Université de Strasbourg

UNIVERSITÉ DE STRASBOURG

École Doctorale des Sciences de la Vie et de la Santé STRASBOURG

ÉCOLE DOCTORALE 414 – SCIENCES DE LA VIE ET DE LA TERRE

Institut de Génétique et de Biologie Moléculaire et Cellulaire

(CNRS UMR 7104 – INSERM U1258)

THÈSE présentée par :

Manon BOIVIN

soutenue le : **08 mars 2022**

pour obtenir le grade de : **Docteur de l'Université de Strasbourg**

Discipline/ Spécialité : Sciences de la vie et de la Santé

ETUDE DES MECANISMES MOLECULAIRES RESPONSABLES DE MALADIES DUES A DES EXPANSIONS DE REPETITIONS SITUEES DANS DES REGIONS DU GENOME DITES « NON-CODANTES »

THÈSE dirigée par :

M CHARLET-BERGUERAND Nicolas Directeur de Recherche, Université de Strasbourg

RAPPORTEURS :

Mme MARTINAT Cécile Directrice de Recherche, Université Paris-Saclay

M KABASHI Edor Directeur de Recherche, Université Paris

AUTRE MEMBRE DU JURY :

Mme ROUAUX Caroline Chargée de Recherche, Université de Strasbourg

Remerciements

Je tiens tout d'abord à remercier les membres de mon jury, Dr Rouaux, Dr Martinat, Dr Kabashi, Pr Depienne et Pr Delage-Mourroux, pour avoir accepté d'évaluer mon travail de thèse.

Ce travail a bénéficié du soutien du Ministère de l'Enseignement Supérieur et de la Recherche, de l'Université de Strasbourg et de la Fondation pour la Recherche Médicale.

Je tiens à remercier également les membres de mon comité de suivi de thèse, Dr Dupuis, Dr Alpy et Dr Vitale pour les conseils qu'ils m'ont donnés tout au long de ma thèse.

Je remercie mon directeur de thèse, le Dr Nicolas Charlet.
Chef, merci de m'avoir laissé ma chance pour ce stage de master 2 alors que je faisais du « hamster » comme tu disais (en espérant que ce soit sans regret !) puis, d'avoir supervisé ma thèse. Tu m'as appris énormément de choses pendant ces années, j'ai apprécié les nombreuses discussions scientifiques et (la plupart) de tes idées que tu as pu proposer pour avancer sur certains projets compliqués. Tu as su me guider et rester positif, notamment sur la mitophagie. Merci pour tout !

Je remercie tous les membres actuels et passés de l'équipe « Charlet », Chantal, Angéline, Camille, David et plus particulièrement Véronique. Merci pour ta bonne humeur et tes conseils, tu m'as accompagné tout au long de ma thèse et c'était un véritable plaisir ! Même si tu accapares les fèves des galettes des rois année sur année, on t'aime quand même !

Hé non David, je ne t'oublie pas ! Alala sacré Dave, par où commencer ? Tu resteras le collègue avec qui j'ai le plus rigolé et fait le plus de soirées, mais « entre collègues hein, on n'est pas potes non plus ». J'ai adoré nos pauses vapot' où on refaisait le monde avec des « si », même si ces mondes n'étaient vraiment pas plus beaux. Je n'abandonne pas l'espoir de réussir ce fameux pas de danse ! On a fait un super duo, surtout en chant (« sous le veeeeeeeeent »).

Je remercie l'ensemble de mes stagiaires, soit pour m'avoir bien aidé, soit pour m'avoir bien fait rire (avec le recul). Merci à Tamara pour son expertise sur toutes les maladies que je pense avoir, entre hypochondriaques, on se comprend ! J'ai une pensée particulière pour Léa, qui a fourni un excellent travail sur l'OPDM, et avec qui j'ai passé de supers moments, autant au boulot et qu'à l'extérieur ! Les journées semblaient passées beaucoup plus vite en ta compagnie, surtout à l'animalerie.

Je tiens à remercier l'ensemble des personnes travaillant sur les plateformes de l'IGBMC avec qui j'ai travaillé de près ou de loin.

Merci à l'équipe de spectrométrie de masse, Luc, Franck et Bastien, qui nous a énormément aidé grâce à leur recherche de site d'initiation ou d'interactants.

Merci à toute l'équipe de microscopie, Erwan, Elvire, Bertrand, Marine et Yves qui ont fini par me convaincre d'abandonner ce sublime microscope qu'était le DMRXA2 pour prendre de « vraies images ». Je remercie particulièrement Erwan, pour sa patience et sa bonne humeur contagieuse, lors de nos sessions d'acquisition. Merci à Nadia, Jean-Luc et Coralie pour leur expertise en MET et pour avoir obtenu de si jolis résultats.

Merci à tous les membres de l'équipe d'histopathologie, Hugues, Olivia, Patrice etc. pour leur travail et leur expertise, notamment sur le sujet OPDM et ; à toutes les personnes travaillant dans la zone phénotypage de l'ICS, l'équipe comportement, Fabrice, Aline, Christophe et Raphaël, et l'équipe métabo, ainsi que Isabelle, pour leurs conseils et leur aide.

Je remercie également Claudine et Muriel du FACS pour leur expertise et leur aide pour le développement des protocoles que j'ai utilisé, que ce soit sur le tri ou l'analyse.

Merci au service d'anticorps, notamment Mustapha ; aux personnes de la plateforme de biologie moléculaire, Paola et Pascale, pour le développement des plasmides, des lentivirus et des AAV.

Merci à toutes les personnes du service de culture qui ont largement facilité mon travail durant ma thèse.

Je tiens à remercier également Rachida du service RH pour son efficacité. Je tiens également à remercier les personnes du service achat, du stock et des petits produits qui nous permettent de travailler dans de bonnes conditions. Merci à Eveline et Maïté pour leur bonne humeur et leur aide précieuse avec cet engin du diable qu'est la C60.

Je remercie toutes les personnes que j'ai croisées et côtoyées dans l'institut. Merci à toutes les personnes du 1er, à Pascal pour toutes ces bonnes blagues ; à Florence pour nos discussions en salle culture. ; à Salvatore et Nemanja avec qui j'ai bien rigolé en faisant le « Get Together » et m'ont laissé la grande responsabilité de servir des bières et à Reuben pour nos petites pauses dans le froid.

Nicolas (pas le chef, le petit), merci pour ta bonne humeur de tous les jours ! Avec toi, on passe la journée en chanson, et pas n'importe lesquelles, des vraies chansons françaises monsieur ! Merci pour ton courage et ta gentillesse, je pense au jour où tu es venu me chercher en panne sous une couche de neige impressionnante, et ceci sans pneu neige, ça patinait dur ! J'ai aimé tes cours de danse « au coin du sapin ».

Alexia, merci pour tous les moments qu'on a passés ensemble ! Tu es mon binôme sportif, on le fera ce triathlon. Rentrer sous la pluie à vélo, courir par -10 degrés, nager dans de l'eau gelée, rien ne nous arrête ! Et merci pour toutes ces soirées où on a bien décompressé, où a fait du limbo jusqu'à « s'en casser le nez » !

Merci également aux copains du master, Dr Kenny, Dr Vincent et (bientôt Dr) Amélie. Vous avez brillamment réussi votre thèse, j'espère suivre votre chemin. J'ai passé d'excellents moments avec vous, que ce soit au RU (même s'il y avait beaucoup trop de curry à mon goût).

Je tiens également à remercier tous mes amis qui m'ont permis d'avoir une vie à côté de ma thèse, j'aurais aimé vous voir plus souvent.

Merci à Maxime, Arthur, Anna, Sam, François, Quentin, Guich, Yohann, Julian et Emilie, la bande de Franche-Comté avec qui j'ai eu le plaisir de passer de superbes vacances et des nouvels ans inoubliables depuis des années !

Léopold, merci pour tout. Depuis qu'on se connaît, je n'ai que des moments de rigolade avec toi, et pour ça, je ne regrette pas d'avoir fait le master Ecophy.

J'aimerais terminer par remercier ma famille, qui m'a toujours soutenu et encouragé. Merci à mes grands-parents, Jeannine et Pierre, pour leur générosité. Merci à ma maman, qui m'a constamment encouragé pendant mes recherches sur la sclérose en plaque. Merci à mon papa, qui me propose des destinations intéressantes pour mon post-doc comme l'Australie ou les Maldives pour pouvoir me rendre visite. Merci à mon frère, Léo, qui comprend sur quoi je travaille, et qui n'a donc pas fait de thèse et est allé travailler en Suisse. *Cette thèse est pour vous…*

Table des matières

Abréviation **8**

Liste des Figures **14**

Liste des Tableaux **15**

INTRODUCTION

1. LES MALADIES A EXPANSION DE REPETITIONS **16**

1.1. Généralités **16**

1.2. Les maladies dues à une toxicité au niveau ADN **19**

1.2.1. Le syndrome de l'X fragile (FXS) 19

1.2.2. Le syndrome Baratela-Scott (BSS) 21

1.2.3. L'ataxie de Friedreich (FRDA) 21

1.2.4. La dystonie parkinsonisme liée à l'X (XDP) 22

1.3. Les maladies dues à une toxicité au niveau ARN **23**

1.4. Les maladies dues à une toxicité au niveau protéique **25**

1.4.1. L'initiation de la traduction chez les eucaryotes 26

1.4.1.1. Initiation de la traduction à un codon AUG 27

1.4.1.2. Initiation de la traduction à des codons ressemblant à des AUG 29

1.4.1.3. Initiation de la traduction au niveau de séquences IRES 32

1.4.1.4. Traduction « RAN » des répétition non associées à un codon AUG 33

1.4.2. Les maladies à polyGlutamine (polyQ) 35

1.4.2.1. Les protéines à polyQ sont-elles toxiques d'elles-mêmes ? 36

1.4.2.2. Toxicité nucléaire ou cytoplasmiques des protéines à polyQ ? 37

1.4.2.3. L'agrégation des protéines à polyQ est-elle nécessaire pour leur toxicité ? 39

1.4.2.4. Gain de fonction et / ou perte de fonction ? 40

1.4.2.5. Conclusion 41

1.4.3. Les maladies à polyAlanine (polyA) 42

1.4.4. Les maladies à polyGlycyne (polyG) 44

1.4.4.1. Le syndrome de tremblement et d'ataxie lié à l'X fragile (FXTAS) 44

1.4.4.2. L'insuffisance ovarienne prématurée liée à l-X (FXPOI) 48

1.5. Les maladies complexes **49**

1.6. Les maladies à mécanisme encore inconnu **51**

2. MALADIES A EXPANSIONS CGG DANS LES REGIONS 5'UTR **53**

2.1. La maladie à inclusions intranucléaires neuronaux (NIID) **53**

2.1.1. Epidémiologie et clinique 53
2.1.2. Histopathologie 54
2.1.3. Mutation responsable de NIID 55
2.2. Les myopathies oculo-pharyngo-distales (OPDM) 57
2.2.1. Epidémiologie et clinique 57
2.2.2. Histopathologie 58
2.2.3. Mutations responsables de l'OPDM 59
2.3. La myopathie oculo-pharyngée avec leucodystrophie (OPML) 61
Revue. Trinucleotide CGG repeat diseases: an expanding field of polyglycine proteins ? 61
2.4. Hypothèses et objectifs 62
3. LA SCLEROSE LATERALE AMYOTROPHIQUE (SLA) ET LA DEMENCE FRONTO-TEMPORALE (DFT) ..63
3.1. Epidémiologie et clinique 63
3.1.1. La sclérose latérale amyotrophique (SLA) 63
3.1.2. La démence fronto-temporale (DFT) 64
3.1.3. SLA et DFT : un continuum 65
3.2. Histopathologie 66
3.3. Mutations 68
3.4. Expansion d'hexa-nucléotides GGGGCC dans le gène *C9ORF72* 70
3.4.1. Hypothèse de la perte de fonction de la protéine C9ORF72 72
3.4.1.1. Fonction de la protéine C9ORF72 73
3.4.1.2. Implication de la protéine C9ORF72 dans l'autophagie 76
3.4.1.2.1. L'autophagie 76
3.4.1.2.2. Rôle potentiel de C9ORF72 dans l'autophagie 81
3.4.2. Hypothèse d'un gain de fonction de l'ARN 83
3.4.3. Hypothèse d'un gain de fonction de protéines composées de dipeptides répétés 84
3.4.3.1. Toxicité de la protéine DPR polyGlycine-Alanine (polyGA) 88
3.4.3.2. Les protéines DPR non-toxiques : polyGlycine-Proline et polyProline-Alanine 90
3.4.3.3. Les protéines DPR toxiques : polyGlycine-Arginine et polyProline-Arginine 90
3.4.4. Exemple de modèles exprimant les répétitions d'hexa-nucléotides GGGGCC 92
3.5. Hypothèses et objectifs 94
4. STRATEGIES THERAPEUTIQUES 95
4.1. Molécules liant l'ADN et / ou l'ARN 95
4.2. Molécules activant l'autophagie 97
4.3. Anticorps dirigés contre les protéines toxiques 101
4.4. Oligonucléotides antisens (ASO) 102

RESULTATS

Article 1. Translation of GGC repeat expansions into a toxic polyglycine protein in NIID defines a novel class of human genetic disorders: The polyG diseases 106

Résultat sur l'OPDM2 107

1. Traduction des répétitions CGG présentes dans la région 5'UTR du gène *GIPC1* 107

2. Mécanisme de traduction de la protéine uGIPC1polyG 109

3. Détection de la protéine uGIPC1polyG 111

4. Modèle mutin exprimant la protéine uGIPC1polyG 113

Article 2. Reduced autophagy upon C9ORF72 loss synergizes with dipeptide repeat protein toxicity in G4C2 repeat expansion disorders 114

DISCUSSION

1. LES MALADIES A POLY-GLYCINE 115

1.1. Autres expansions CGG et maladies à polyG 117

1.2. Taille de l'expansion CGG et sévérité 118

1.3. Mécanisme de toxicité des protéines à polyG 120

1.3.1. Niveau d'expression des protéines à polyG 120

1.3.2. Toxicité tissu spécifique des protéines à polyG ? 120

1.3.3. Les protéines à polyG sont-elles toxiques d'elles-mêmes ? 122

1.3.4. Localisation des protéines à polyG et toxicité 124

1.3.5. L'agrégation des protéines à polyG est-elle nécessaire pour leur toxicité ? 125

1.3.6. Perte de fonction et / ou gain de fonction ? 125

1.3.7. Mécanisme de toxicité des protéines à polyG 127

1.3.8. Conclusion 127

2. EXPANSION DE REPETITIONS GGGGCC DANS LE GENE *C9ORF72* DANS LA SLA/DFT 129

2.1. Synergie de toxicité entre perte de C9ORF72 et protéines DPR 129

2.2. Rôles de protéine C9ORF72 131

2.2.1. Rôle de la protéine C9ORF72 au lysosome 131

2.2.1. Rôle de la protéine C9ORF72 aux LDVC 134

2.3. La toxicité des protéines DPR 137

2.3.1. Toxicité de la protéines DPR polyGA 138

2.3.2. Toxicité de la protéines DPR polyGP 140

2.3.3. Détection des protéines DPR chez les patients C9-SLA/DFT 142

3. TRAITEMENTS.....143
3.1. Thérapie par oligonucléotides antisens (ASO) 143
3.1. Thérapie par molécules activant l'autophagie 144
4. MECANISMES DE TRADUCTION DES REPETITIONS.....145
4.1. Mécanisme de traduction des protéines DPR dans la SLA/DFT 145
4.2. Mécanisme de traduction des protéines à polyG 147
4.3. La traduction de régions dites « non-codantes » 150
5. CONCLUSION.....154

Références bibliographiques.....155

Abréviation

A

AD	Maladie d'Alzheimer
ALR	Reformation des lysosomes après autophagie
ADARB2	Adenosine Deaminase RNA Specific B2
ALYREF	Aly/REF Export Factor
AMPK	AMP-activated protein kinase
AR	*Androgen Receptor*
ARF	ADP-ribosylation factor
ARL	ADP-ribosylation factor-like
ARNlnc	ARN long et non-codant
ARX	Aristaless related homeobox
ASO	Oligonucléotides antisens
ATF	Activating Transcription Factor
ATG	*Autophagy Related Gene*
ATN1	*Atrophin 1*
ATXN	Ataxin

B

B6	Black-6
BAC	Chromosome Artificiel Bactérien
BEAN1	*Brain Expressed Associated With NEDD4 1*
BECN1	Beclin-1
BER	Réparation par Excision de Bases
BIN1	Bridging Integrator 1
BiP	Binding immunoglobulin protein
BPES	Syndrome de Blépharophimosis-Ptosis-Epicanthus inversus
BSS	Syndrome Baratela-Scott

C

C9ORF72	*Chromosome 9 Open Reading Frame 72*
CACNA1A	*Calcium voltage-gated channel subunit alpha1 A*
CANVAS	Syndrome ataxie cérébelleuse avec neuropathie, aréflexie vestibulaire bilatérale
CCHS	Syndrome d'hypoventilation alvéolaire central congénital
ChGA	Chromogranine A
CIC	Capicua
CLCN1	Chloride channel protein, skeletal muscle
CMA	Autophagie Médiée par les protéines Chaperonnes
Cryo-EM	Cryo-Microscopie Electronique
CSW	C9ORF72-SMCR8-WDR41
CUGBP1	CUG triplet repeat, RNA binding protein 1

D

DAB1	*Disabled-1*
DAG	Diacylglycérol
DBQD2	Dysplasie de Dubesquois de type 2
DENN	Differentially Expressed in Normal and Neoplastic cells
DEPTOR	DEP Domain Containing MTOR Interacting Protein

DFCP1	Double FYVE-containing protein 1
DFT	Démence Fronto-Temporale
DGKκ	Diacylglycerol kinase kappa
DM	Dystrophie Myotonique
DMPK	*DM1 protein kinase*
DPR	DiPeptides Répétés
DRPLA	Atrophie Dentato-Rubro-Pallido-Luysienne
DWORF	Dwarf Open Reading Frame

E

eIF	Facteur d'Initiation eucaryote
EMCV	Virus de l'encéphalomyocardite

F

FAME	Epilepsies Myocloniques Familiales de l'Adulte
FDA	Administration de la nourriture et des médicaments
FECD3	Dystrophie cornéenne endothéliale de Fuchs de type 3
FGF2	Fibroblast Growth Factor 2
FIP200	FAK family kinase-Interacting Protein of 200 kDa
FLNC	Filamin C
FMR1	*Fragile X Mental Retardation 1*
FMRP	Fragile-X Mental Retardation Protein
FNIP2	Folliculin Interacting Protein 2
FOXL2	*Forkhead Box L2*
FRAXE	Syndrome de retard mental non-symptomatique lié au chromosome X
FRDA	Ataxie de Friedreich
FUS	Fused in Sarcoma
FVB	Friend leukemia virus B
FXPOI	Syndrome de l'insuffisance ovarienne liée l'X
FXN	Frataxine
FXS	Syndrome de l'X fragile
FXTAS	Syndrome de tremblement et d'ataxie lié à l'X fragile

G

GABARAP	GABA Type A Receptor-Associated Protein
GABARAPL	GABA Type A Receptor-Associated Protein Like
GAP	GTPase-activating protein
GCN4	*General Control Nondepressible*
GEF	GDP/GTP Exchange Factor
GFP	Green Fluorescent Protein
GIPC1	*GIPC* PDZ Domain Containing Family Member 1
GRN	*Granulin Precursor*
GT1-7	Mouse Hypothalamic GnRH Neuronal

H

HD	Maladie de Huntington
HDL2	Maladie de Huntington de type 2
HEK293	human embryonic kidney 293
HeLa	Henrietta Lacks
HFGS	Syndrome main-pied-génital

HOXA13	*Homeobox A13*
HOXD13	*Homeobox D13*
hnRNP	Heterogeneous nuclear Ribonucleoprotein
HPE5	Holoprosencéphalie de type 5
HPRT	*Hypoxanthine Phosphoribosyltransferase 1*
HR23	Ubiquitin receptor RAD23
HTT	Huntingtin

I

IHC	Immunohistochemistry
iMN	Induced MotoNeuron
INSR	Insulin Receptor
iPS	Induced Pluripotent Stem Cells
IREs	Internal Ribosome Entry Site
IRM	Imagerie par Résonnance Magnétique
ISG	Granules de Sécrétion Immatures

J

JPH3	*Junctophilin 3*

K

KH	Homologie K
KI	Knock In
KO	Knock Out
KU70	ATP-dependent DNA helicase 2 subunit KU70
KU80	ATP-dependent DNA helicase 2 subunit KU80

L

Lap2β	Lamina-associated polypeptide 2β
LC3	Microtubule Associated Protein 1 Light Chain 3
LDCV	Large Dense Core Vesicle
LRP12	*LDL Receptor Related Protein 12*

M

m^7G	7-méthyl-guanosine
MAPT	*Microtubule Associated Protein Tau*
MARCHF6	*Membrane Associated Ring-CH-Type Finger 6*
MBNL	Muscle blind-like
MIEF1	Mitochondrial dynamics protein MID51
MEF	Mouse Embryonic Fibroblast
mLST8	MTOR Associated Protein, LST8 Homolog
MOTS-c	mitochondrial ORF of the 12S rRNA type-**c**
MMR	Réparation des Mésappariements
MRXS33	Retard mental lié à l'X, syndromique 33
mTOR	Mammalian Target Of Rapamycin
mTORC1	Complexe mTOR 1

N

NA	Naphthyridine-Azaquinolone

NBR1	Neighbor of BRCA1 gene 1
NCP	10-(4′-(N-diethylamino)butyl)-2-chlorophenoxazine
NDP52	Nuclear Domain 10 Protein 52
NIID	Maladie Neuronale à Inclusions Intranucléaires
NLS	Séquence de Localisation Nucléaire
NOTCH2NLC	*Notch 2 N-Terminal Like C*
NPM1	Nucleophosmin 1
NS-XLID	Déficience Intellectuelle Non-Syndromique Liée à l'X

O

OPDM	Myopathie Oculo-Pharyngo-Distale
OPMD	Dystrophie Musculaire Oculo-Pharyngée
OPML	Myopathie Oculo-Pharyngée avec Leucodystrophie
OPTN	Optineurin
ORF	Open Reading Frame

P

PA	Acide Phosphatidique
PABP	Poly(A) Binding Protein
PABPN1	*Poly(A) Binding Protein Nuclear 1*
PE	Phosphatidyléthanolamine
PERK	PKR-like ER kinase
PHOX2B	*Paired Like Homeobox 2B*
PI	Phosphoinositide
PI3KC3	Class III phosphatidylinositol 3-kinase complex 3
PI3P	Phosphatidylinositol-3-phosphate
PIC	Complexe de Pré-Initiation
PIGBOS	PIGB opposite strand 1
POF	Insuffisances ovariennes primaires
polyA	PolyAlanine
polyG	PolyGlcyine
polyQ	PolyGlutamine
polyS	PolySérine
PQLC2	PQ loop repeat containing 2
PRAS40	Proline-rich Akt substrate of 40 kDa
PSCM4	Proteasome 26S subunit, ATPase 4
Purα	Purine rich element binding protein A

R

RA	Récepteur aux Androgènes
RAB	Ras-related proteins in brain
RAD001	Everolimus
Rag	Ragulator-Rag complex
Ran	Ras-related nuclear protein
RAN	Repeat Associated Non-AUG translation
RAPGEF2	*Rap Guanine Nucleotide Exchange Factor 2*
RAPTOR	Regulatory Associated Protein Of MTOR Complex 1
Ras	Virus du sarcome du rat
RILPL1	*Rab Interacting Lysosomal Protein Like 1*
RFC1	*Replication Factor C Subunit 1*

rbFOX1	RNA Binding Fox-1 Homolog 1
RUNX2	*RUNX Family Transcription Factor 2*

S

SAM68	Src associated in mitosis, of 68 kDa
SAMD12	*Sterile Alpha Motif Domain Containing 12*
SBMA	Atrophie Musculaire Spinale et Bulbaire
SCA	Ataxie spinocérébelleuse
SCN5A	Sodium voltage-gated channel alpha subunit 5
SFPQ	Splicing Factor Proline And Glutamine Rich
SLA	Sclérose Latérale Amyotrophique
SLAF	Sclérose Latérale Amyotrophique Familiale
SLAS	Sclérose Latérale Amyotrophique Sporadique
SMA	Amyotrophie spinale proximale
SMCR8	Smith Magenis syndrome Chromosome Region candidate 8
SMN	*Survival of motor neuron*
smORF	Small Open Reading Frame
SNAP29	Synaptosome Associated Protein 29
SNARE	Soluble N-éthylmaleimide-sensitive-factor Attachment protein REceptor
SOD1	Superoxide Dismutase 1
SOX3	*SRY-Box Transcription Factor 3*
SPD	Syndactilie de type 1 et 5
SQSTM1	*Sequestosome 1*
SRSF	Serine/arginine-rich splicing factor
STARD7	*Star Related Lipid Transfer Domain Containing 7*
STX17	Syntaxin 17
SV2	Synaptic vesicle-associated protein 2

T

TAF1	Transcription initiation factor TFIID subunit 1
TARDBP	TAR DNA binding protein
TAX1BP1	Tax1 Binding Protein 1
TBK1	*TANK Binding Kinase 1*
TBP	TATA-binding protein
TC	Complexe Ternaire
TCF4	Transcription factor 4
TDP-43	Transactivating response DNA binding protein of 43 kD
TE	Tremblement Essentiel
TFIID	Transcription Factor II D
TFEB	Transcription Factor EB
TK2	*Thymidine Kinase 2*
TMPyP4	Tetra-(N-methyl-4-pyridyl)porphyrin
TNNT2	Troponin T2, cardiac type
TNRC6A	*Trinucleotide Repeat Containing Adaptor 6A*

U

U2OS	Homo sapiens bone osteosarcoma
Ube3A	Ubiquitin Protein Ligase E3A
UBQLN2	Ubiquilin 2
ULK	Unc-51 Like Autophagy Activating Kinase

Unc	Uncoordinated
uORF	Upstream Open Reading Frame
UTR	Untranslated Region

V

VAMP	Vesicle Associated Membrane Protein
VCP	Valosin Containing Protein
VHC	Virus de l'Hépatite C
VPS15	Phosphoinositide-3-Kinase Regulatory Subunit 4
VPS34	Phosphatidylinositol 3-kinase catalytic subunit type 3

W

WDR41	WD repeat containing protein 41
WIPI	WD Repeat Domain, Phosphoinositide Interacting 1
WT	Wild Type (sauvage)

X

XDP	Dystonie-Parkinsonisme liée à l'X
XH	Hypopituitarisme lié à l'X
XYLT1	Xylosyltransferase 1

Y

YEATS2	*YEATS Domain Containing 2*

Z

ZIC2	*Zinc finger protein of the cerebellum 2*

Liste des Figures

Figure 1. Maladies causées par des expansions de répétitions 16
Figure 2. Mécanisme de toxicité au niveau ADN 19
Figure 3. Mécanisme de toxicité au niveau ARN 23
Figure 4. Toxicité au niveau protéique 25
Figure 5. Traduction chez les eucaryotes 26
Figure 6. Initiation de la traduction chez les eucaryotes 28
Figure 7. Utilisation des codons pour l'initiation de la traduction dans les ORF et uORF 30
Figure 8. Régulation positive en condition de stress des uORF sur les gènes *ATF4* et *ATF5* 31
Figure 9. Les 4 principales types d'IRES 33
Figure 10. Mécanisme de traduction « RAN » 34
Figure 11. Histopathologie des patients touchés par les maladies à polyQ 35
Figure 12. Fonctionnement du récepteur aux androgènes (RA) 38
Figure 13. La toxicité de l'ATXN1-polyQ dépend de sa liaison à la protéine CIC 40
Figure 14. Caractéristiques histopathologiques des patients OPMD 43
Figure 15. Hyper-intensité en T2 des pédoncules cérébelleux dans FXTAS 44
Figure 16. Inclusions intranucléaires chez des patients FXTAS 45
Figure 17. FMRpolyG dans les inclusions intranucléaires chez les patients FXTAS 46
Figure 18. Modèles murins de FXTAS prouvant la toxicité de FMRpolyG 47
Figure 19. Mécanisme de toxicité proposé pour SCA31 49
Figure 20. IRM de patients FXTAS et NIID 54
Figure 21. Inclusions intranucléaires dans le cerveau des patients NIID 55
Figure 22. Rôle des paralogues de NOTCH2 dans le développement cortical humain 56
Figure 23. Evolution d'un patient atteint d'OPDM 57
Figure 24. Caractéristiques histopathologiques des fibres musculaires des patients OPDM 58
Figure 25. IRM de cerveaux de patients FXTAS, NIID et OPML. 61
Figure 26. Les 2 formes de SLA 64
Figure 27. Zones du cerveau atteintes chez les patients DFT 65
Figure 28. Présence d'agrégats de la protéine TDP-43 chez des patients SLA et DFT 66
Figure 29. Perte des neurones moteurs chez un patient atteint de SLA 67
Figure 30. Gènes mutés chez les patients atteins de SLA, de DFT ou de SLA/DFT 68
Figure 31. Principaux gènes impliqués dans la SLA et la DFT 68
Figure 32. Mécanismes de toxicité de l'expansion GGGGCC dans le gène *C9ORF72* 71
Figure 33. Epissage alternatif de C9ORF72 produisant 2 isoformes protéiques 72
Figure 34. La superfamille des GTPases Ras 74
Figure 35. Conformation des petites GTPases Ras 75
Figure 36. Structure des complexes C9ORF72-SMCR7-WDR41 et FLNC-FNIP2 75
Figure 37. Les 3 formes d'autophagie 76
Figure 38. Protéines impliquées dans la cascade de signalisation dans l'autophagie 77
Figure 39. Incorporation des lipides au niveau de l'autophagosome 78
Figure 40. Systèmes de conjugaison impliqués dans le recrutement des protéines LC3 à l'autophagosome 79
Figure 41. Foci d'ARN compose de répétitions GGGGCC chez les patients SLA/DFT 83
Figure 42. Traduction des ARN sens et antisens en protéines DPR 84
Figure 43. Présence des différentes protéines DPR chez un patient C9-SLA/DFT 85
Figure 44. Quantification des agrégats de protéines DPR dans le cerveau de patients SLA 86

Figure 45. Voies cellulaires altérées par les protéines DPR. 87
Figure 46. PolyGA séquestre le protéasome 88
Figure 47. La protéine NPM1 est délocalisée par les protéines polyGR et polyPR 91
Figure 48. Molécules activant l'autophagie 97
Figure 49. Composés de la classe des tricycles activant ou non l'autophagie 99
Figure 50. Mécanismes d'action des ASO 102
Figure 51. Action du Nusinersen sur l'épissage de SMN2 103
Figure 52. Traduction des expansions de répétitions CGG du gène *GIPC1* 108
Figure 53. Traduction canonique des répétitions CGG dans le cadre de lecture glycine 110
Figure 54. Impact de l'expression des protéines à polyGlycine 112
Figure 55. Expansions de répétitions traduites en protéines toxiques 116
Figure 56. Arbre généalogique d'une famille OPDM 119
Figure 57. Les maladies à polyG : maladies neurodégénératives et neuromusculaires 121
Figure 58. Expression de uN2CpolyG-GFP est ATG-polyG-GFP dans le cerveau de souris 123
Figure 59. Mécanisme de toxicité des protéines à polyG 128
Figure 60. Implication du complexe CSW dans l'autophagie 133
Figure 61. Rôle du complexe CSW pour les vésicules LDCV 135
Figure 62. Phénotype et histopathologie des souris WT et *C9orf72* SNC-KO exprimant la protéine polyGA 139
Figure 63. Phénotype locomoteur des souris exprimant la protéine polyGP 141
Figure 64. Anticorps dirigés contre les protéines polyGA et polyGP 142
Figure 65. Traduction des répétitions CGG dans le gène LRP12 149
Figure 66. Profil de migration des protéines à polyA 150

Liste des Tableaux

Tableau 1. Maladies à expansion de répétitions 18
Tableau 2. Efficacité de l'initiation de la traduction selon le codon initiateur 29
Tableau 3. Gènes mutés responsables de la SLA, la DFT ou la SLA/DFT 69

INTRODUCTION

INTRODUCTION

- CHAPITRE 1 -

Les maladies à expansion de répétitions

1. LES MALADIES A EXPANSION DE REPETITIONS

1.1. Généralités

C'est il y a maintenant 30 ans qu'a été démontré, pour la première fois, que des expansions de microsatellites étaient responsables de maladies génétiques chez l'Homme. Ces équipes ont respectivement mis en lumière que l'amyotrophie bulbospinale liée à l'X, ou maladie de Kennedy, est causée par une expansion de nucléotides CAG dans le récepteur aux androgènes (La Spada et al, 1991), et que le syndrome de l'X fragile (FXS) est dû à une expansion de nucléotides CGG dans le gène *FMR1* (Oberlé et al., 1991 ; Verkerk et al., 1991). Ces travaux ont permis d'établir une nouvelle catégorie de maladies génétiques : les **maladies à expansion de répétitions**. Ces expansions de répétitions peuvent être composées de tri-, tétra-, penta-, hexa-nucléotides dans les régions codantes ou non-codantes (5'UTR, intron, 3'UTR) des gènes (Figure 1).

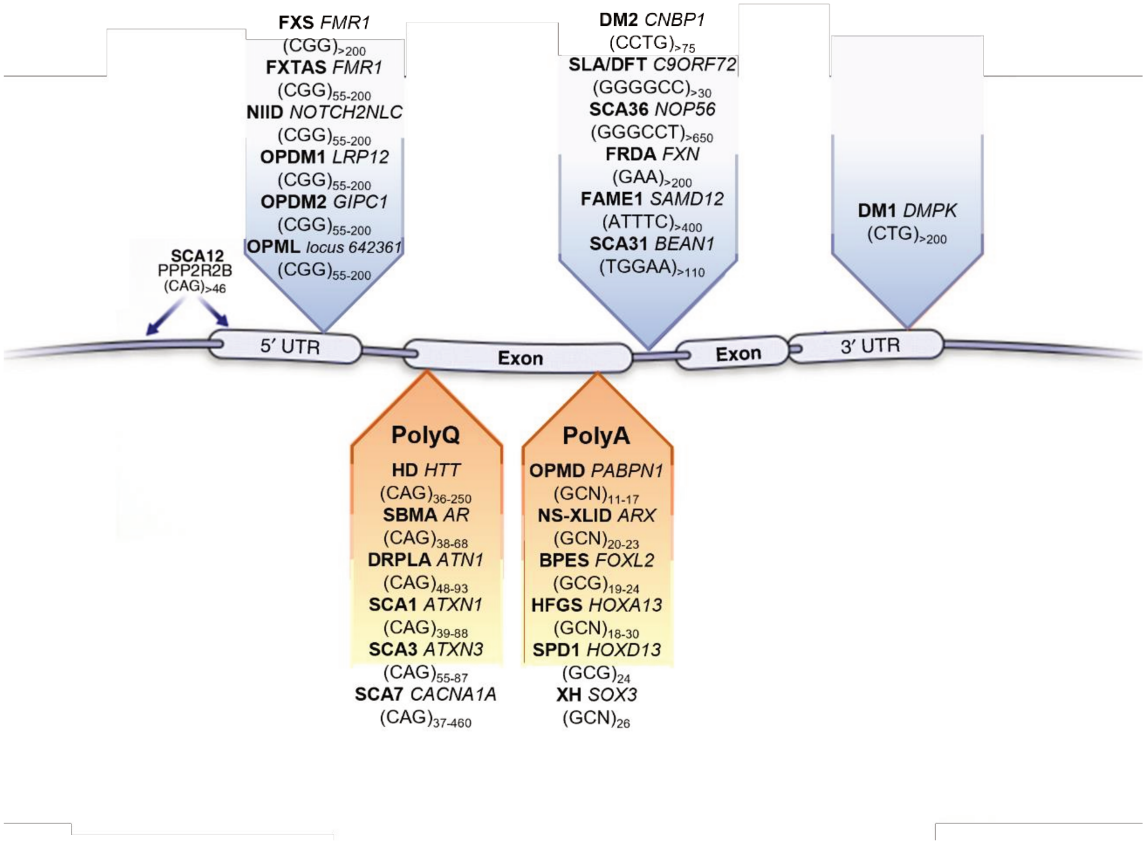


Figure 1. Maladies causées par des expansions de répétitions
Composition, localisation et taille des expansions de répétitions impliquées dans différentes pathologies. Adapté de Ellerby, 2019.

Le nombre de répétitions varie en fonction de la maladie, du patient et de son âge, du tissu considéré mais aussi lors de la transmission aux descendants, permettant ainsi de qualifier ces répétitions d'instables et dynamiques. De ce fait, des familles atteintes par certaines de ces mutations dynamiques peuvent connaitre un phénomène d'anticipation caractérisé par une augmentation de la taille des répétitions conduisant à une plus grande sévérité de la maladie et à une diminution de l'âge d'apparition des premiers symptômes lors du passage d'une génération à la suivante (ex : FXS, HD, DM1). La naissance, l'allongement, ou plus rarement le raccourcissement des expansions de répétitions fait intervenir des mécanismes de réplication et de réparation de l'ADN, notamment la réparation des mésappariements (MMR) et la réparation par excision de bases (BER) (pour revue, Zhao et al., 2015a).

A ce jour, pas moins de 50 maladies sont associées à des expansions de répétitions dont la moitié identifiée lors de ces 10 dernières années grâce au progrès en termes de séquençage du génome (Tableau 1 ; pour revues, Depienne et Mandel, 2021 ; Malik et al., 2021). De nombreux efforts de recherche ont permis d'identifier 3 mécanismes moléculaires à l'origine de ces pathologies, selon la localisation et le nombre de répétitions :

- Le **mécanisme de toxicité au niveau ADN** (FXS, BSS, FRDA, XDP, etc) ;
- Le **mécanisme de toxicité au niveau ARN** (DM1 et DM2) ;
- Le **mécanisme de toxicité au niveau protéique** (HD, OPMD, SLA, FXTAS, etc).

Année de découverte	Maladie	Gène	Région génique	Motif répété	Taille d'expansion normale	Taille d'expansion pathologique	Equipe
1991	SBMA	*AR*	Exon	CAG	9 - 36	38 - 68	La Spada et al.
1991	FXS	*FMR1*	5' UTR	CGG	5 -50	> 200	Oberlé et al. ; Verkerk et al.
1992	DM1	*DMPK*	3' UTR	CTG	5 - 37	50 - 10 000	Mahadevan et al.
1993	HD	*HTT*	Exon	CAG	6 - 35	36 - 250	The Huntington's Collaborative Research Group
1993	SCA1	*ATXN1*	Exon	CAG	6 - 38	39 - 88	Orr et al.
1993	FRAXE	*AFF2*	5' UTR	CCG	4 - 39	200 - 900	Knight et al.
1994	DRPLA	*ATN1*	Exon	CAG	3 - 35	48 - 93	Koide et al. ; Nagafuchi et al.
1994	SCA3	*ATXN3*	Exon	CAG	12 - 44	55 - 87	Kawaguchi et al.
1996	SCA2	*ATXN2*	Exon	CAG	13 - 31	32 - 500	Pulst et al.
1996	SCA7	*ATXN7*	Exon	CAG	4 - 33	37 - 460	Lindblad et al.
1996	SPD1	*HOXD13*	Exon	GCG	15	24	Akarsu et al.
1996	FRDA	*FXN*	Intron	GAA	5 - 34	65 - 1300	Campuzano et al.
1997	SCA6	*CACNA1A*	Exon	CAG	4 - 18	20 - 33	Zhuchenko et al.
1997	BCCD	*RUNX2*	Exon	GCN	17	27	Mundlos et al.
1997	EPM1	*CSTB*	5' UTR	CCCCGCCCCGCG	2 - 3	30 - 75	Lalioto et al.
1998	OPMD	*PABPN1*	Exon	GCG	6 - 10	12 - 17	Brais et al.
1998	FXPOI	*FMR1*	5' UTR	CGG	5 - 50	55 - 200	Conway et al.
1999	SCA17	*TBP*	Exon	CAG	25 - 40	43 - 66	Koide et al.
1999	SCA8	*ATXN8*	Exon	CAG	15 - 50	74 - 250	Koob et al.
1999	SCA12	*PPP2R2B*	5' UTR	CAG	4 - 32	43 - 78	Holmes et al.
2000	HFGS	*HOXA13*	Exon	GCN	12 - 18	18 - 30	Goodman et al.
2000	SCA10	*ATXN10*	Intron	ATTCT	10 - 32	280 - 4500	Matsuura et al.
2001	HDL2	*JPH3* AS	Exon	CAG	6 - 28	41 - 58	Margolis et al.
2001	BPES	*FOXL2*	Exon	GCN	14	19 - 24	De Baere et al.
2001	HPES	*ZIC2*	Exon	GCN	15	25	Brown et al.
2001	DM2	*CNBP*	Intron	CCTG	11 - 30	50 - 11 000	Liquori et al.
2001	FXTAS	*FMR1*	5' UTR	CGG	5 - 50	55 - 200	Hagerman et al.
2002	NS-XLID	*ARX*	Exon	GCN	12 - 16	20 - 23	Stromme et al.
2002	XH	*SOX3*	Exon	GCN	15	26	Laumonnier et al.
2003	CCHS	*PHOX2B*	Exon	GCN	20	25 - 29	Amiel et al.
2009	SCA31	*BEAN1 / TK2*	Intron	TGGAA / TTCCA	variable	110 - 760	Sato et al.
2011	SCA36	*NOP56*	Intron	GGGCCT	5 - 14	650 - 2500	Kobayashi et al.
2011	SLA/DFT	*C9ORF72*	Intron	GGGGCC	3 - 25	> 30	Dejesus-Hernandez et al. ; Renton et al.
2015	FECD3	*TCF4*	Intron	CTG	5 - 31	> 50	Mootha et al.
2017	XDP	*TAF1*	Intron	CCCTCT	absent	30 - 55	Bragg et al.
2017	SCA37	*DAB1*	Intron	ATTTC	absent	31 -75	Seixas et al.
2018	FAME1	*SAMD12*	Intron	ATTTC	absent	440 - 3680	Ishiura et al.
2018	FAME6	*TNRC6A*	Intron	ATTTC	absent		Ishiura et al.
2018	FAME7	*RAPGEF2*	Intron	ATTTC	absent		Ishiura et al.
2019	BSS	*XYLT1*	Promoteur	CGG	9 - 20	120 - 800	LaCroix et al.
2019	CANVAS	*RFC1*	Intron	AAGGG	variable	400 - 2000	Cortese et al. ; Rafehi et al.
2019	GDPAG	*GLS*	5' UTR	GCA	8 - 16	680 - 1400	Van Kuilenburg et al.
2019	NIID	*NOTCH2NLC*	5' UTR	CGG	7 - 60	61 - 200	Ishiura et al. ; Sone et al. ; Tian et al.
2019	OPDM1	*LRP12*	5' UTR	CGG	13 - 45	90 - 130	Ishiura et al.
2019	OPML1	*Locus 642361*	ARN lnc	CGG	3 - 16	40 - 60	Ishiura et al.
2019	FAME4	*YEATS2*	Intron	ATTTC	absent		Yeetong et al.
2019	FAME2	*STARD7*	Intron	ATTTC	absent	661 - 735	Corbett et al.
2019	FAME3	*MARCHF6*	Intron	ATTTC	absent	660 - 2800	Florian et al.
2020	OPDM2	*GIPC1*	5' UTR	CGG	12 - 32	97 - 120	Deng et al.
2021	OPDM3	*NOTCH2NLC*	5' UTR	CGG	7 - 60	61 - 200	Yu et al.

Tableau 1. Maladies à expansion de répétitions

Taille et composition de l'expansion de répétitions ainsi que leur localisation dans le gène impliqué dans les différentes pathologies. Adapté de Depienne et Mandel, 2021.

1.2. Les maladies dues à une toxicité au niveau ADN

Ces pathologies sont causées par des expansions de répétitions de nucléotides situées le plus souvent proche du promoteur de leur gène hôte. Il en résulte une diminution de la transcription de ce gène et, donc, à une diminution de l'expression de la protéine codée par ce gène (Figure 2). Cette hypothèse est confirmée par l'observation de patients aux symptômes identiques et présentant des mutations perte de fonction dans la séquence codante du gène, mais ne présentant pas l'expansion de répétitions. De plus, le développement de modèles animaux Knock Out (KO) pour ces gènes touchés par les expansions a permis de confirmer cette théorie.

Parmi ces maladies, nous pouvons citer le syndrome de l'X Fragile (FXS) (Oberlé et al., 1991), l'ataxie de Friedreich (FRDA) (Campuzano et al., 1996), le syndrome de retard mental non-symptomatique lié au chromosome X (FRAXE) (Knight et al., 1993), la dystonie-parkinsonisme liée à l'X (XDP) (Bragg et al., 2017) ou encore le syndrome Baratela-Scott (BSS) (LaCroix et al., 2019).

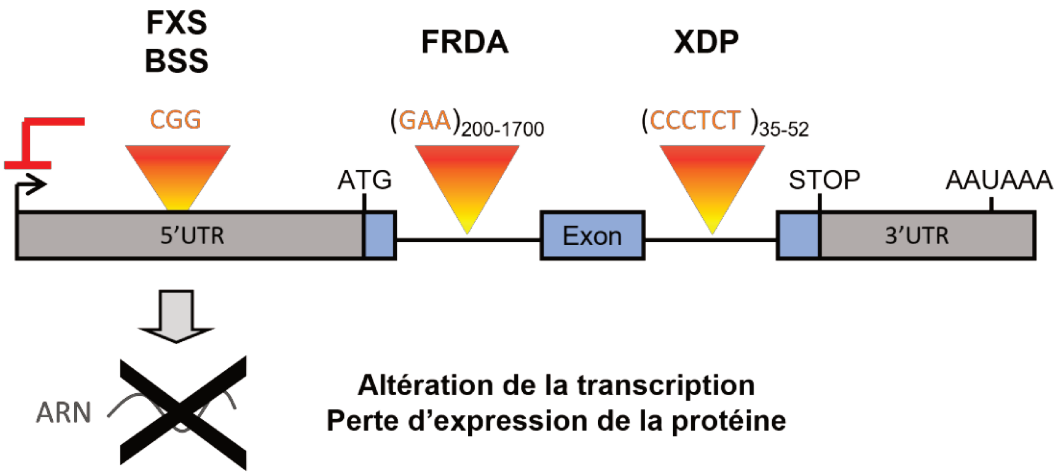


Figure 2. Mécanisme de toxicité au niveau ADN
Des expansions de répétitions, généralement proches du promoteur du gène, peuvent altérer la transcription de leur gène hôte par des modification épigénétiques.

1.2.1. Le syndrome de l'X fragile (FXS)

Le syndrome de l'X fragile (FXS) est la cause génétique la plus commune de retard mental héréditaire chez l'enfant avec une prévalence de 1/4000 pour les garçons et de 1/8000 pour les filles (Crawford et al., 2001). Ce syndrome se caractérise par un retard mental, plus prononcé chez les garçons (QI<50) que chez les filles (QI<80-90), par des troubles cognitifs comme des

troubles de l'attention ainsi que des signes d'autisme, une dysmorphie faciale ainsi qu'une macro-orchidie.

Dans la majeure partie des cas, FXS est causé par une expansion de nucléotides CGG dans la région 5'UTR du gène *FMR1* (Fragile X Mental Retardation 1) localisée sur le chromosome X (Oberlé et al., 1991 ; Verkerk et al., 1991 ; Fu et al., 1991). Chez les individus non atteints, le nombre de répétitions CGG est inférieur à 50 alors que chez les patients FXS, le nombre de répétitions CGG est supérieur à 200. De manière intéressante, une taille d'expansion de répétitions intermédiaire comprise entre 50 et 200 conduit à d'autres pathologies, le syndrome de tremblement et d'ataxie lié à l'X fragile (FXTAS) (Hagerman et al., 2001) et le syndrome de l'insuffisance ovarienne liée l'X (FXPOI) (Bretherick et al., 2005) qui seront présenté par la suite (cf Introduction 4.4.4). Lorsque les expansions CGG sont supérieures à 200 répétitions, elles conduisent à une hyper-méthylation des îlots CpG et à une diminution de l'acétylation des histones proches du promoteur du gène *FMR1*, entraînant une perte de l'expression de la protéine FMRP (De Boulle et al., 1993 ; Coffee et al., 1999 ; Tarleton et al., 2002). De plus, dans 2% des cas, FXS est dû à des mutations ponctuelles ou à des délétions dans la région codante du gène *FMR1*, et ceci sans expansion des nucléotides CGG (De Boulle et al., 1993 ; Tarleton et al., 2002).

La diminution d'expression de FMRP conduit à des dendrites et des synapses immatures et est considérée comme la cause de FXS (Hirst et al., 1995) cependant, le rôle exact de cette protéine reste débattu. La protéine FMRP possède 2 domaines KH et un domaine RGG liant les ARN au niveau de structures particulières dont les quadruplets de guanosine (quadruplexe-G) et les « kissing loop » (Darnell et al., 2001 ; Darnell et al., 2005 ; Didiot et al., 2008), suggérant un rôle de FMRP dans la régulation du métabolisme des ARN. Ainsi, cette protéine pourrait être impliquée dans la liaison et l'export des ARN vers le cytoplasme, mais aussi dans le transport d'ARN messagers (ARNm) du cytoplasme aux épines dendritiques et / ou régulant localement la synthèse de protéines en réponse à des signaux de neurotransmetteurs (Eberhart et al., 1996 ; Laggerbauer et al., 2001 ; Villace et al., 2004 ; Kim et al., 2009). Ces résultats pourraient ainsi expliquer les anomalies morphologiques des épines dendritiques chez les patients (Rudelli et al., 1985), qui sont également retrouvées dans les souris invalidées pour le gène *Fmr1* (Comery et al., 1997). Plus récemment, la protéine FMRP a été décrite comme activant la traduction d'un ARN en particulier, DGKκ. L'absence de FMRP conduit à une diminution de l'expression de la protéine DGKκ, qui convertit le diacylglycérol (DAG) en acide phosphatidique (PA), permettant ainsi de réguler ces voies de signalisation. De plus, la surexpression de DGKκ chez

des souris KO pour le gène *Fmr1* restaure l'altération des épines dendritiques (Tabet et al., 2016).

1.2.2. Le syndrome Baratela-Scott (BSS)

En 2012, un nouveau syndrome a été décrit par Baratela et al., où les patients présentent des anomalies et un retard du développement osseux, une petite taille et une dysmorphie de la face (Baratela et al., 2012). La découverte de différentes mutations perte de fonction sur les 2 allèles du gène *XYLT1* a permis d'identifier la cause d'apparition de la maladie, alors appelée la dysplasie de Dubesquois de type 2 (DBQD2), chez une grande partie des patients (Bui et al., 2014 ; Jamsheer et al., 2016 ; Guo et al., 2017). Néanmoins, une partie des patients ne présentaient une mutation perte de fonction du gène *XYLT1* que sur un seul allèle, ne permettant pas d'expliquer la gravité des symptômes. En 2019, il a été montré que ces patients atteints du syndrome Baratela-Scott (BSS) étaient porteurs d'une expansion de tri-nucléotides CGG située dans la région promotrice du gène *XYLT1* (LaCroix et al., 2019). Tout comme dans FXS, cette expansion conduit à une hyper-méthylation du promoteur du gène *XYLT1*, ce qui aboutit à une perte d'expression de la protéine Xylosyltransférase 1 (XYLT1), une enzyme impliquée dans la biosynthèse des protéoglycanes (Müller et al., 2005). Il est à noter que quelques rares patients sont porteurs de l'expansion de nucléotides CGG sur les 2 allèles.

1.2.3. L'ataxie de Friedreich (FRDA)

Les ataxies sont caractérisées par des troubles de la coordination des mouvements, de l'équilibre et de l'expression. De plus, les patients atteints de FRDA présentent des signes ostéoarticulaires et une cardiomyopathie. Cette pathologie, affectant 1 personne sur 50 000, est autosomale récessive ; le plus souvent, l'un des allèles présente une mutation perte de fonction et l'autre allèle est porteur d'une expansion de nucléotides GAA dans le premier intron du gène *FRDA*. Le nombre de répétitions GAA peut varier de 200 à 1700 et est inversement proportionnel à l'âge d'apparition des premiers symptômes (Dürr et al., 1996). Ces mutations conduisent à une forte diminution de l'expression de la Frataxine, codée par le gène *FXN* (Campuzano et al., 1996). Cette petite protéine mitochondriale joue un rôle dans l'assemblage et la structuration des centres Fer-Soufre retrouvés dans de nombreuses protéines (pour revue, Castro et al., 2019).

1.2.4. La dystonie parkinsonisme liée à l'X (XDP)

La dystonie parkinsonisme liée à l'X (XDP) est une maladie neurodégénérative rare caractérisée par une dystonie focale, c'est-à-dire des contractions involontaires et soutenues des muscles, associée à des tremblements et à de la rigidité musculaire. Cette pathologie est retrouvée principalement sur l'île de Panay aux Philippines où sa prévalence atteint 1 homme sur 20 000 (environ 500 cas). L'âge d'apparition de XDP est extrêmement variable, les premiers symptômes pouvant apparaître chez l'enfant comme chez une personne âgée (Makino et al., 2007).

Cette pathologie est causée par l'insertion d'un rétrotransposon composé d'une expansion de 35 à 52 nucléotides CCCTCT dans un intron du gène *TAF1*, le nombre de répétitions corrèle inversement avec l'âge d'apparition de la maladie (Bragg et al., 2017). Ce gène code pour la protéine TAF1 qui, tout comme la protéine TBP (« TATA-binding protein »), fait partie du complexe TFIID nécessaire pour le recrutement de l'ARN polymérase II sur le promoteur de gènes lors de l'initiation de la transcription (Bragg et al., 2017 ; Westenberger et al., 2019). Ce rétrotransposon est composé d'hexa-nucléotides CCCTCT et est donc composé de nucléotides GGGAGA sur le brin d'ADN complémentaire, cette séquence est prédite pour former des structures en quadruplexe-G qui interfèrent avec la transcription de *TAF1*, conduisant à une perte de fonction partielle de cette protéine uniquement dans les neurones, par un mécanisme encore mal compris (Makino et al., 2007). Il est intéressant de noter que des duplications et / ou des mutations faux-sens dans le gène *TAF1* provoquent le retard mental lié à l'X, syndromique 33 (MRXS33), confirmant ainsi que des altérations d'expression ou des mutations dans la protéine TAF1 conduisent à des pathologies du système nerveux central (O'Rawe et al., 2015).

1.3. Les maladies dues à une toxicité au niveau ARN

Les maladies dues à un gain de fonction toxique de l'ARN sont caractérisées par de longues expansions de plusieurs centaines, voire plusieurs milliers, de répétitions situées dans des régions non-codantes du génome, principalement dans les introns et les 3'UTR. Ces répétitions sont éloignées des promoteurs des gènes et ne conduisent pas à l'hyper-méthylation de celui-ci. Ces répétitions sont donc transcrites mais ne sont pas traduites et s'accumulent sous forme de foci d'ARN dans le noyau des cellules (Figure 3).

Les dystrophies myotoniques de type 1 et 2 (DM1 et DM2) sont les maladies les mieux caractérisées pour ce mécanisme de toxicité au niveau ARN. Ces pathologies sont caractérisées par un affaiblissement et une atrophie musculaires, une myotonie, c'est-à-dire une difficulté à relâcher un muscle contracté, des troubles cardiaques, une cataracte, des troubles cognitifs ainsi qu'une résistance à l'insuline.

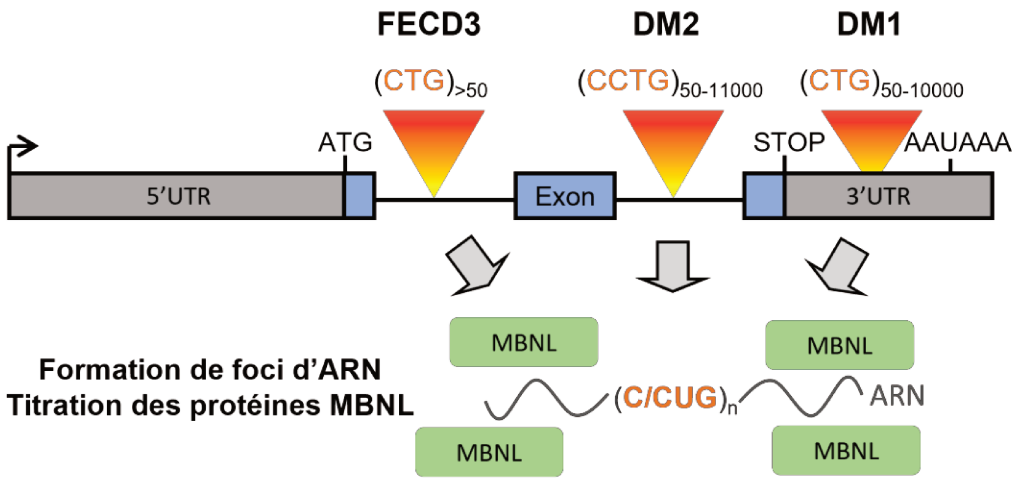


Figure 3. Mécanisme de toxicité au niveau ARN
Des expansions de répétitions, éloignées du promoteur et situées dans des régions non-codantes, sont transcrites en ARN qui forment des foci nucléaires et séquestrent les facteurs d'épissage MBNL.

La dystrophie myotonique de type 1 (DM1), ou dystrophie de Steinert, est classée sous 4 formes selon la sévérité et l'âge d'apparition des premiers symptômes : la forme congénitale, la forme infantile, la forme adulte et la forme tardive. Cette maladie est causée par une expansion de quelques centaines à plusieurs milliers de répétitions de tri-nucléotides CTG située dans la région 3'UTR du gène *DMPK*. Ces répétitions peuvent s'étendre d'une génération à l'autre, ce qui explique le mécanisme d'anticipation généralement observé dans les familles atteintes par la DM1 avec une diminution de l'âge d'apparition de la maladie et une augmentation de la

sévérité des symptômes. Ces répétitions, éloignées du promoteur, sont transcrites en un ARN contenant alors des milliers de répétitions CUG, formant des foci d'ARN dans le noyau des cellules. En 2000, il a été montré que les protéines « Muscle blind-like 1/2 » (MBNL1 et MBNL2) se lient spécifiquement aux répétitions CUG (Miller et al., 2001, Ho et al. 2004) et qu'elle colocalise avec les foci d'ARN composés de répétitions CUG (Dansithong et al., 2005). La séquestration de la protéine MBNL1 dans les foci d'ARN entraîne une perte de fonction de ce facteur d'épissage, conduisant à une altération de l'épissage de nombreux ARN tels que la troponine T2 de type cardiaque (TNNT2) et la sous-unité α du canal sodique voltage dépendant de type V (SCN5A) responsables des altérations cardiaques (Philips et al., 1998 ; Freyermuth et al., 2016), le récepteur à l'insuline (INSR) responsable de la résistance à l'insuline (Savkur et al., 2001), le canal chlore musculaire (CLCN1) responsable de la myotonie (Charlet et al., 2002 ; Mankodi et al., 2002), l'amphiphysine 2 (BIN1) responsable de la faiblesse musculaire (Fugier et al., 2011), ou encore la dystrophine (DMD) responsable d'anormalités structurales des fibres musculaires (Rau et al., 2015).

La dystrophie myotonique de type 2 (DM2) est causée par une expansion de 75 à 11 0000 répétitions de nucléotides CCTG dans le premier intron du gène *CNBP* (Liquori et al., 2001). Tout comme dans la DM1, ces répétitions CCTG sont transcrites en un ARN qui forment des foci et titrent le facteur d'épissage MBNL1, conduisant donc à des altérations moléculaires et des symptômes similaires. Toutefois, la DM2 est moins sévère que la DM1. Cette plus faible sévérité de la DM2 pourrait s'expliquer par le fait qu'une autre protéine, rbFOX1, est capable de lier des répétitions CCUG et non CUG, ce qui entraînerait une moindre séquestration de MBNL1 et donc, un plus faible niveau d'altération de l'épissage des ARN cibles de MBNL1 (Sellier et al., 2018).

Enfin, la dystrophie cornéenne endothéliale de Fuchs de type 3 (FECD3), est également due à une expansion de répétitions de tri-nucléotides CTG, cette fois dans une région intronique du gène *TCF4*, codant pour le Facteur de Transcription 4. Les FECD affectent plus de 4% de la population âgée de plus de 40 ans et conduisent à des altérations de la vision dues au dysfonctionnement des cellules endothéliales de la cornée (Sundin et al., 2006). Tout comme dans les DM1 et DM2, ces répétitions CTG sont transcrites en ARN formant des foci et séquestrant le facteur d'épissage MBNL1 conduisant à l'altération de l'épissage de nombreux ARNm de l'endothélium de la cornée (Du et al., 2015 ; Wieben et al., 2017). Il est intéressant de noter que 40% des patients atteints de la DM1 présentent des altérations oculaires similaires aux patients touchés par la FECD3 (Gattey et al., 2014 ; Mootha et al., 2017).

1.4. Les maladies dues à une toxicité au niveau protéique

Les maladies dues à un gain de fonction protéique sont causées par des expansions de nucléotides qui sont situées, dans la majeure partie des cas, dans la séquence codante du gène mais, sujet de cette thèse, elles peuvent également être localisées dans des régions dites « non-codantes » comme dans les régions 5'UTR ou introniques (Figure 4). Elles sont essentiellement provoquées par :

- Des expansions CAG dans la phase codante du gène traduites en **polyGlutamine** (polyQ) (HD, SCA1, SCA2, SCA3, SCA6, SCA7, SCA17, DRPLA, SBMA, etc) ;
- Des expansions GCN dans la phase codante du gène traduites en **polyAlanine** (polyA) (OPMD, NS-XLID, BPES, HFGS, CCHS, etc) ;
- Des expansions CGG situées dans le 5'UTR du gène et traduites en **polyGlycine** (polyG) (FXTAS).

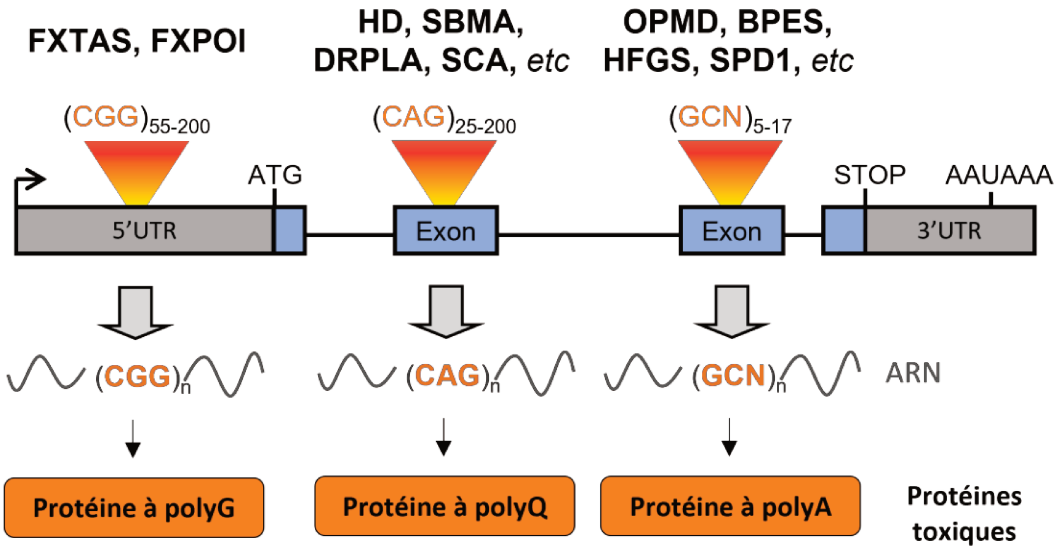


Figure 4. Toxicité au niveau protéique
Des expansions de répétitions, situées généralement dans la phase codante ou dans les régions 5'UTR des gènes, sont traduites en protéines toxiques.

La traduction de répétitions situées dans des régions dites « non-codantes » du génome faisant l'objet de ce travail de thèse, je vais m'attacher à présenter les généralités ainsi que les spécificités de l'initiation de la traduction chez les eucaryotes dans le sous chapitre suivant.

1.4.1. L'initiation de la traduction chez les eucaryotes

La traduction est le processus biologique permettant de décoder les ARNm en protéines et, où la séquence de nucléotides est convertie en une séquence d'acides aminés. Chaque acide aminé est généré à partir d'un triplet de nucléotides, appelé codon, la correspondance entre le codon et l'acide aminé se faisant selon le code génétique et par l'intervention d'un ARN de transfert (ARNt). Cette étape est cytoplasmique et fait intervenir le ribosome, un complexe ribonucléoprotéique, c'est-à-dire composé à la fois de protéines et d'ARN, composé de 2 sous-unités : la petite sous-unité 40S et la grande sous-unité 60S. La petite sous-unité 40S se lie à l'ARNm et recrute la grande sous-unité 60S qui contrôle l'arrivée des ARNt porteurs des acides aminés. La sélection de l'acide aminé se fait par appariement entre l'anti-codon de l'ARNt et le codon de l'ARNm. Lorsqu'un ARNt arrive au niveau du ribosome, il se trouve en attente au niveau du site A (pour Aminoacyl-ARNt) puis, il est transloqué dans le site P (pour Peptidyl-ARNt) où leur acide aminé est lié à la chaîne peptidique grâce à la grande sous-unité 60S et, il est enfin libéré par le site E (pour Exit) de la grande sous-unité du ribosome. Le ribosome avance alors de 5' vers 3' et par tranche de 3 nucléotides sur l'ARNm pour lire un nouveau codon et l'apparier avec l'anti-codon d'un nouvel ARNt, procédant ainsi à l'élongation de la traduction (Figure 5).

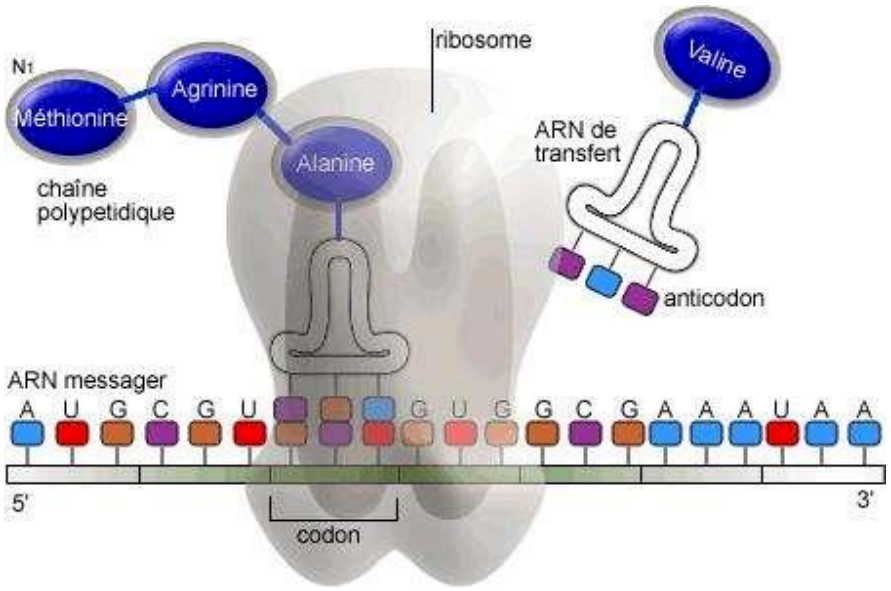


Figure 5. Traduction chez les eucaryotes
Le ribosome se déplace de 5' vers 3' sur l'ARNm où il sélectionne un ARNt grâce à l'appariement de son anti-codon avec le codon de l'ARNm. Le ribosome transfère l'acide aminé porté par l'ARNt vers la chaîne polypeptidique générant ainsi une nouvelle protéine. Image provenant de © Guillaume Bokiau, Wikimedia, Licence GFDL.

1.4.1.1. *Initiation de la traduction à un codon AUG*

L'initiation de la traduction fait intervenir des facteurs d'initiation eucaryotes (eIF) et comprend 2 étapes : la formation du complexe d'initiation 48S du ribosome puis, l'assemblage des sous-unités 48S et 60S du ribosome.

Tout d'abord, le complexe ternaire (TC) se forme : eIF2-GTP se lie à l'ARNt initiateur chargé de l'acide aminé méthionine (Met-ARNt^{Met}). Puis, le TC interagit avec la sous-unité 40S du ribosome associée à des facteurs d'initiation eucaryotes dont eIF1, eIF1A, eIF5 et le complexe eIF3 pour ainsi former le complexe de pré-initiation 43S (PIC-43S).

En parallèle, l'ARNm est activé et circularisé via son interaction avec eIF4B, une protéines de liaison à l'ARN, et le complexe eIF4F qui est composé d'eIF4A, une hélicase DEAD-box, d'eIF4E, une protéine liant la coiffe 7-méthyl-guanosine (m^{7}G) située à l'extrémité 5' de l'ARNm et eIF4G, une protéine qui lie la protéine de liaison au polyA (PABP) située à l'extrémité 3' de l'ARNm. Cette étape nécessite de l'énergie par l'hydrolyse d'ATP afin d'aplanir les structures secondaires de l'ARNm par l'hélicase eIF4A. Le complexe eIF4F recrute alors le PIC-43S au niveau de l'ARNm pour former le PIC-48S (Figure 6).

Dépendamment de l'hydrolyse d'ATP, le PIC-48S scanne alors l'ARNm de 5' vers 3' à la recherche d'un codon d'initiation, le plus souvent un codon AUG, situé dans une séquence Kozak. Chez les mammifères, la séquence Kozak consensus est C(**A/G**)CCAUG**G** (Kozak et al., 1989), où la purine en position -3 et la guanosine en position +4 (en gras) sont particulièrement importantes pour la stabilisation de l'interaction du PIC-48S avec l'ARNm (Peabody, 1987 ; Pisarev et al., 2006). Une fois l'anticodon du Met-ARNt^{Met} porté par eIF2-GTP apparié au codon initiateur sélectionné au niveau du site P du PIC-48S, eIF5 facilite l'hydrolyse du GTP de eIF2 en GDP + un phosphate qui est relâché. S'en suit le départ des facteurs eIF1, eIF1A, eIF2-GDP, eIF3 et eIF5, ce qui permet le recrutement de la sous-unité 60S du ribosome, formant ainsi le ribosome 80S nécessaire à l'élongation (Figure 6 ; pour revues, Jackson et al., 2010 ; Kearse et Wilusz, 2017).

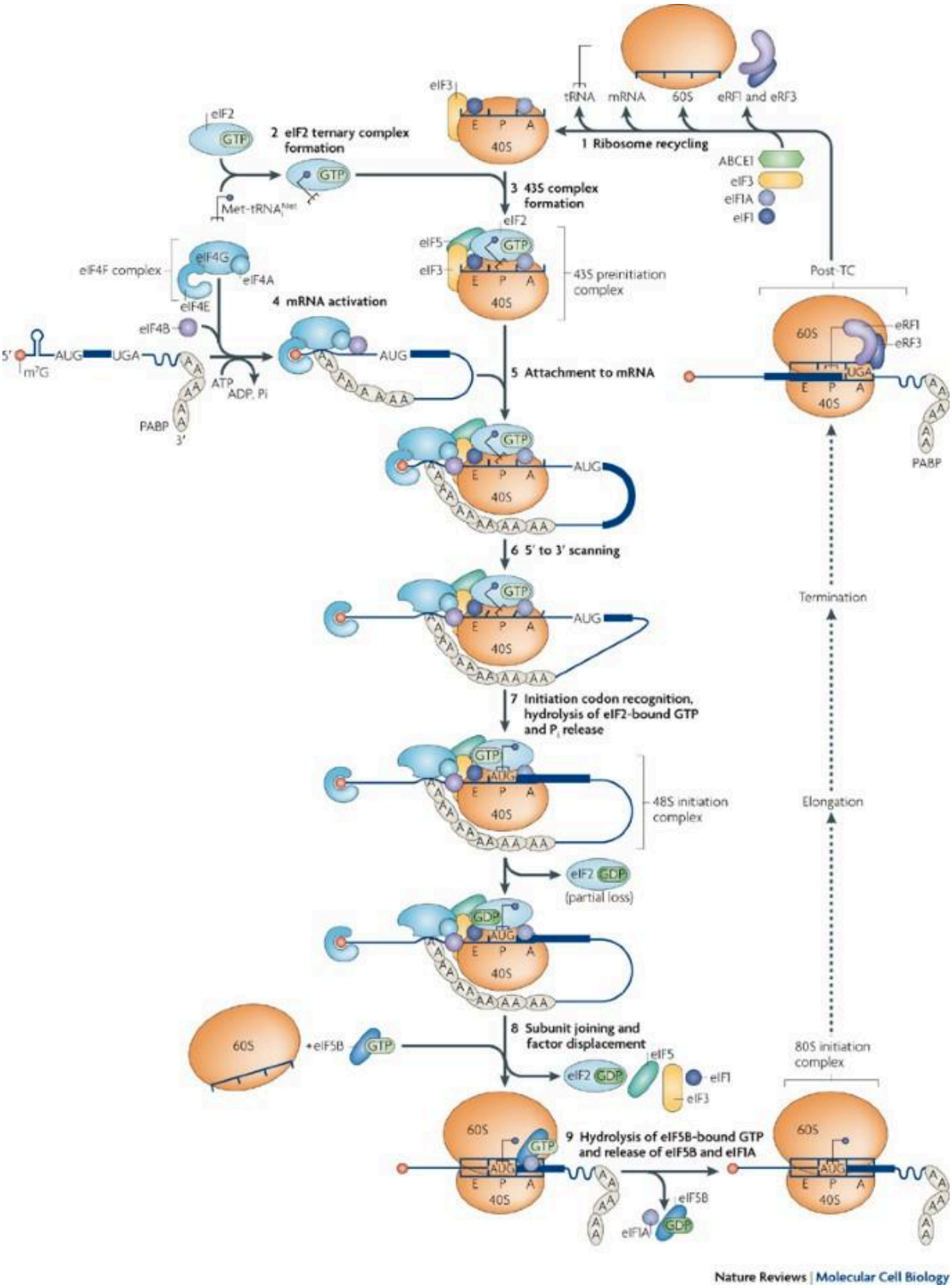


Figure 6. Initiation de la traduction chez les eucaryotes
Etapes et facteurs intervenant dans l'initiation de la traduction. Image de Jackson et al., 2010.

1.4.1.2. *Initiation de la traduction à des codons ressemblant à des AUG*

Depuis les années 1980, il est connu que l'initiation de la traduction ne se fait pas uniquement à des codons AUG. En effet, des codons ne différant que d'un nucléotide par rapport au codon AUG comme des codons CUG, GUG, UUG, ACG, etc. sont capables d'initier, par mésappariement avec l'ARNt initiateur Met-ARNt^{Met}, la traduction mais avec une efficacité moindre de 1 à 20% par rapport à un codon canonique AUG (Tableau 2 ; Peabody, 1987 ; Clements et al, 1988 ; Peabody, 1989 ; Ivanov et al., 2010).

Biological source	Rabbit reticulocyte lysate (Peabody 1989)	Rabbit reticulocyte lysate (Wei et al. 2013)	Wheat germ extract (Peabody 1989)	*Saccharomyces cerevisiae* (Clements et al. 1988)	*Neurospora crassa* (Wei et al. 2013)	HEK293T cells (Ivanov et al. 2010)
Reporter used	Dihydrofolate reductase	Firefly luciferase	Dihydrofolate reductase	β-Galactosidase	Firefly luciferase	Firefly luciferase
AUG	100	100	100	100	100	100
CUG	82	18	36	0.22	10	19
GUG	36	11	8	0.5	6	9
UUG	39	8	10	0.37	2.5	2
ACG	84	5	45	0.39	3.5	7
AUC	47	3	17	0.05	<0.5	2
AUU	67	6	14	0.38	1	3
AAG	14	0.01	3	0.02	<0.5	<0.5
AUA	59	5	30	0.29	3	3
AGG	17	0.02	3	0.04	<0.5	<0.5

Tableau 2. Efficacité de l'initiation de la traduction selon le codon initiateur
Les codons non-AUG permettent d'initier la traduction mais à plus faible niveau que le codon AUG. Cette efficacité varie selon le modèle cellulaire. Tableau de Kearse et Wilusz, 2017.

Ce mésappariement de l'anticodon de l'ARNt initiateur avec un codon ressemblant à un AUG sur l'ARNm dépend notamment de la séquence Kozak entourant ce codon. Si cette séquence est défavorable, par l'absence d'une guanosine en +4 par exemple, le taux d'initiation sera plus faible qu'avec la séquence Kozak consensus. De plus, l'utilisation de codons non-canoniques pour l'initiation de la traduction est un mécanisme finement régulé notamment par eIF1, eIF2 et eIF5. En effet, des mutations dans eIF1, eIF1A et / ou eIF5, conduisant à l'hydrolyse prématurée d'eIF2-GTP, augmentent l'initiation de la traduction à des codons non-AUG (Huang et al., 1997). De plus, la surexpression d'eIF5 et eIF5B augmente l'utilisation de codons non-AUG alors que la surexpression d'eIF1 et eIF1A diminue l'utilisation de codons non-AUG (Barth-Baus et al., 2013). Enfin, la structure des ARNm influe également sur l'utilisation des codons non-AUG. Par exemple, la présence de structures en quadruplexe-G ou en épingle à cheveux situées en aval des codons non-AUG ralentit la vitesse de scan du ribosome, augmentant donc le temps de pause sur ces codons favorisant ainsi leur utilisation pour l'initiation de la traduction (Kozak, 1989 ; Kearse et al., 2016).

Initialement considéré comme un mécanisme rare et limité à quelques protéines, par exemple l'initiation de la traduction de la protéine TEAD1 à un codon AUU (Xiao et al., 1991), des expériences récentes de « ribosome profiling » à l'échelle du génome ont mis en évidence que l'initiation de la traduction des uORF pour « upstream Open Reading Frame », se fait majoritairement à des codons non-AUG : **CUG**, **GUG**, **UUG** et **ACG** étant les plus fréquents (Figure 7 ; Ingolia et al., 2011).

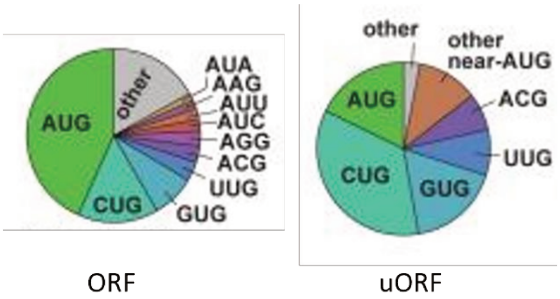


Figure 7. Utilisation des codons pour l'initiation de la traduction dans les ORF et uORF
L'initiation de la traduction se fait majoritairement à un codon AUG pour les ORF principales et à un codon CUG pour les uORF. Image de Ingolia et al. 2011.

Les uORF sont des ORF de petites tailles, présentes dans 30 à 50% des gènes de mammifères et situées en amont des ORF principales et qui sont donc localisées dans la région 5'UTR des gènes (Ingolia et al., 2011). Ces uORF peuvent avoir plusieurs rôles. Lorsque l'uORF se trouve dans le même cadre de lecture ouvert que l'ORF principale, l'initiation de la traduction au niveau de l'uORF produit une protéine plus longue avec une séquence N-terminale différente pouvant modifier la fonction de la protéine. Par exemple, la protéine FGF2 est une protéine cytoplasmique impliquée dans la prolifération cellulaire et la différenciation. Au moins 4 uORF initiant à un codon CTG ont été décrites dans ce gène (Renko et al., 1990 ; Arnaud et al., 1999). Il a été montré que les cellules cancéreuses favorisent l'utilisation de ces codons CTG, générant alors la production d'isoformes de FGF2 plus longues, régulant alors positivement les propriétés cancéreuses de ces cellules (Vagner et al. 1996).

Les uORF peuvent être chevauchantes ou séparées de l'ORF principale mais dans un cadre de lecture ouvert qui diffère de l'ORF principale. Ce type d'uORF est connu pour réguler, souvent de manière négative, la traduction en aval de l'ORF principale. En effet, le ribosome scannant l'ARNm peut rencontrer un codon AUG ou non-AUG initiant la traduction d'une uORF. Arrivant à un codon stop UAG, UGA ou UAA, le ribosome se décroche de l'ARNm et, si cette uORF est chevauchante de l'ORF principale ou, si la fin de cette uORF est proche du codon AUG de l'ORF principale, il ne peut réinitier, aboutissant donc à l'absence de la traduction de

l'ORF principale. Il est à noter que si l'ORF principale est éloignée (> 100 nucléotides) de la fin de l'uORF, le ribosome peut réinitier la traduction de cette ORF principale.

Toutefois, des cas de régulation positive en condition de stress existent également. Ainsi, une des régulations positives de la traduction par l'uORF sur l'ORF principale les mieux décrites chez les levures et les mammifères concernent les gènes *GCN4/ATF4* et *ATF5*. Ces gènes codent pour les protéines ATF4 et ATF5 impliquées dans l'adaptation et la survie cellulaire qui, après un stress cellulaire, sont 5 à 10 fois plus exprimées. La régulation de la traduction de ces protéines fait intervenir au moins 2 uORF : la première se trouve entièrement en amont du gène et, la seconde est chevauchante à l'ORF principale (Figure 8).

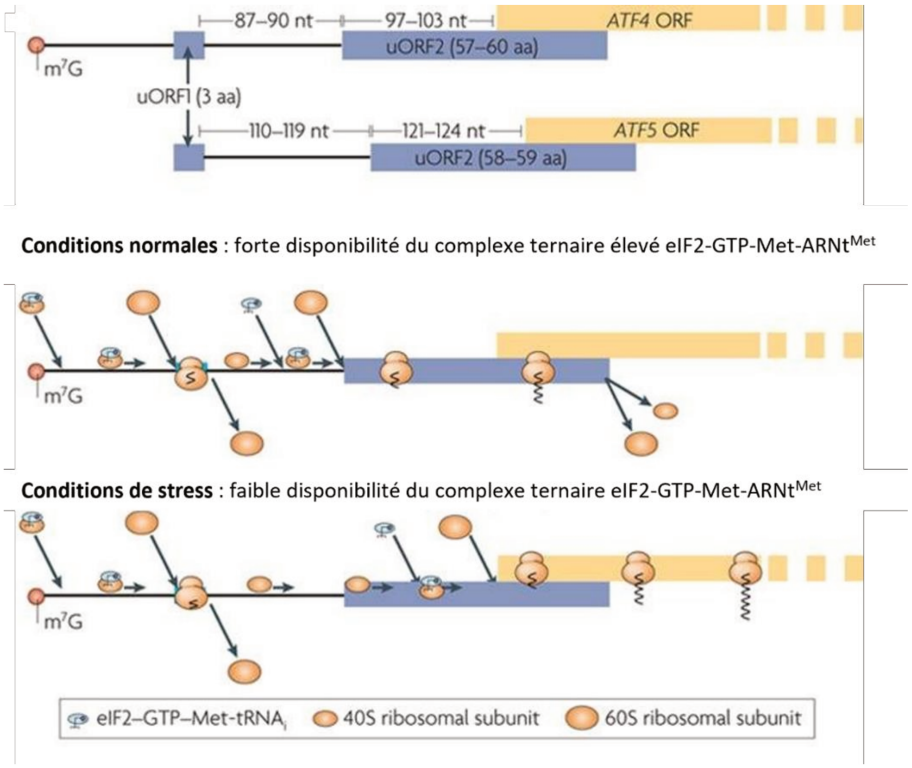


Figure 8. Régulation positive en condition de stress des uORF sur les gènes *ATF4* et *ATF5*
Les gènes *ATF4* et *ATF5* possèdent 2 uORF qui régulent la traduction de l'ORF principale située en aval en fonction des conditions de stress. Image de Jackson et al., 2010.

En condition « normale », le ribosome initie à la première uORF puis, initie à nouveau à la seconde uORF. Cette seconde uORF chevauche l'ORF principale, ce qui ne permet donc pas au ribosome d'initier la traduction des protéines ATF (Figure 8). Lors d'un stress cellulaire, comme une privation en acides aminés, la protéine kinase PERK est activée et phosphoryle alors eIF2 en sérine 51. Cette phosphorylation induit une augmentation de l'affinité d'eIF2 pour son facteur d'échange GDP/GTP (GEF) eIF2B, ce qui séquestre alors eIF2B qui ne peut plus

assurer son activité GEF auprès d'eIF2, diminuant donc la quantité de eIF2-GTP disponible. Comme précédemment, le ribosome initie à la première uORF mais ne peut pas réinitier à la seconde uORF car la petite sous-unité ribosomique n'a pas pu se lier à temps à un nouvel eIF2-GTP. Il en résulte que le ribosome continue à scanner l'ARNm et s'arrête au niveau de l'ATG de l'ORF principale, permettant alors la traduction des protéines ATF4 et ATF5 (Figure 8 ; Vattem et Wek, 2004 ; Zhou et al., 2008).

1.4.1.3. *Initiation de la traduction au niveau de séquences IRES*

Les séquences IRES, pour « Internal Ribosome Entry Site », sont des structures formées par les ARN permettant le recrutement du ribosome directement à l'intérieur des ARN qui ne sont pas coiffés en 5', et donc indépendamment du scan du ribosome. Ce type d'initiation de la traduction, majoritairement utilisé par les virus à ARN monocaténaire, est basé sur des interactions non-canoniques entre l'ARN viral et le ribosome et / ou les facteurs eIF (pour revue, Jackson et al., 2010). Selon les structures de l'ARN viral, on dénombre 4 types d'IRES principales (Figure 9) :

- Types 1 et 2 où l'ARN recrute directement les facteurs eIF4A et eIF4G (mais pas eIF4E), conduisant ainsi au recrutement du complexe PIC-43S pour former le complexe PIC-48S. Ce type d'initiation de la traduction est notamment utilisé par les picornavirus comme le poliovirus causant la poliomyélite ou le virus de l'encéphalomyocardite (EMCV).
- Type 3 où l'ARN interagit directement avec le facteur eIF3 et le sous-unité 40S du ribosome du complexe PIC-43S. Cette stratégie est utilisée par les virus de type VHC responsables de l'hépatite C.
- Type 4 où l'ARN recrute directement la sous-unité 40S du ribosome. Par sa structure, l'ARN mime la présence de l'ARNt initiateur Met-ARNt^{Met} au niveau du site P du ribosome, lançant ainsi directement l'élongation de la traduction. Ce type d'IRES est utilisé par les dicistrovirus comme le virus de la paralysie du criquet (Figure 9).

Outre leur présence chez certains virus, des éléments IRES ont aussi été identifiés dans certains ARNm eucaryotiques. Par exemple, lorsque des cellules HeLa sont infectées par un picornavirus, elles ne sont plus capables d'utiliser le mécanisme d'initiation de la traduction traditionnelle. Or, il a été montré que la protéine liant les chaines lourdes de

l'immunoglobuline (BiP) était toujours traduite lors d'une infection virale, suggérant que certains ARNm cellulaires peuvent être traduits via un mécanisme de type IRES (Macejak et Sarnow, 1991). Par la suite, plusieurs ARNm ont été décrits comme comportant une séquence IRES putative, notamment des gènes impliqués dans le développement, la différenciation cellulaire, le cycle cellulaire ou encore l'apoptose (pour revue, Komar et Hatzoglou, 2005).

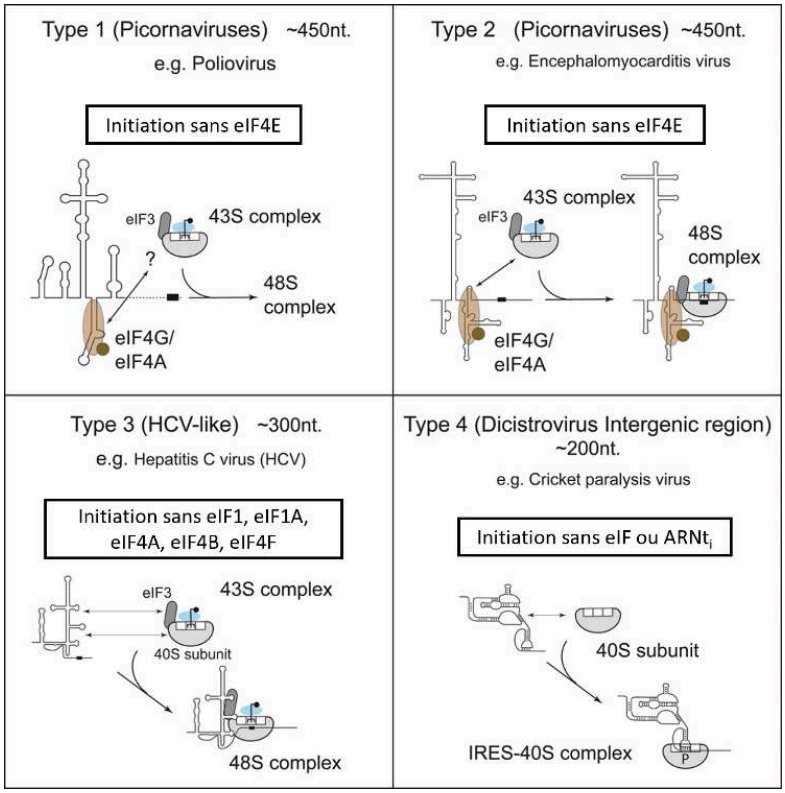


Figure 9. Les 4 principales types d'IRES
Suivant le virus à ARN monocaténaire considéré, les séquences IRES utilisées pour recruter le ribosome diffèrent. Image de Jackson et al., 2010.

1.4.1.4. *Traduction « RAN » des répétitions non associées à un codon AUG*

La traduction « RAN » pour « Repeat Associated Non-AUG translation » est un mécanisme qui a été décrit pour la première fois en 2011 lors de l'étude de la traduction des expansions de répétitions de nucléotides dans la dystrophie de type 1 (DM1) et l'ataxie spinocérébelleuse de type 8 (SCA8) (Zu et al., 2011). Concernant SCA8, il a été montré dans un travail antérieur qu'une expansion de répétitions CAG dans le gène *ATXN8* était traduite en une protéine polyGlutamine (polyQ) grâce à un codon ATG situé en amont de ces répétitions (Moseley et al., 2006). Néanmoins, Zu et al., en ôtant ce codon ATG, a pu mettre en évidence que ces

répétitions CAG étaient également traduites en protéine polyAlanine (polyA) et polySérine (polyS), et ceci sans codon d'initiation de la traduction apparent. De même, dans la DM1, ils ont pu montrer que l'ARN antisens de DMPK, composé de répétitions CAG, était traduit en protéine polyQ et ce, sans codon d'initiation de la traduction. Ces résultats suggèrent que les expansions de répétitions peuvent être traduites dans les trois cadres de lecture, générant plusieurs protéines anormales et potentiellement toxiques. Les auteurs ont alors proposé un nouveau mécanisme d'initiation de la traduction appelé « RAN » où, les expansions de répétitions au niveau ARN seraient capables de former des structures de type épingle à cheveux et ainsi de recruter potentiellement certains facteurs eIF et / ou directement le ribosome, tout comme une séquence IRES virales (Figure 10 ; Zu et al., 2011).

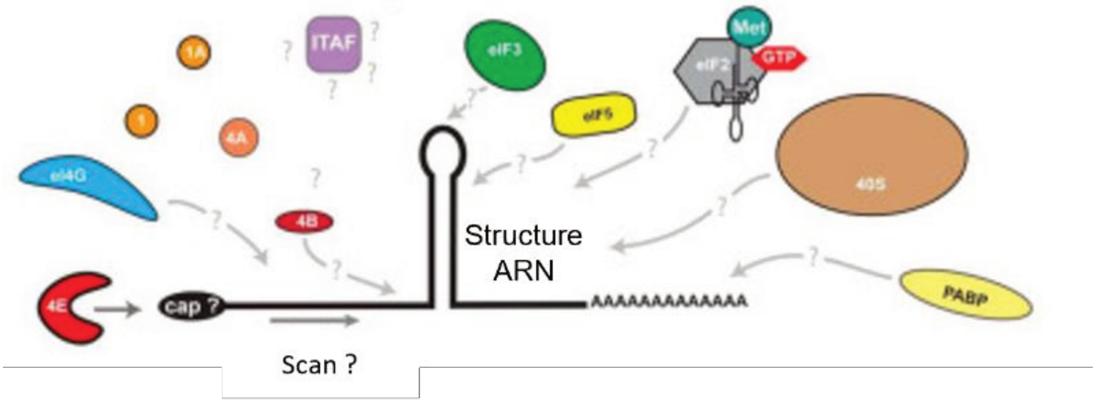


Figure 10. Mécanisme de traduction « RAN »
Les expansions de répétitions au niveau ARN formeraient des structures secondaires capables de recruter certains facteurs impliqués dans l'initiation de la traduction ou le ribosome. Image de Cleary et al., 2018

Depuis cette découverte, la traduction « RAN » a été décrite dans de nombreuses maladies comme dans la maladie de Huntington (Bañez-Coronel et al., 2015), la dystrophie myotonique de type 2 (Zu et al., 2017), la dystrophie cornéenne endothéliale de Fuchs (Soragni et al., 2018), le syndrome de tremblement et d'ataxie lié à l'X fragile (Todd et al., 2013 ; Krans et al., 2016) ou la sclérose latérale amyotrophique associée à la démence fronto-temporale due à une expansion de répétitions dans le gène *C9ORF72* (Ash et al., 2013 ; Gendron et al., 2013 ; Mackenzie et al., 2013 ; Mori et al., 2013a ; Mori et al., 2013b ; Zu et al., 2013).

Suite à cette présentation des différents mécanismes de traduction existants, je vais maintenant m'attacher à décrire quelques exemples de maladies causées par une toxicité au niveau protéique.

1.4.2. Les maladies à polyGlutamine (polyQ)

Actuellement, une dizaine de maladies autosomales dominantes, exceptée l'atrophie musculaire spinale et bulbaire (SBMA) qui est liée à l'X, est causée par une expansion de tri-nucléotides CAG traduites en polyQ a été répertoriée : l'atrophie musculaire spinale et bulbaire (SBMA), la maladie de Huntington (HD), la maladie de Huntington de type 2 (HDL2), l'atrophie dentato-rubro-pallido-luysienne (DRPLA) et les ataxies spinocérébelleuses (SCA) de type 1, 2, 3, 6, 7, 8 et 17 dues respectivement à des mutations dans la phase codante des gènes *AR*, *HTT*, *JPH3AS*, *ATN1*, *ATXN1*, *ATXN2*, *ATXN3*, *CACNA1A*, *ATXN7*, *ATXN8* et *TBP*. Le seuil de pathogénicité de ces répétitions CAG varie de 21 à 55 selon la pathologie considérée, et n'excèdent pas 200 répétitions. Bien que ces gènes soient exprimés de manière ubiquitaire, des populations neuronales spécifiques semblent plus sensibles que les autres cellules de l'organisme aux protéines à polyQ, expliquant que ces pathologies soient essentiellement neurodégénératives. Toutefois, il est intéressant de noter que selon leur gène hôte, ces répétitions sont toxiques pour des populations neuronales spécifiques. En effet, les patients touchés par des SCA présentent une atrophie du cervelet (Figure 11A, 11B et 11C) et les patients atteints de HD et HDL2 développent une atrophie du striatum conduisant à une hypertrophie des ventricules (Figure 11D).

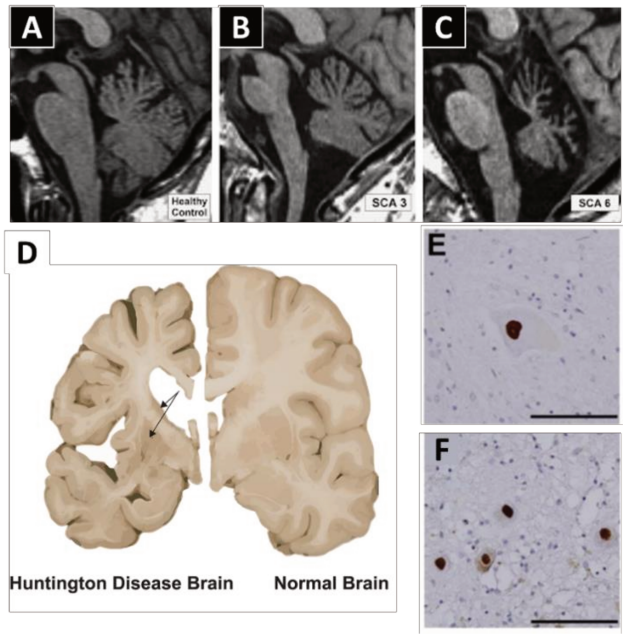


Figure 11. Histopathologie des patients touchés par les maladies à polyQ
(**A**, **B**, **C**) IRM d'un individu (**A**) contrôle, (**B**) atteint de SCA3 et (**C**) de SCA6. Images de Eichler et al., 2011. (**D**) Atrophie du cerveau d'un patient atteint de HD en comparaison à un individu contrôle. Image de « Courtesy of Dr Jean-Paul Vonsattel, Columbia University ». (**E**, **F**) IHC révélant la presence d'agrégats de polyQ dans (**E**) la corne antérieure de la moelle épinière d'un patient SBMA et (**F**) les noyaux dentelés d'un patient DRPLA. Images de Banno et al., 2009.

D'un point de vue génétique, plus l'expansion de répétitions CAG est grande, plus la maladie est sévère et plus les premiers symptômes de ces pathologies se développent tôt chez les patients. De plus, on observe un mécanisme d'anticipation où les enfants d'un individu malade présentent généralement un nombre de répétitions plus élevé et, par conséquent, développent des manifestations cliniques plus précocement (pour revue, Lieberman et al., 2019). D'un point de vue histopathologique, les patients porteurs de ces répétitions CAG dans la phase codante des gènes présentent des inclusions typiques de protéines composées de polyQ (Figure 11E et 11F). Bien qu'aujourd'hui le gain de fonction toxique des protéines composées de polyQ soit admis, les mécanismes moléculaires entraînant la mort neuronale sont complexes.

1.4.2.1. *Les protéines à polyQ sont-elles toxiques d'elles-mêmes ?*

Le développement des premiers modèles murins exprimant des polyQ ont permis de mettre en évidence la toxicité de ces protéines. En effet, des souris exprimant l'exon 1 de la Huntingtin (HTT) avec une expansion de polyQ est suffisante pour provoquer une neuropathologie regroupant une partie des symptômes de la maladie de Huntington (HD), comme des mouvements stéréotypiques involontaires et des tremblements (Mangiarini et al., 1996). De manière similaire, l'expression chez la souris d'une expansion de polyQ, dérivée du gène muté de l'ataxie spinocérebelleuse de type 3 (SCA3), uniquement dans les cellules de Purkinje provoque une ataxie chez ces souris (Ikeda et al., 1996). De façon intéressante, l'insertion d'une expansion de 146 tri-nucléotides CAG dans le gène *HPRT*, non-impliqué dans ces différentes pathologies, conduit aussi à un phénotype neurodégénératif sévère chez la souris, confirmant une toxicité intrinsèque des polyQ (Ordway et al., 1997).

Depuis, de nombreux autres modèles animaux ont été développé ce qui a permis de mettre en évidence, par exemple, une plus forte toxicité des polyQ dans des protéines tronquées par rapport aux protéines entières. En effet, l'insertion de répétitions CAG par Knock In (KI) dans le gène de *HTT* chez la souris conduit à de faibles altérations comportementales (Wheeler et al., 2000) en comparaison à la seule expression de l'exon 1 contenant les polyQ (Mangiarini et al., 1996). Des résultats similaires ont été rapporté dans l'étude de modèles murins modélisant des ataxies spinocérébelleuses où une forme tronquée de l'ATXN3-polyQ développe rapidement une sévère atrophie du cervelet (Ikeda et al., 1996), alors que des souris exprimant la protéine

entière ATXN3-polyQ ne développe des altérations neurologiques qu'à partir de 10 mois (Switonski et al., 2015).

De plus, la mutation de certains acides aminés proches des répétitions de polyGlutamine semble moduler leur toxicité. Par exemple, la mutation de résidus sérines pour mimer une phosphorylation constitutive (S -> D), localisés proches des polyQ dans la HTT permet d'améliorer nettement le phénotype des souris HD (Gu et al., 2009). De manière similaire, la mutation S776A de l'ATXN1 dans un modèle murin de SCA1 permet une moindre agrégation de l'ATXN1 et une meilleure survie des souris (Emamian et al., 2003). Enfin, dans SCA8, les répétitions CAG peuvent être interrompues par des codons CGG codant pour une arginine et, plus le nombre de résidus arginine est élevé, plus l'âge d'apparition de la pathologie est faible, suggérant que les acides aminés arginines augmentent la toxicité des polyQ (Perez et al., 2021). Ces données ont ainsi permis de mettre en évidence une toxicité intrinsèque des polyQ, mais dont la toxicité est modulée par la taille et / ou la composition en acides aminés des séquences entourant les polyQ.

1.4.2.2. *Toxicité nucléaire ou cytoplasmique des protéines à polyQ ?*

Une similarité entre les patients atteints de HD, DRPLA, SBMA et plusieurs SCA est la présence caractéristique d'inclusions intranucléaires formées par les protéines composées de polyGlutamine (Trottier et al., 1995). Il est toutefois à noter que des agrégats cytoplasmiques dans les neurones ont également été observés (DiFiglia et al., 1997). Le fait que ces agrégats soient majoritairement nucléaires est un résultat surprenant dans le cas de HD, car la protéine hôte des polyQ, la Huntingtin, est une protéine cytoplasmique. Ces données ont poussé les scientifiques à investiguer la contribution d'une toxicité nucléaire vs cytoplasmique de ces protéines.

La SBMA, une maladie caractérisée par une atrophie et une faiblesse musculaire, est causée par une expansion de répétitions CAG dans le gène *AR* codant pour le récepteur aux androgènes (La Spada et al, 1991). Les androgènes sont des hormones sexuelles masculines, dont la principale chez l'Homme est la testostérone, produite dans les testicules par les cellules de Leydig. La testostérone, lorsqu'elle entre dans une cellule cible, se lie au récepteur aux androgènes (RA) situé dans le cytoplasme, conduisant à la translocation de ce facteur de transcription dans le noyau des cellules, où il va se lier à l'ADN et stimuler la transcription de

ses gènes cibles (Figure 12 ; pour revue, Centenera et al., 2008). Comme chez l'Homme, uniquement les souris mâles porteuses de l'expansion de CAG dans le gène *AR* développent la pathologie. De manière intéressante, l'administration de testostérone chez des souris femelles, induisant la localisation nucléaire du RA-polyQ, conduit au développement de la maladie ; alors que la castration des mâles en empêche le développement, puisque le récepteur aux androgènes avec une expansion de polyQ reste cytoplasmique (Katsuno et al. 2002).

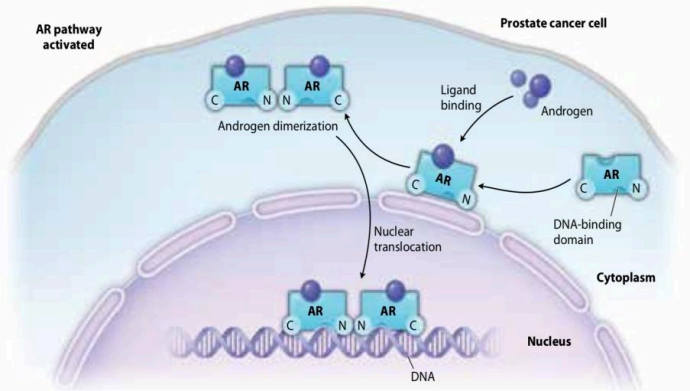


Figure 12. Fonctionnement du récepteur aux androgènes (RA)
Lorsque la testostérone se lie au RA, il est transloqué dans le noyau où il active la transcription de ses gènes cibles. Image de prost8blog.com.

De manière similaire, la protéine ATXN1, mutée dans SCA1, est un facteur de transcription nucléaire dont la localisation dépend d'une séquence de localisation nucléaire (NLS). De façon intéressante, la génération de souris transgéniques exprimant l'ATXN1-polyQ dépourvue de cette séquence NLS montre que ces animaux ne développent pas d'ataxie, suggérant ainsi que la localisation nucléaire est nécessaire pour la pathogénicité des protéines polyQ (Klement et al., 1998). Enfin, il a été montré dans HD que la délétion des 17 premiers acides aminés de la protéine HTT-polyQ module sa capacité d'agrégation et sa localisation. *In vitro*, la délétion de ces 17 acides aminés conduit à l'export du noyau vers le cytoplasmique de la HTT-polyQ et à une moindre toxicité (Zheng et al., 2013). Cependant, *in vivo*, la délétion de cette région accélère la migration de la HTT-polyQ dans le noyau et, les souris exprimant ce mutant présentent un phénotype plus sévère que les souris exprimant la protéine HTT-polyQ entière (Gu et al., 2015). Ces résultats indiquent que dans le cas des maladies SBMA, HD, SCA1, etc. les protéines à polyQ sont toxiques principalement lorsqu'elles sont nucléaires.

Néanmoins, dans les ataxies spinocérébelleuses de type 2 et 6 (SCA2 et SCA6), les inclusions de protéines à polyQ sont principalement cytoplasmiques chez les patients (Ishikawa et al., 1999 ; Huynh et al., 2000), suggérant qu'une localisation nucléaire des polyQ nucléaire n'est pas systématiquement nécessaire à leur toxicité.

1.4.2.3. *L'agrégation des protéines à polyQ est-elle nécessaire pour leur toxicité ?*

Comme nous venons de le voir, toutes les maladies à polyQ sont caractérisées par la présence d'agrégats protéiques, nucléaires et / ou cytoplasmiques. Toutefois, ces inclusions ne sont pas systématiquement retrouvées dans les tissus atteints chez les patients. En effet, bien que dans HD, les inclusions protéiques soient majoritairement localisées dans les zones atteintes ; ce n'est pas le cas par exemple dans l'ataxie spinocérébelleuse de type 7 (SCA7), où les inclusions sont retrouvées dans de nombreuses régions du cerveau et non uniquement dans le cervelet (Holmberger et al., 1998). Ces observations soulèvent donc l'hypothèse que ces agrégats ne sont pas forcément responsables de la neurodégénérescence.

Grâce à la génération de cultures primaires de neurones de striatum de rats pour l'étude de HD, il a été démontré que l'expression de la HTT-polyQ conduisait bien à une mort neuronale, mais qui ne corrèle pas avec la dynamique d'agrégation de cette protéine. En effet, les neurones ne développant pas d'agrégats meurent prématurément alors que les neurones présentant des agrégats voient leur temps de survie augmenté (Arrasate et al., 2004). De plus, l'utilisation d'un mutant de la HTT-polyQ ne s'agrégeant pas conduit à une mort des cultures neuronales plus importante, suggérant un effet protecteur de l'agrégation (Saudou et al., 1998).

De même, dans SCA1, un modèle murin exprimant un mutant de la protéine ATXN1-polyQ incapable de s'agréger (par la suppression de sa région d'auto-association), révèle une toxicité aussi élevée de ce mutant que celle observée avec l'expression de la protéine ATXN1-polyQ WT, confirmant que le développement de la pathologie n'est pas dépendant de l'agrégation des protéines à polyQ (Klement et al., 1998). De plus, l'expression de l'ATXN1-polyQ chez un modèle animal déficient pour l'ubiquitine ligase E6-AP (Ube3A), une protéine impliquée dans la dégradation des protéines par le protéasome, permet la solubilisation de l'ATXN1-polyQ. Néanmoins, ces souris présentent une dégénération exacerbée des cellules de Purkinje (Cummings et al., 1999).

En conclusion, ces expériences démontrent une toxicité particulièrement forte des protéines à polyQ sous leur forme soluble et suggèrent un effet protecteur de l'agrégation.

1.4.2.4. *Gain de fonction et / ou perte de fonction ?*

Outre SCA8 et HDL2, les expansions de répétitions de nucléotides CAG sont localisées dans la phase codante de protéines bien décrites, un mécanisme de perte de fonction ou d'altération de la fonction normale de ces protéines n'est donc pas à exclure. Ainsi, la protéine mutée dans HD, la Huntingtin (HTT), joue un rôle dans le trafic vésiculaire et son KO est léthal à l'état embryonnaire chez la souris (Duyao et al., 1995 ; Nasir et al., 1995 ; Zeitlin et al., 1995). Récemment, il a été montré que la HTT était impliquée dans le neurodéveloppement humain et que, la présence de la HTT-polyQ altérait le développement cortical embryonnaire, apportant ainsi une composante neurodéveloppementale à la maladie de Huntington (Barnat et al., 2020).

En ce qui concerne SCA1, il est décrit que la protéine ATXN1, native ou avec une expansion de polyQ, forme un complexe avec la protéine Capicua (CIC). Ce complexe agit comme un répresseur transcriptionnel (Lam et al., 2006). L'identification du site de liaison entre l'ATXN1 et la CIC a permis de générer un mutant ATXN1 incapable de lier CIC. La surexpression de ce mutant ATXN1-polyQ n'est plus toxique chez la souris, suggérant un gain de fonction toxique du complexe ATXN1-polyQ-CIC dans les cellules de Purkinje (Figure 13 ; Rousseaux et al., 2018). Ces résultats expliquent d'ailleurs la plus faible toxicité du mutant ATXN1 S776A (cf Introduction 1.4.2.2) qui perd en partie sa capacité de liaison à CIC (Emamian et al., 2003 ; Lam et al., 2006). Enfin, la perte de la protéine CIC conduit des à une déficience intellectuelle associée à un spectre autistique chez l'Homme et à des altérations de l'apprentissage et de la mémoire en modèle murin, soulignant ainsi l'importance de ces protéines pour le fonctionnement neuronal (Lu et al., 2017).

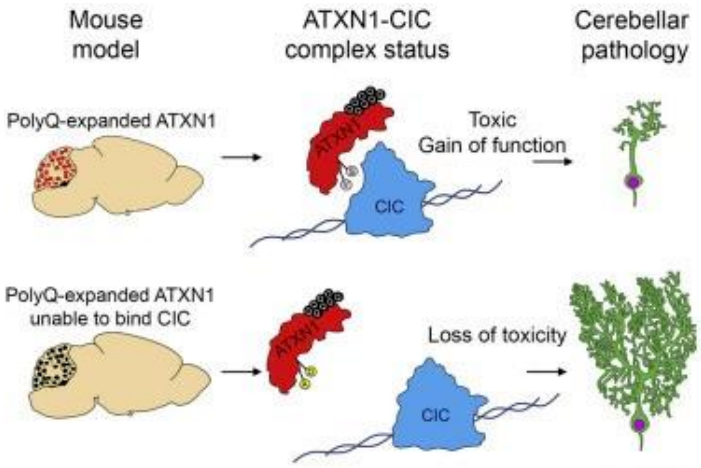


Figure 13. La toxicité de l'ATXN1-polyQ dépend de sa liaison à la protéine CIC
La neurodégénérescence du cervelet provoquée par l'ATXN1-polyQ est abolie lorsque cette protéine est incapable de lier la protéine CIC. Image de Rousseaux et al., 2018.

Pour terminer, SCA6 est due à une courte expansion de nucléotides CAG dans le gène *CACNA1A*, codant pour un canal calcique voltage-dépendant. La présence d'une expansion de polyQ dans cette protéine n'altère pas systématiquement sa localisation, mais altère les propriétés calciques de ce canal pouvant expliquer la mort des cellules de Purkinje (Restituito et al., 2000).

1.4.2.5. *Conclusion*

Ces nombreux travaux mettent en lumière la complexité des différents mécanismes pouvant intervenir dans les maladies à polyQ. En effet, bien qu'une toxicité intrinsèque des polyQ soit admise, cette toxicité peut être modulée par les séquences entourant ces polyQ. Ces séquences bordant cette expansion peuvent également influer sur la localisation et / ou sur la capacité d'agrégation des protéines à polyQ. A ce titre, l'agrégation de ces protéines semble avoir un effet protecteur ; les formes solubles des protéines à polyQ étant particulièrement toxiques. De plus, selon la pathologie considérée, la toxicité des protéines à polyQ dépend de leur localisation cellulaire, nucléaire ou cytoplasmique. Enfin, une perte de fonction partielle des protéines ou un gain de fonction toxique des protéines à polyQ peuvent jouer un rôle plus ou moins important dans le développement des différentes pathologies.

1.4.3. Les maladies à polyAlanine (polyA)

A ce jour, on dénombre 9 maladies à polyA : la dystrophie musculaire oculo-pharyngée (OPMD) due à une expansion de répétitions des tri-nucléotides GCG dans le gène *PABPN1* ; et 8 maladies développementales dues à des expansions GCN dans les gènes *ARX*, *FOXL2*, *HOXA13*, *HOXD13*, *PHOX2B*, *RUNX2*, *SOX3* et *ZIC2* codant tous pour des facteurs de transcription et responsables respectivement de la déficience intellectuelle non-syndromique liée à l'X (NS-XLID), le syndrome de blépharophimosis-ptosis-épicanthus inversus (BPES), le syndrome main-pied-génital (HFGS), la syndactilie de type 1 et 5 (SPD), le syndrome d'hypoventilation alvéolaire central congénital (CCHS), la brachydactylie, l'hypopituitarisme lié à l'X (XH) et l'holoprosencéphalie de type 5 (HPE5). Ces maladies sont dues à un court allongement, de 5 à 17 répétitions GCN, de séquences de microsatellites préexistantes de 3 à 20 nucléotides GCN. Toutefois, il est important de noter que les 8 maladies développementales citées précédemment ne sont pas uniquement dues à des expansions GCN. En effet, de nombreuses mutations perte de fonction de type faux-sens, non-sens, décalage du cadre de lecture, etc. sont également décrites comme provoquant ces pathologies. De ce fait, il est probable que le mécanisme de toxicité sous-jacent à ces expansions de polyAlanine consiste en la perte de fonction de ces facteurs de transcription et / ou, pour les expansions les plus longues, en la naissance de protéine ayant un effet dominant négatif sur la transcription de gènes impliqués dans le développement. Enfin, ces protéines avec les expansions les plus longues ont tendance à former des agrégats cytoplasmiques et / ou nucléaires (pour revue, Messaed et Rouleau, 2009).

Enfin, je vais m'attacher à décrire plus en avant l'OPMD, car cette maladie présente des similitudes cliniques avec les myopathies oculo-pharyngo-distales (OPDM), abordées dans ce travail de thèse. Les personnes affectées par l'OPMD présentent une faiblesse musculaire des membres distaux, un ptosis, c'est-à-dire un abaissement des paupières, et une dysphagie. D'un point de vue histopathologique, les patients présentent dans leurs fibres musculaires des vacuoles ainsi que des inclusions intranucléaires caractéristiques (Figure 14 ; Tomé et Fardeau, 1980).

L'OPMD est due à l'allongement à l'état hétérozygote de 2 à 8 répétitions ou, à l'état homozygote d'une seule répétition, de tri-nucléotides GCG dans le premier exon du gène *PABPN1* ; la population générale présentant 10 répétitions GCG (Brais et al., 1998). Ce gène

code pour la « poly(A) binding-protein nuclear 1 » (PABPN1), une protéine nucléaire qui contrôle la formation de la queue poly(A) à l'extrémité 3' des pré-ARNm.

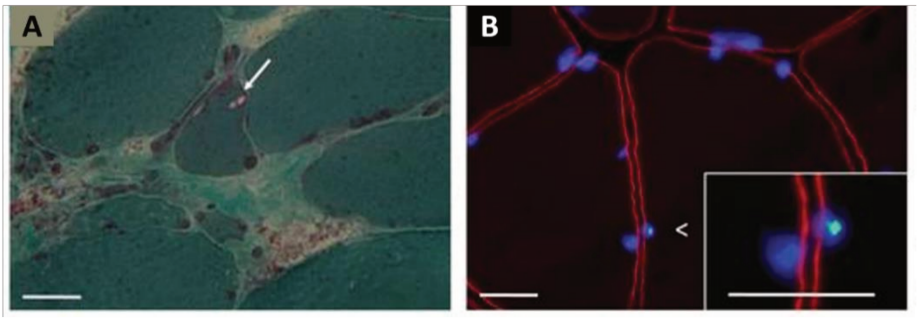


Figure 14. Caractéristiques histopathologiques des patients OPMD
(A, B) Fibres musculaires de patients OPMD présentant (A) des vacuoles et, (B) des inclusions intranucléaires en vert. Echelle 50 µm. Images de Richard et al., 2015.

L'élongation du nombre de résidus alanines dans la partie N-terminale de la protéine PABPN1 promeut son agrégation sous forme d'inclusions insolubles, caractéristiques des muscles des patients OPMD (Calado et al., 2000). Le mécanisme de toxicité de cette expansion reste encore mal compris. Des données suggèrent un gain de fonction toxique où, dans des myoblastes de patients OPMD, la protéine PABPN1-polyA séquestrerait différents ARNm sans modifier la taille de leur queue poly(A) (Calado et al., 2000). D'autres études suggèrent une perte de fonction partielle de PAPBN1. En effet, des analyses de transcriptomiques à partir de muscles d'un modèle murin transgénique exprimant 17 résidus alanine dans le gène *PAPBN1* révèlent une dérégulation massive de la polyadénylation des ARN, pouvant ainsi expliquer l'atrophie musculaire (Trollet et al., 2010).

1.4.4. Les maladies à polyGlycine (polyG)

Lorsque j'ai commencé ma thèse, les maladies à polyG étaient au nombre de 2 : le syndrome de tremblement et d'ataxie lié à l'X fragile (FXTAS) et, de manière incertaine, l'insuffisance ovarienne prématurée liée à l'X (FXPOI). Ces 2 pathologies sont causées par la même mutation, une expansion de 55 à 200 répétitions de tri-nucléotides CGG dans la région 5'UTR du gène *FMR1* (Murray et al., 1998 ; Hagerman et al., 2001).

1.4.4.1. *Le syndrome de tremblement et d'ataxie lié à l'X fragile (FXTAS)*

Le syndrome de tremblement et d'ataxie lié à l'X fragile (FXTAS) a été décrit pour la première fois en 2001 par la Professeure Randi Hagerman (Hagerman et al., 2001). Il s'agit d'une maladie neurodégénérative caractérisée par une démarche ataxique et / ou un tremblement intentionnel ; les patients peuvent également présenter des altérations neuropsychiatriques et cognitifs ainsi que des symptômes parkinsoniens (Jacquemont et al., 2003). Cette maladie touche principalement les hommes et apparait généralement entre 60 et 65 ans avec une espérance de vie comprise entre 5 et 25 ans. Des analyses d'IRM cérébrales de patients FXTAS ont révélé une hyper-intensité en T2 des pédoncules cérébelleux moyens et de la matière blanche adjacente (Figure 15 ; Jacquemont et al., 2003 ; Leehey et al., 2003).

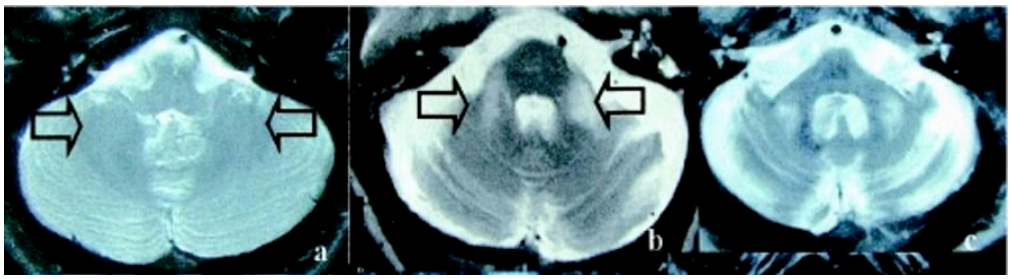


Figure 15. Hyper-intensité en T2 des pédoncules cérébelleux dans FXTAS
(**a**, **b**, **c**) IRM de cervelets (**a**) d'un individu contrôle et (**b**, **c**) d'un patient FXTAS présentant une hyper-intensité en T2 des pédoncules cérébelleux. Images de Jacquemont et al. 2003.

D'un point de vue histopathologique, l'analyse des cerveaux de patients FXTAS montrent une diminution du nombre de cellules de Purkinje, une gliose des cellules de Bergmann (Greco et al., 2002) ainsi qu'une spongiose de la substance blanche des pédoncules moyens du cervelet (Greco et al., 2006). De plus, des inclusions intranucléaires éosinophiliques et positives à l'ubiquitine et p62 sont retrouvées dans les neurones et les astrocytes (Figure 16 ; Greco et al.,

2006), mais aussi dans de nombreux autres tissus à l'exception des muscles squelettiques (Hunsaker et al., 2011).

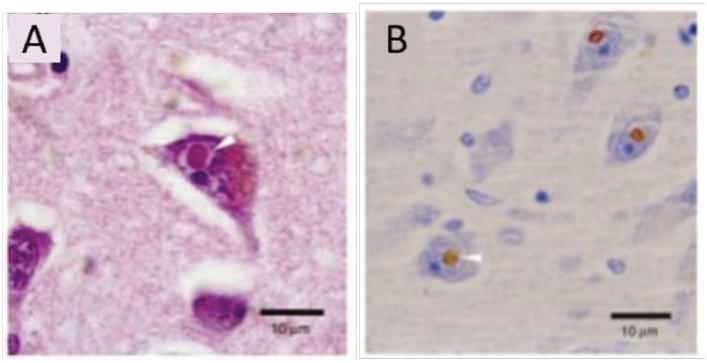


Figure 16. Inclusions intranucléaires chez des patients FXTAS
(**A**, **B**) Présence d'inclusions intranucléaires dans les cellules de Purkinje du cervelet révélées par (**A**) une coloration hématoxyline / éosine et (**B**) par un anticorps dirigé contre l'ubiquitine. Images de Greco et al., 2002.

La cause génétique de FXTAS est une expansion de 55 à 200 répétitions de tri-nucléotides CGG dans la région 5'UTR du gène *FMR1*. Contrairement à FXS où ces expansions excèdent 200 répétitions CGG, FXTAS présente des répétitions plus courtes, qu'on appelle alors prémutation. Cette prémutation ne conduit pas à une hyper-méthylation du promoteur du gène *FMR1* comme dans FXS. En effet, les patients FXTAS présentant une quantité d'ARNm de FMR1 2 à 8 fois plus élevée en comparaison à la population générale (Tassone et al., 2000), écartant ainsi l'hypothèse d'une toxicité au niveau ADN. Par la suite, l'hypothèse d'une toxicité au niveau à ARN a été largement investiguée avec plus d'une vingtaine de protéines décrites comme liant les ARN composées de répétitions de nucléotides CGG comme hnRNP A2, MBNL, Purα, CUGBP1, SAM68, etc. (Iwahashi et al., 2006 ; Jin et al., 2007 ; Sofola et al., 2007 ; Sellier at al., 2010). Néanmoins, l'incidence d'une éventuelle perte de fonction de ces protéines liant l'ARN dans FXTAS reste à démontrer. De plus, les agrégats intranucléaires retrouvés chez les patients atteints de FXTAS sont différents de ceux retrouvés chez les patients DM par leur taille et leur composition. En effet, les agrégats des patients FXTAS sont au moins 10 fois plus larges que les foci d'ARN nucléaires typiquement retrouvés chez les patients DM et surtout, ils sont enrichis en protéines impliquées dans la dégradation des protéines (ubiquitine et p62) contrairement aux foci d'ARN typiques des DM (Greco et al., 2002). Ces données, en plus de la découverte que certaines répétitions situées dans des régions non-codantes des gènes pouvaient être traduites (Zu et al., 2011), ont conduit à s'interroger sur une possible traduction « RAN » des répétitions CGG dans FXTAS. Des expériences de transfection de culture

cellulaire ont ainsi montré que ces répétitions CGG pouvaient être traduites en protéines composées de polyAlanine (polyA) ou de polyGlycine (polyG) (Todd et al., 2013), et ceci en absence de codon d'initiation de la traduction ATG. Grâce au développement d'anticorps dirigés spécifiquement contre ces protéines à polyA ou à polyG et au développement de modèles de drosophiles, il a alors été montré que les inclusions des patients FXTAS sont majoritairement composées de protéines à polyG, appelées alors FMRpolyG ; l'expression de FMRpolyA étant très faible et cette protéine étant détectée uniquement après sa surexpression en modèles cellulaires (Figure 17 ; Buijsen et al., 2014 ; Sellier et al. 2017 ; Krans et al., 2019).

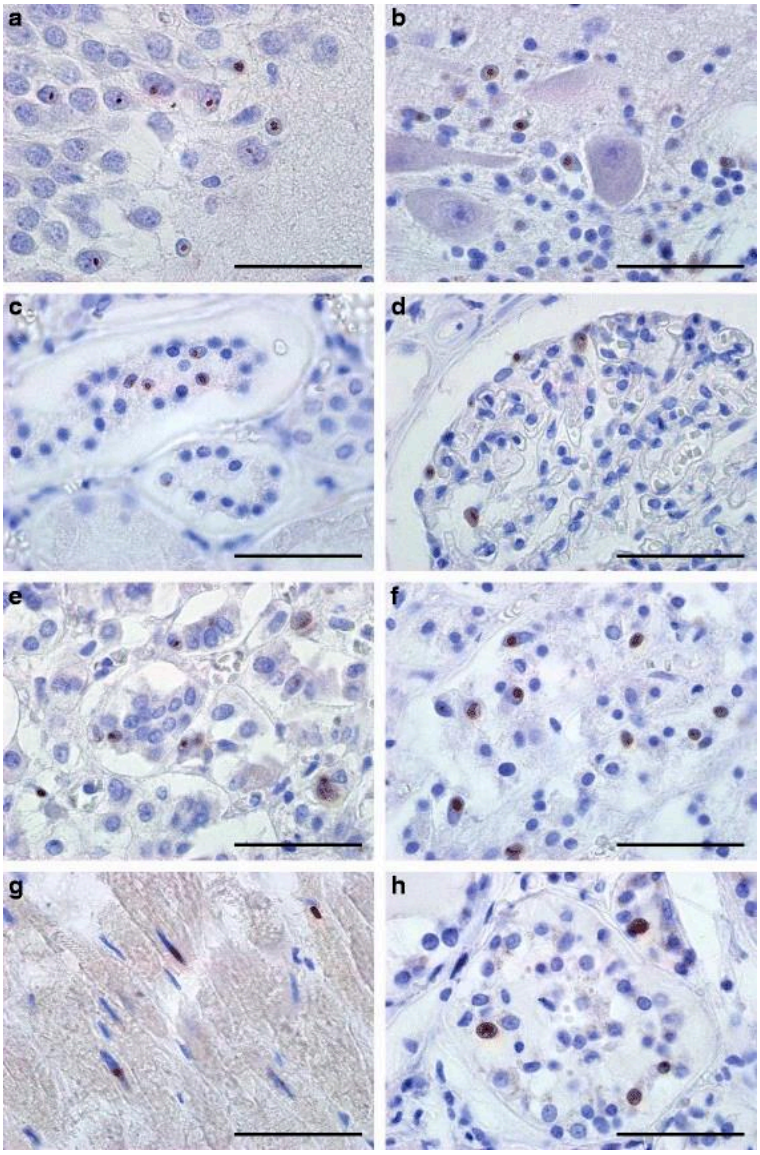


Figure 17. FMRpolyG dans les inclusions intranucléaires chez les patients FXTAS
(**a**, **b**, **c**, **d**, **e**, **f**, **g**, **h**) IHC dirigée contre la protéine FMRpolyG révélant de nombreuses inclusions intranucléaires chez les patients FXTAS dans (**a**) l'hippocampe, (**b**) le cervelet, (**c**) les glomérules et (**d**) les tubes distaux des reins, (**e**) la zone glomérulaire et (**f**) la zone réticulaire des glandes surrénales, (**g**) les cardiomyocytes et (**h**) la thyroïde. Echelle 50 µm. Images de Buijsen et al., 2014.

Des analyses de spectrométrie de masse sur cette protéine FMRpolyG ont ensuite permis d'identifier que l'initiation de la traduction avait lieu en amont des répétitions CCG à un codon ressemblant à un ATG : un codon ACG, révélant ainsi la présence d'une petite uORF présente en amont du gène *FMR1*. Afin de discerner le gain de fonction au niveau ARN du gain de fonction protéique, deux modèles murins transgéniques ont alors été généré : le premier avec la séquence 5'UTR du gène *FMR1* avec 100 répétitions CGG permettant la transcription en un ARN contenant les répétitions et leur traduction en FMRpolyG et ; le second, délété de la région contenant le codon ACG qui permet la traduction de la protéine FMRpolyG, ce modèle exprimant donc uniquement l'ARN contenant les répétitions CGG (Figure 18). De façon intéressante, seul le premier modèle développe une ataxie et des agrégats intranucléaires typiques des patients FXTAS, suggérant que la protéine FMRpolyG est responsable des symptômes et des caractéristiques histopathologiques de cette pathologie. De plus, outre la confirmation que l'initiation de la traduction de la protéine FMRpolyG a lieu en amont des répétitions CGG grâce à un codon non-AUG, ces résultats éliminent la possibilité d'une toxicité de protéines qui seraient traduites par un mécanisme « RAN » initiant directement à l'intérieur des répétitions. Enfin, ces résultats excluent la toxicité au niveau ARN, puisque la seule délétion du codon ACG ne modifie pas l'ARN composé de répétitions CGG (Sellier et al., 2017).

Figure 18. Modèles murins de FXTAS prouvant la toxicité de FMRpolyG
Les 2 modèles murins sont porteurs d'une expansion de 100 répétitions CGG fusionnées à la GFP dans le cadre de lecture glycine ; ils expriment donc les ARN contenant les répétitions. Le premier modèle présente, en amont des répétitions, la région 5'UTR du gène *FMR1* permettant la traduction de la protéine FMRpolyG, ce qui n'est pas le cas du second. Image de Sellier et al. 2017.

Enfin, le mécanisme par lequel les expansions de répétitions CGG dans FXTAS est toxique reste encore mal compris. Il a été montré que les ARN contenant des répétitions CGG pouvaient créer des boucles-R interagissant et provoquant des cassures de l'ADN (Loomis et al., 2014 ; Robin et al., 2017), altérer les fonctions mitochondriales (Hukema et al., 2014 ; Robien et al.,

2017) ou encore diminuer la production des micro-ARN mitochondriaux (Gohel et al. 2021). La protéine FMRpolyG, quant à elle, est décrite pour altérer la structure de l'enveloppe nucléaire via son interaction avec Lap2β (Sellier et al., 2017), le système ubiquitine-protéasome (Oh et al., 2015) ou encore l'intégrité mitochondriale via l'assemblage des complexes de la chaîne respiratoire (Gohel et al., 2019). Il reste donc à déterminer les causes précises de la mort neuronale dans FXTAS, et quelles sont les voies moléculaires directement affectées par la protéine FMRpolyG.

1.4.4.2. *L'insuffisance ovarienne prématurée liée à l'X (FXPOI)*

L'insuffisance ovarienne prématurée liée à l'X (FXPOI) est caractérisée par une ménopause précoce et fait partie d'un syndrome plus vaste, les insuffisances ovariennes primaires (POF). Les causes génétiques de FXPOI sont multiples et, parmi elles, une expansion de 55 à 200 répétitions CGG dans la région 5'UTR du gène *FMR1* est responsable de 20% des cas (Man et al., 2017). Les cellules ovariennes de patientes FXPOI, et plus spécifiquement celles de la granulosa, montrent des inclusions intranucléaires positives pour FMRpolyG (Buijsen et al., 2016 ; Friedman-Gohas et al., 2020). De plus, un modèle murin surexprimant la protéine FMRpolyG présente un arrêt précoce de leur capacité reproductive dû à une cessation de l'ovulation (Shelly et al., 2021). Il est à noter que la seule expression de l'ARN contenant les répétitions CGG conduit également à des altérations de la réponse ovulatoire face à des stimulations hormonales. Toutefois, il s'agit d'anomalies mineures comparées au modèle exprimant la protéine FMRpolyG (Shelly et al., 2021).

Ces résultats suggèrent que FXTAS, et peut-être FXPOI, sont causés par une traduction des expansions de répétitions CGG en une protéine composée de polyGlycine toxique.

1.5. Les maladies complexes

Les différents mécanismes de toxicité (au niveau ADN, ARN ou protéique) proposés pour expliquer la pathogénicité des expansions de répétitions ne sont pas exclusifs. Par exemple, l'ataxie spinocérébelleuse de type 31 (SCA31) est causée par une expansion intronique de répétitions de penta-nucléotides soit TGGAA dans le gène *BEAN1*, soit TTCCA dans le gène *TK2*, selon le brin d'ADN considéré (Figure 19 ; Sato et al., 2009). L'impact de la transcription des répétitions TTCCA localisées dans le gène *TK2* reste à éclaircir mais, dans le sens de lecture du gène *BEAN1*, la transcription de ces répétitions produit des ARN contenant des répétitions UGGAA qui forment des foci d'ARN nucléaires dans les cellules de Purkinje des patients (Niimi et al., 2013). Plusieurs protéines liant l'ARN sont associées à ces ARN dont TDP-43, FUS et hnRNPA2. De plus, il a été montré que ces ARN étaient traduits en protéines composées de polyWNGME présentes chez les patients (Figure 19). De manière intéressante, la surexpression de FUS et hnRNPA2, deux protéines liant l'ARN, diminue le nombre de foci d'ARN et d'agrégats protéiques polyWNGME, suggérant un effet protecteur de certaines protéines chaperonnes face à la toxicité des expansions de répétitions (Ishiguro et al., 2017).

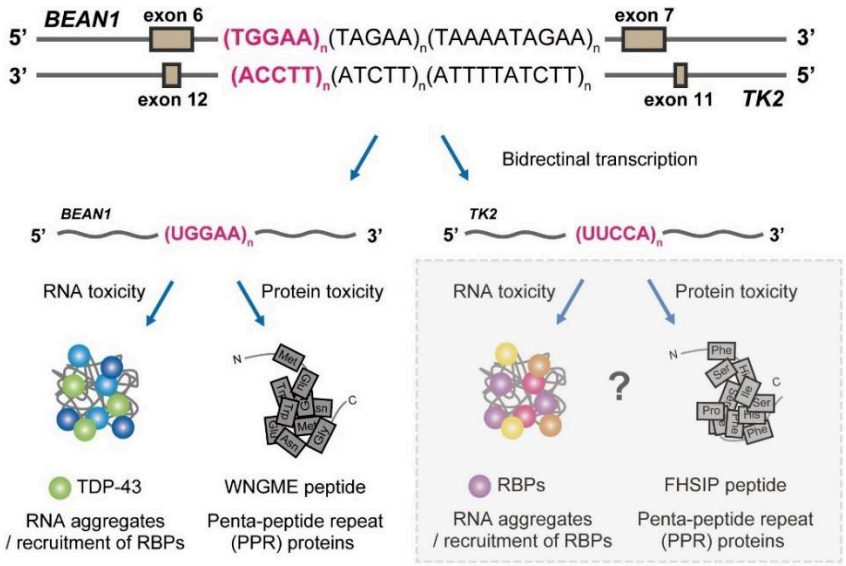


Figure 19. Mécanisme de toxicité proposé pour SCA31
Les répétitions TGGAA dans le gène *BEAN1* sont transcrites en ARN formant des foci séquestrant des protéines de liaison à l'ARN, qui sont traduits en protéines composées de polyWNGME toxiques. L'impact de répétitions ACCTT dans le gène *TK2* reste à éclaircir. Image de Ishiguro et al., 2021.

De manière similaire, il a été montré dans la DM2 qu'en plus d'une toxicité au niveau ARN, les répétitions sens CCTG et antisens CAGG peuvent être traduites grâce au mécanisme de traduction « RAN » en protéines toxiques composées de polyLPAC et polyQAGR respectivement. Il est important de noter que la présence des foci d'ARN est inversement corrélée à la présence des agrégats protéiques, suggérant que les protéines MBNL liant l'ARN permettent la rétention dans le noyau des ARN contenant les répétitions et préviennent ainsi leur traduction (Zu et al., 2017).

Un troisième exemple de maladie à mécanisme complexe que je développerais plus en détail par la suite (cf Introduction 3) est l'expansion de répétitions GGGGCC située dans le premier intron du gène *C9ORF72* comme première cause de sclérose latérale amyotrophique (SLA) et de démence fronto-temporale (DFT).

1.6. Les maladies à mécanisme encore inconnu

Ces dernières années, de nouvelles expansions de répétitions ont été identifiées dans différentes pathologies, dont le syndrome d'ataxie cérébelleuse avec neuropathie et aréflexie vestibulaire bilatérale (CANVAS) ou encore les épilepsies myocloniques familiales de l'adulte (FAME).

CANVAS est une maladie autosomale récessive causée par une expansion de 400 à 2000 répétitions de nucléotides AAGGG dans l'intron 2 du gène *RFC1* (Cortese et al., 2019 ; Rafehi et al., 2019). De manière intéressante, ces expansions AAGGG sont pathogènes à l'état homozygote et remplacent une expansion de microsatellites existante dans le gène composée de $AAAAG_{11x}$ / $AAAGG_{nx}$ chez les individus contrôles. La présence de cette expansion ne modifie pas le niveau d'expression de l'ARN RFC1, écartant ainsi une possible perte de fonction de la protéine RFC1 (Cortese et al., 2019). De plus, l'absence d'inclusions protéiques dans le cerveau des patients CANVAS pourrait laisser penser à un gain de fonction au niveau ARN, sans exclure la possibilité d'un nouveau mécanisme de toxicité de cette expansion.

Les épilepsies myocloniques familiales de l'adulte (FAME, également appelée BAFME, ADCME et FMCTE) sont des maladies autosomales dominantes caractérisées par des épilepsies et des tremblements des muscles (myoclonus). On distingue 6 types de FAME selon le gène muté : *SAMD12* dans FAME1, *STARD7* dans FAME2, *MARCHF6* dans FAME3, *YEATS2* dans FAME4, *TNRC6A* dans FAME6 et *RAPGEF2* dans FAME7 (Ishiura et al., 2018 ; Corbett et al., 2019 ; Florian et al., 2019 ; Yeetong et al., 2019). Toutefois, la mutation responsable de ces FAME est identique ; il s'agit d'une expansion de plusieurs centaines à plusieurs milliers de répétitions ATTTC situées dans les régions introniques de ces gènes. La taille (plusieurs milliers de répétitions), la localisation intronique ainsi que la diversité des gènes mutés suggèrent un mécanisme de toxicité commun indépendant du gène hôte, rendant la perte de fonction de ces protéines peu probable. En revanche, ces grandes expansions situées dans des régions non-codantes rappellent celles observées dans les DM, suggérant peut-être un mécanisme toxique commun au niveau ARN. De manière intéressante, bien que certains gènes comme *STARD7* et *MARCHF6* soient exprimés de manière ubiquitaire, les patients présentent uniquement des altérations neuronales. Ces données suggèrent une toxicité des expansions ATTTC seulement dans le système nerveux central, via, peut-être, l'altération de la fonction d'une protéine spécifiquement neuronale. De plus, des expansions proches dans SCA10 (ATTCT dans le gène *ATXN10*), SCA31 (ATTCC dans le gène *TK2*) et SCA37 (ATTTC dans le gène *DAB1*) pourrait laisser penser à un mécanisme de toxicité similaire aux FAME.

Enfin, parmi les maladies à mécanisme encore inconnu, plusieurs expansions comprises entre 50 et 200 répétitions de tri-nucléotides CGG localisées dans les régions 5'UTR de différents gènes (*LRP12*, *GIPC1*, *NOTCH2NLC*, etc.) ont récemment été identifiées et, ces différentes pathologies sont présentées dans le chapitre suivant (Deng et al., 2019 ; Ishiura et al., 2019 ; Sone et al., 2019 ; Tian et al., 2019 ; Deng et al., 2020 ; Yu et al., 2021).

INTRODUCTION

- CHAPITRE 2 -

Maladies à expansions CGG dans les régions 5’UTR

2. MALADIES A EXPANSIONS CGG DANS LES REGIONS 5'UTR

Après la découverte en 2001 d'une expansion de 50 à 200 répétitions CGG dans la région 5'UTR du gène *FMR1* comme cause de FXTAS (cf Introduction 1.4.4.1), d'autres expansions similaires ont été reportées, notamment dans la maladie neuronale à inclusions intranucléaires (NIID), la myopathie oculo-pharyngo-distale (OPDM) et la myopathie oculo-pharyngée avec leucodystrophie (OPML). La pathologie FXTAS ayant été décrite précédemment (cf Introduction 1.4.4), je vais maintenant décrire les maladies NIID, OPDM et OPML. Cette description a fait l'objet d'une revue présentée à la suite de ce chapitre.

2.1. La maladie à inclusions intranucléaires neuronaux (NIID)

2.1.1. Epidémiologie et clinique

La maladie à inclusions intranucléaires neuronaux (NIID, « Neuronal Intranuclear Inclusion Disease ») est une pathologie d'origine génétique à transmission autosomale dominante. Cette maladie neurodégénérative et neuromusculaire présente des manifestations cliniques nombreuses et diverses : démence, neuropathie périphérique, dystonie neurovégétative, ataxie cérébelleuse, parkinsonisme, tremblement essentiel, convulsions, accident vasculaire cérébral, trouble de la conscience, encéphalopathie, faiblesse des muscles squelettiques, etc., qui peuvent varier grandement d'un patient à l'autre (Sone et al., 2016). De même, l'âge d'apparition des premiers symptômes varie grandement et des formes infantiles, juvéniles et adultes sont décrites. Une classification a été envisagée selon le symptôme initial majeur et l'âge d'apparition des symptômes où la pathologie apparaît (Sone et al., 2016) :

- Chez l'adolescent ou le jeune adulte où le symptôme prédominant est une faiblesse musculaire des membres distaux ;
- Chez l'adulte où les symptômes principaux sont un parkinsonisme et / ou un tremblement essentiel ;
- Chez l'adulte où le symptôme prédominant est une démence.

2.1.2. Histopathologie

Des analyses d'IRM cérébrales de patients NIID montrent que cette pathologie est caractérisée, tout comme FXTAS, par des lésions hyper-intenses en T2 des pédoncules cérébelleux moyens et de la substance blanche (leuco-encéphalopathie) (Figure 20 ; Sasaki et al., 2015 ; Morimoto et al., 2017 ; Takumida et al, 2017).

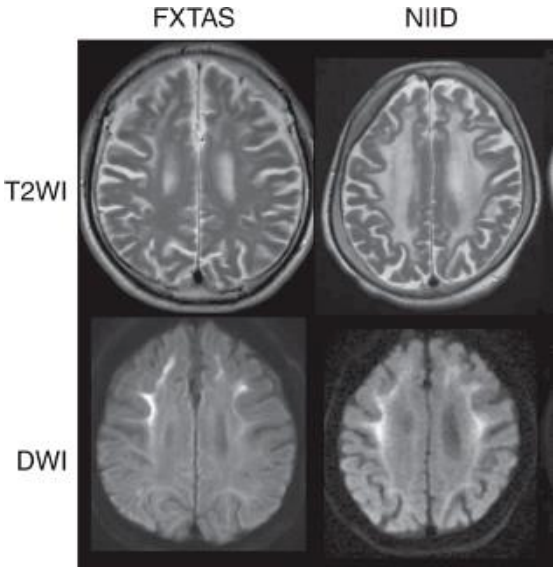


Figure 20. IRM de patients FXTAS et NIID
IRM d'un patient FXTAS (à gauche) et NIID (à droite) où une hyper-intensité en T2 dans les jonctions cortico-médullaires est visible par DWI et T2WI. Images de Ishiura et al., 2019.

Enfin, comme son nom l'indique, cette maladie est caractérisée par la présence caractéristique d'inclusions intranucléaires éosinophiliques positives à l'ubiquitine et à p62. De façon intéressant, ces inclusions sont strictement identiques à celles décrites dans FXTAS (Figure 21, Gelpi et al., 2017 ; Lim et al., 2020 ; Toko et al., 2021). Ces agrégats sont présents dans les neurones et astrocytes des systèmes nerveux central, périphérique et autonome ainsi que dans les cellules des organes viscéraux (pour revue, Kimber et al., 1998). Toutefois, il est aussi possible d'observer ces inclusions dans les fibroblastes, les adipocytes et les glandes sébacées, permettant ainsi de confirmer le diagnostic de ce syndrome par biopsie cutanée (Sone et al., 2011). Cette observation a permis de mieux diagnostiquer la maladie. En effet, le premier patient atteint de NIID a été décrit en 1968 (Lindenberg et al., 1968) et seulement une quarantaine de cas, à la fois sporadiques et familiaux, ont été répertoriés jusqu'en 2011. Depuis, le diagnostic de la maladie grâce aux inclusions intranucléaires présentes dans la peau des patients puis, l'identification de la mutation responsable de NIID, ont permis de révéler la présence de plusieurs centaines de cas à travers le monde.

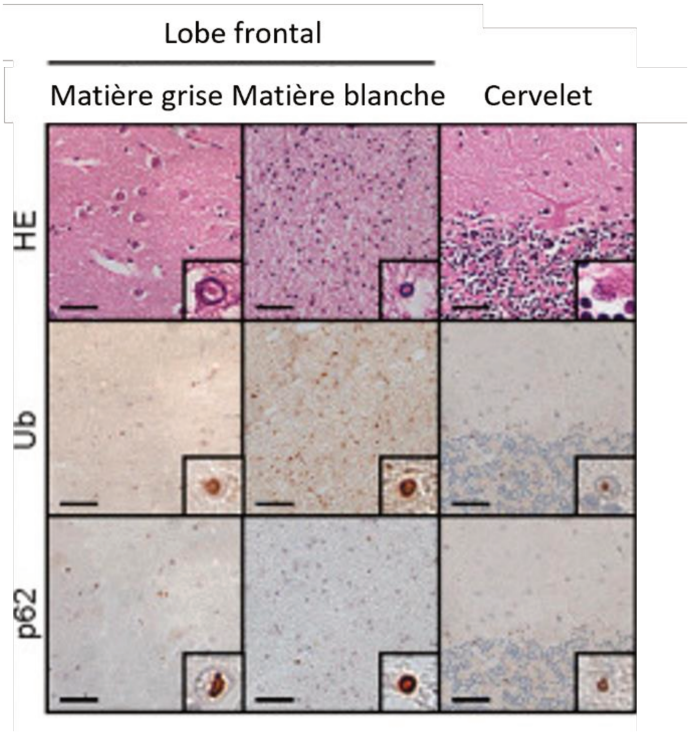


Figure 21. Inclusions intranucléaires dans le cerveau des patients NIID
Présence d'inclusions intranucléaires révélées par une coloration hématoxyline / éosine, et qui sont positives pour l'ubiquitine et p62, dans les matières grise et blanche du lobe frontal et dans le cervelet d'un patient NIID. Images de Yamaguchi et al. 2018.

2.1.3. Mutation responsable de NIID

La cause génétique de cette maladie a été récemment identifiée comme une expansion comprise entre 60 et 200 répétitions de tri-nucléotides CGG située dans la région 5'UTR du gène *NOTCH2NLC* (Deng et al., 2019 ; Ishiura et al., 2019 ; Sone et al., 2019 ; Tian et al., 2019). De plus, cette expansion de répétitions est également retrouvée chez des patients initialement diagnostiqués pour d'autres pathologies, comme le tremblement essentiel (TE), ou la myopathie oculo-pharyngo-distale de type 3 (OPDM3), montrant la grande variabilité des symptômes des patients NIID (Ng et al., 2020 ; Yu et al., 2021). Il est important de noter que cette mutation est retrouvée chez les populations de l'est de l'Asie, et non chez les populations européennes, suggérant qu'un autre locus est impliqué dans le développement de la pathologie en Europe (Chen et al., 2020a).

NOTCH2NLC, avec *NOTCHNLA* et *NOTCH2NLB*, sont des paralogues tronqués du gène *NOTCH2*, et qui sont présents uniquement chez l'Homme (Fiddes et al., 2018 ; Suzuki et al.,

2018a). Les 4 protéines NOTCH (NOTCH1, NOTCH2, NOTCH3 et NOTCH4) sont hautement conservées dans le règne animal. Ces protéines transmembranaires font toutes partie de la voie de signalisation NOTCH intervenant dans la différenciation, la prolifération et l'apoptose durant la morphogénèse (pour revue, Artavanis-Tsakonas et al., 1999). Il a été montré que les paralogues de NOTCH2, notamment NOTCH2NLA/B, sont des régulateurs de la neurogénèse intervenant notamment dans le développement cortical en favorisant la prolifération des progéniteurs neuronaux chez l'Homme (Figure 22). De plus, les gènes *NOTCH2NLA/B* sont situés sur la région chromosomique responsable du syndrome 1q21.1, caractérisé par un retard mental et (1) une microcéphalie lorsque cette région subit une délétion ou, (2) une macrocéphalie lorsque cette région est dupliquée (Fiddes et al., 2018 ; Suzuki et al., 2018a).

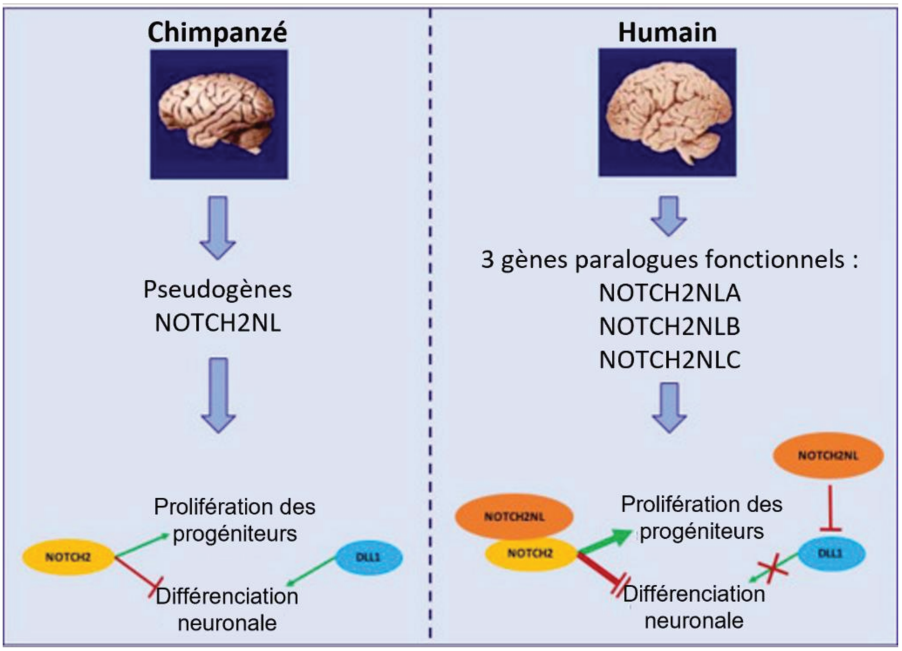


Figure 22. Rôle des paralogues de NOTCH2 dans le développement cortical humain
Les chimpanzés sont porteurs de gènes non-fonctionnels, appelés pseudogènes, de NOTCH2NL. Chez les humains, 3 gènes paralogues, et donc fonctionnels, codent pour des protéines permettant de favoriser la prolifération des progéniteurs corticaux et d'inhiber la différenciation neuronale, conduisant au développement d'un cortex plus complexe. Image de Bizzotto et al., 2018.

Mon travail de thèse a participé à déterminer comment cette expansion de répétitions CGG située dans la région non-codante du gène *NOTCH2NLC* peut se révéler pathogénique chez l'Homme.

2.2. Les myopathies oculo-pharyngo-distales (OPDM)

2.2.1. Epidémiologie et clinique

Les myopathies oculo-pharyngo-distales (OPDM) sont des maladies génétiques autosomales dominantes affectant les muscles squelettiques et se manifestant à l'âge adulte. Depuis la première description de cette maladie en 1977, environ 300 patients atteints d'OPDM ont été recensés dans le monde, essentiellement en Chine, au Japon, en Europe et aux Etats-Unis (Satoyoshi et Kinoshita, 1977 ; Deng et al., 2020). D'un point de vue clinique, les OPDM sont caractérisées par un ptosis progressif (abaissement des paupières), une ophtalmoplégie ainsi qu'une faiblesse des muscles faciaux, pharyngés et des membres distaux (Figure 23). Cet affaiblissement musculaire est associé à une augmentation modérée du taux de créatine kinase sérique, une insuffisance respiratoire précoce et des risques de problèmes cardiovasculaires sur le long terme. L'âge moyen d'apparition de la maladie est estimé à 28 ans, plus ou moins 11 ans, et le phénotype complet de l'OPDM se développe généralement sur une dizaine d'années (Durmus et al., 2011).

Figure 23. Evolution d'un patient atteint d'OPDM
(**A**) A 16 ans, ptosis bilatéral comme premier symptôme. (**B**) A 20 ans, ptosis et atrophie de la face. (**C**-**E**) A 43 ans, sévère atrophie des muscles faciaux et distaux des brase et des jambes (durée de la maladie : 27 ans). Images de Durmus et al., 2011.

2.2.2. Histopathologie

Les muscles de patients atteints d'OPDM présentent une variation de la taille des fibres musculaires, des vacuoles, une fibrose endomysiale et un remplacement graisseux léger à sévère (Figure 24A, 24B, 24C, 24D ; Durmus et al., 2011 ; Zhao et al., 2015b). Toutefois, ces caractéristiques sont retrouvées dans de multiples myopathies et ne sont pas spécifiques de l'OPDM. En revanche, l'OPDM est caractérisée par la présence d'inclusions intranucléaires éosinophiliques de 1 à 2 µm de diamètre, positives à p62 et à l'ubiquitine, et d'aspect tubulo-filamenteux en microscopie électronique (Figure 24E, 24F ; Zhao et al., 2015b ; Ogasawara et al., 2020). Ces inclusions très particulières ne sont pas retrouvées dans d'autres types de myopathies, mais rappellent les inclusions intranucléaires observées dans les maladies neurodégénératives NIID et FXTAS.

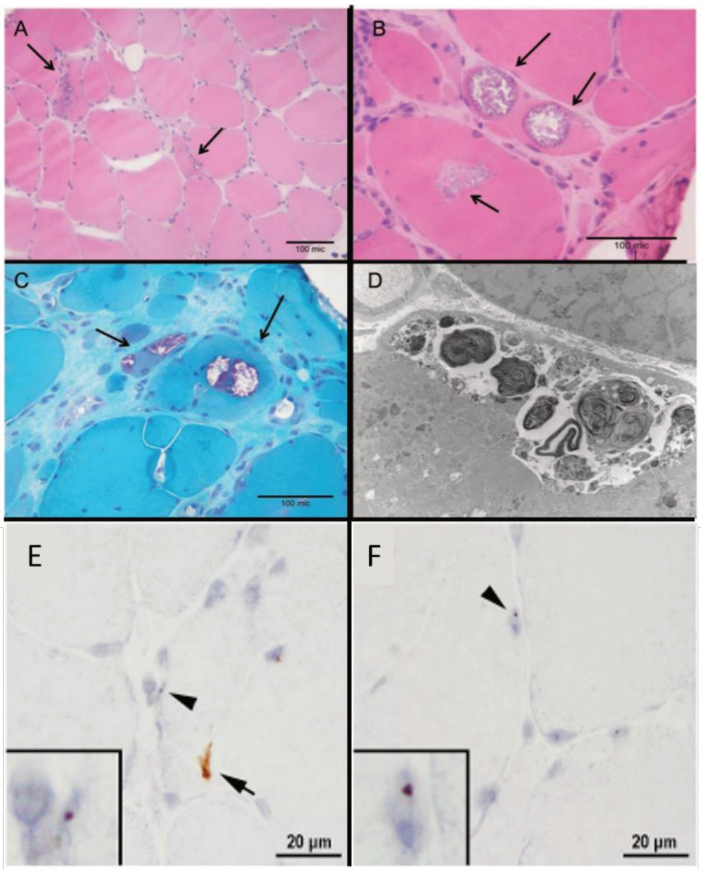


Figure 24. Caractéristiques histopathologiques des fibres musculaires des patients OPDM (**A**, **B**) Coloration hématoxyline / éosine montrant la variation de la taille des fibres et les vacuoles du tibialis antérieur. (**C**) Coloration au trichome de Gomori révélant les vacuoles en rouge du biceps. (**D**) Microscopie électronique des vacuoles autophagiques contenant des structures multi-lamellées du tibialis antérieur. (**E**, **F**) Immunomarquage de p62 dans des fibres musculaires révélant la présence de d'inclusions et intranucléaires et de vacuoles. Images de Durmus et al., 2011 et Deng et al., 2020.

2.2.3. Mutations responsables de l'OPDM

Les progrès du séquençage du génome humain ont récemment permis d'identifier les causes génétiques de l'OPDM. Tout d'abord, en 2019, une expansion anormale de répétitions CGG dans la région 5'UTR du gène *LRP12* a été identifiée chez 22 patients OPDM japonais (Ischiura et al., 2019). Le gène *LRP12* code pour la protéine reliée aux récepteurs de lipoprotéines de basse densité 12 (LRP12), une protéine portant les caractéristiques d'un récepteur membranaire et qui serait impliquer dans la transduction des signaux par l'endocytose de molécules lipophiles (Battle et al., 2003). La déplétion de LRP12 lors du développement conduit à un retard de la migration neuronale, à une plus faible arborisation neuronale et à des altérations de la lamination corticale (Schneider et al., 2011 ; Grotte et al., 2016). Cependant, il est à noter que cette expansion dans le gène *LRP12* n'a été retrouvée que dans ~40% des cas d'OPDM, suggérant l'existence d'autres causes génétiques responsables de cette maladie.

Effectivement, en 2020, Deng et ses collaborateurs ont montré qu'une expansion anormale de répétitions CGG dans le 5'UTR du gène *GIPC1* était associée à l'OPDM chez 12 patients chinois (Deng et al., 2020). Cette découverte a ensuite été confirmée par une seconde étude réalisée sur une cohorte plus grande de patients atteints d'OPDM (Xi et al., 2021). La protéine GIPC1 contient un domaine PDZ et interagit avec la Myosine-6 qui est nécessaire au recrutement des vésicules d'endocytose non-coatées (Naccache et al., 2006).

Enfin, une expansion de répétitions CGG a été très récemment identifiée dans le 5'UTR du gène *NOTCH2NLC* chez des patients atteints d'OPDM (Yu et al., 2021).

Quel que soit le gène hôte, un nombre de répétitions CGG inférieur à 30 a été observé chez les individus non atteints, tandis que les individus atteints d'OPDM présentaient une expansion de répétitions comprise entre 60 et 200 (Xi et al., 2021). À la suite de ces découvertes, l'OPDM a donc été divisée en 3 sous-types : l'OPDM1 due à une expansion de répétitions CGG dans le 5'UTR du gène *LRP12*, l'OPDM2 causée par une expansion de répétitions CGG dans le 5'UTR du gène *GIPC1*, et désormais l'OPDM3 due à une expansion de répétitions CGG dans le 5'UTR du gène *NOTCH2NLC*. Il est à noter qu'une étude publiée sur MedRxiv rapporte une expansion de répétitions CGG dans le 5'UTR du gène *RILPL1* chez des patients atteints d'OPDM (Yang et al., 2021). Ces 3 (voire 4) sous-types sont cliniquement et histologiquement identiques, ils sont causés par des mutations similaires en taille et en séquence (expansions de 60 à 200 répétitions CGG) mais localisées dans des gènes différents. Ces données suggèrent donc que

les expansions de répétitions CGG jouent un rôle essentiel dans la pathogénèse de l'OPDM, indépendamment des gènes dans lesquels ces répétitions sont situées.

Enfin, dans l'OPDM3, l'expansion de répétitions CGG retrouvée dans la région 5'UTR du gène *NOTCH2NLC* est identique à celle observée dans NIID (cf Introduction 2.1.3). Ceci suggère qu'une même mutation peut causer 2 maladies différentes, l'OPDM et NIID, soulignant un continuum entre ces 2 pathologies neuromusculaires. L'hypothèse d'un continuum est renforcée par la découverte d'un autre syndrome, la myopathie oculo-pharyngée avec leucodystrophie (OPML), affectant à la fois les muscles et le système nerveux central qui est décrite ci-dessous.

Mon travail de thèse a tenté de déterminer par quel mécanisme ces expansions de répétitions CGG, notamment dans les régions 5'UTR des gènes *GIPC1* et *NOTCH2NLC*, peuvent se révéler toxiques pour les muscles.

2.3. La myopathie oculo-pharyngée avec leucodystrophie (OPML)

Actuellement dans le monde, 7 individus issus d'une même famille ont été décrit comme développant une pathologie musculaire autosomale dominante est caractérisée par un ptosis, des mouvements des yeux réduits, une dysphagie et une dysarthrie ainsi que des faiblesses musculaires. De plus, certains patients présentent des symptômes neurodégénératifs tels qu'une ataxie et des tremblements (Ishiura et al., 2019). Cette maladie a donc été nommée myopathie oculo-pharyngée avec leucodystrophie (OPML) en référence aux symptômes musculaires et neurologiques.

Tout comme NIID et FXTAS, des IRM réalisées sur les patients OPML révèlent une hyper-intensité en T2 de la matière blanche (leucodystrophie) ainsi qu'une atrophie du cerveau (Figure 25). Toutefois, les analyses histopathologiques sur des biopsies de muscle de 2 patients n'a pas révélé la présence d'inclusions intranucléaires (Ishiura et al., 2019).

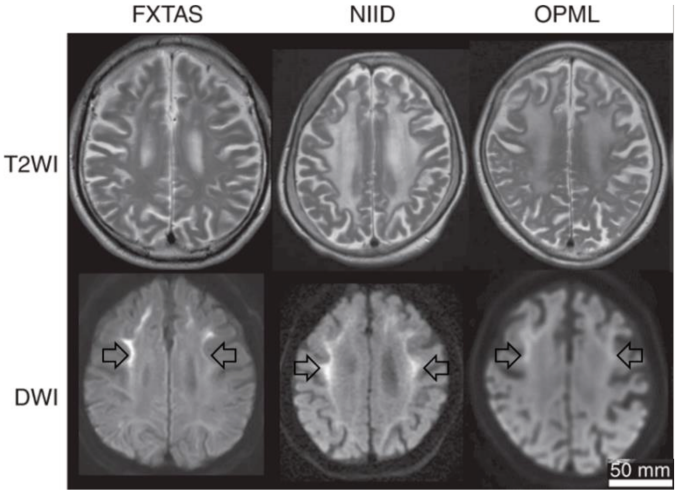


Figure 25. IRM de cerveaux de patients FXTAS, NIID et OPML.
IRM d'un patient FXTAS (à gauche), NIID (au centre) et OPML (à droite) révèle en T2WI des lésions de la matière blanche, notamment chez les patients NIID et OPML. Les images en DWI montrent une hyper-intensité en T2 des jonctions cortico-médullaires. Images de Ishiura et al., 2019.

De façon intéressante, une expansion d'une centaine de répétitions de tri-nucléotides CGG a été identifiée au niveau du locus *642361* chez ces patients OPML, alors que les individus sains présentent entre 3 et 16 répétitions CGG (Ishiura et al., 2019). Toutefois, il reste à déterminer comment ces expansions de répétitions CGG sont toxiques.

Enfin, pour plus de détails sur ces différentes maladies à expansion de répétitions CGG dans les régions 5'UTR, une revue est présentée page suivante.

Review

Trinucleotide CGG repeat diseases: an expanding field of polyglycine proteins?

Manon Boivin[1], **Nicolas Charlet-Berguerand**[1#]

[1]Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), INSERM U 1258, CNRS UMR 7104, University of Strasbourg, F-67404 Illkirch, France.

[#]Correspondence to Nicolas Charlet-Berguerand, Department of Translational Medicine and neurogenetics, IGBMC, 1 rue Laurent Fries, 67400 Illkirch, France. Phone: +33 388 653 309, e-mail: ncharlet@igbmc.fr

Running title

polyG diseases

Keywords

Microsatellite, trinucleotide repeat disorder, protein aggregates, RAN translation, neurodegeneration.

ABSTRACT

Microsatellites are repeated DNA sequences of 3 to 6 nucleotides highly variable in size and sequence and that have important roles in genomes regulation and evolution. However, expansion of a subset of these microsatellites over a threshold size is responsible of more than 50 human genetic diseases. Interestingly, some of these disorders are caused by expansions of similar sequences, sizes and localizations and present striking similarities in clinical manifestations and histopathological features, which suggest a common mechanism of disease. Notably, five identical CGG repeat expansions, but located in different genes, are the genetic causes of fragile X-associated tremor/ataxia syndrome (FXTAS), neuronal intranuclear inclusion disease (NIID), oculopharyngodistal myopathy type 1 to 3 (OPDM1-3) and oculopharyngeal myopathy with leukoencephalopathy (OPML), which are neuromuscular and neurodegenerative syndromes with overlapping symptoms and similar histopathological features, including the presence of characteristic eosinophilic ubiquitin-positive intranuclear inclusions. In this review we summarize recent finding in NIID and FXTAS, where the causing CGG expansions were found to be embedded within small upstream ORFs (uORFs), resulting in their translation into novel proteins containing a stretch of polyglycine (polyG). Importantly, expression of these polyG proteins is toxic in animal models and is sufficient to reproduce the formation of ubiquitin-positive intranuclear inclusions. These data suggest the existence of a novel class of human genetic pathology, the polyG diseases, and question whether a similar mechanism may exist in other diseases, notably OPDM and OPML.

Introduction

DNA short tandem repeats (STR), also known as microsatellites, are short sequences of 3 to 6 nucleotides repeated multiple time so that they occupy as much as 2 to 3 % of the human genome. These microsatellites are highly variable in size and sequence and have important roles in genomes regulation and evolution. However, expansion of a subset of these microsatellite sequences over a threshold size can also be the leading cause of human genetic diseases, and 2021 celebrates the 30th anniversary of the discovery of the two first pathogenic trinucleotide expansions, namely CGG and CAG repeats located in the fragile X mental retardation (*FMR1*) and androgen receptor (AR) genes and that cause fragile X syndrome (FXS) and spinal and bulbar muscular atrophy (SBMA), respectively (Fu et al., 1991; Oberlé et al., 1991; Verkerk et al., 1991; Kremer et al., 1991; La Spada et al., 1991). Since then, expansion of tri-, tetra-, penta- and hexa-nucleotide sequences were identified in more than 50 neuro developmental, neuromuscular and neuro degenerative genetic disorders (review in Malik et al., 2021; Depienne and Mandel 2021; table 1). As a consequence of their inherent variability, these microsatellite mutations are instable and dynamic with expansion lengths changing among populations, generation and individuals, and even according to tissues, cells and/or age. Thus, in a subgroup of microsatellite diseases where clinical manifestations correlate with expansion sizes, symptoms may greatly diverge between individuals and/or between generations, resulting in atypical heritability such as anticipation (increased disease severity and/or decreased age of onset with increased repeat size among generations) or its opposite mechanism, contraction where the expansion may shrink below its pathogenic threshold size from one generation to another. Moreover, instability at a given microsatellite locus is a rare event, which may result in a profound genetic founder effect, restricting observation of some of these mutations to specific human populations. For example, the G4C2 repeat expansion in the *C9ORF72* gene, which causes amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD), is amply observed in North Europe and North America but rare in other populations; while a CGG repeat expansion in the *NOTCH2NLC* gene that causes the neuronal intranuclear inclusion disease (NIID) is increasingly reported in Asian populations, but rare in Europe. These microsatellite expansions can be located in either non-transcribed and non-coding sequences, transcribed but untranslated regions, or within translated sequences, and thirty years of researches have now unveiled that these expansions are pathogenic by both loss and/or gain of function mechanisms at the DNA, RNA and protein levels (review in Malik et al., 2021; Depienne and Mandel 2021; table 1).

First, repeat expansions located within or close to promoters can promote DNA epigenetic changes that inhibit transcription, resulting in haploinsufficiency of the allele carrying the repeats. As indication that these expansions are pathogenic through a loss of function mechanism, identical clinical manifestations can be observed in individuals carrying genetic deletions and/or loss of function coding mutations in the gene hosting these microsatellites. Archetype of microsatellite expansions shutting down gene expression are the GGC, CGG and GAA repeat expansions located within the promoter, 5'UTR and first intron of the *XYLT1*, *FMR1* and frataxin genes and associated with Baratela-Scott syndrome (BBS), fragile X syndrome (FXS) and Friedreich ataxia, respectively (Fu et al., 1991; Oberlé et al., 1991; Verkerk et al., 1991; Kremer et al., 1991; Campuzano et al. 1996; laCroix et al., 2019; table 1).

Second, these expansions can be pathogenic at the RNA level through a toxic RNA gain of function mechanism, which depend of the expression, but is relatively independent of the protein functions, of the gene hosting the repeats. In these disorders, large repeat expansions are transcribed into pathogenic RNA that accumulate into nuclear RNA foci, which recruit and consequently alter the localization and function of specific RNA binding proteins, ultimately resulting in multiple RNA processing defects that are responsible of the diverse clinical manifestations observed in these diseases. This RNA gain of function mechanism is best exemplified by the titration of the muscleblind (MBNL) RNA binding proteins by expanded CUG or CCUG RNA repeats in myotonic dystrophy type 1 and 2 (DM1 & 2), as well as in Fuchs endothelial corneal dystrophy (FECD), which are caused by large expansions of CTG, CCTG and CTG repeats embedded within the 3'UTR of *DMPK* or within introns of the *CNBP* and *TCF4* genes, respectively (Brook et al. 1992; Fu et al. 1992; Harley et al. 1992; Mahadevan et al. 1992; Liquori et al., 2001; Wieben et al., 2012; Mootha et al., 2014; Miller et al., 2000; Fardaei et al., 2002; Wieben et al., 2017; Rong et al., 2019; review in Wheeler and Thornton, 2007; Morriss and Cooper, 2017; Sznajder and Swanson, 2019; table 1).

Third, these repeat expansions can be translated into pathogenic proteins containing a stretch of repeated amino acids, resulting in a toxic gain of function, a loss of function, a dominant negative effect and/or a mix of these mechanisms for the protein hosting the expansion. Importantly, translation of these repeat expansions can occur through two main mechanisms. The first one is based on classical translation initiation at a canonical AUG, or alternatively at a near-cognate (CUG, GUG, UUG or ACG), start codon, and thus results in expression of a pathogenic protein encoded by one predominant coding frame. The second mechanism, named repeat-associated non-AUG (RAN) translation, uses unconventional translation initiation that starts directly within the repeat expansion, resulting in expression of diverse proteins encoded in the three possible frames (Zu et al., 2011; review in Gao et al., 2017; Cleary et al., 2018; Nguyen et al., 2019). Archetype of this protein gain of function mechanism is the polyglutamine (polyQ) group of diseases, where expansions of CAG repeats, embedded within the main ORF of various genes (Huntingtin, Androgen receptor, Atrophin-1, ATXN1 to 3, 6, 7 and 17) (review in Lieberman et al., 2019), or alternatively within small ATG-driven ORFs located in ill-defined transcribed genetic regions (*JPH3AS*, *ATXN8*) (Moseley et al., 2006; Wilburn et al., 2011), are translated in

Proposed mechanism	Disease	Gene	Localization	Repeat	Normal size	Pathogenic size	reference
LOF	BSS	*XYLT1*	Promoter	CGG	9 - 20	120 - 800	LaCroix et al. 2019
LOF	FXS	*FMR1*	5' UTR	CGG	5 - 50	> 200	Verkerk et al. 1991; Oberlé et al. 1991; Fu et al., 1991
LOF	FRAXE	*AFF2*	5' UTR	CCG	4 - 39	200 - 900	Knight et al. 1993
LOF	EPM1	*CSTB*	5' UTR	C4GC4GCG	2 - 3	30 - 75	Lalioto et al. 1997
LOF	GDPAG	*GLS*	5' UTR	GCA	8 - 16	680 - 1400	Van Kuilenburg et al. 2019
LOF	FRDA	*FXN*	Intron	GAA	5 - 34	65 - 1300	Campuzano et al. 1996
LOF	XDP	*TAF1*	Intron	C3TCT	absent	30 - 55	Bragg et al. 2017
polyAla	SPD1	*HOXD13*	Exon	GCG	15	24	Akarsu et al. 1996
polyAla	BCCD	*RUNX2*	Exon	GCN	17	27	Mundlos et al. 1997
polyAla	HFGS	*HOXA13*	Exon	GCN	12 - 18	18 - 30	Goodman et al. 2000
polyAla	BPES	*FOXL2*	Exon	GCN	14	19 - 24	De Baere et al. 2001
polyAla	HPE5	*ZIC2*	Exon	GCN	15	25	Brown et al. 2001
polyAla	EIEE1	*ARX*	Exon	GCN	12 - 16	20 - 23	Stromme et al. 2002
polyAla	MRGH	*SOX3*	Exon	GCN	15	26	Laumonnier et al. 2002
polyAla	CCHS	*PHOX2B*	Exon	GCN	20	25 - 29	Amiel et al. 2003
polyAla	OPMD	*PABPN1*	Exon	GCG	6 - 10	11- 18	Brais et al. 1998
polyQ	SBMA	*AR*	Exon	CAG	9 - 36	38 - 68	La Spada et al. 1991
polyQ	DRPLA	*ATN1*	Exon	CAG	3 - 35	48 - 93	Koide et al. 1994; Nagafuchi et al. 1994
polyQ	HD	*HTT*	Exon	CAG	6 - 35	36 - 200	Huntington's Collaborative Group 1993
polyQ	HDL2	*JPH3 AS*	Exon	CAG	6 - 28	41 - 58	Margolis et al. 2001
polyQ	SCA1	*ATXN1*	Exon	CAG	6 - 38	39 - 88	Orr et al. 1993
polyQ	SCA2	*ATXN2*	Exon	CAG	13 - 31	32 - 500	Pulst et al. 1996
polyQ	SCA3	*ATXN3*	Exon	CAG	12 - 44	55 - 87	Kawaguchi et al. 1994
polyQ	SCA6	*CACNA1A*	Exon	CAG	4 - 18	20 - 33	Zhuchenko et al. 1997
polyQ	SCA7	*ATXN7*	Exon	CAG	4 - 33	37 - 460	Lindblad et al. 1996
polyQ	SCA8	*ATXN8*	Exon	CAG	15 - 50	74 - 250	Koob et al. 1999
polyQ	SCA17	*TBP*	Exon	CAG	25 - 40	43 - 66	Koide et al. 1999
?	SCA12	*PPP2R2B*	5' UTR	CAG	4 - 32	43 - 78	Holmes et al. 1999
polyGly	FXTAS	*FMR1*	5' UTR	CGG	5 - 50	55 - 200	Hagerman et al. 2001
polyGly	NIID	*NOTCH2NLC*	5' UTR	CGG	7 - 60	60 - 200	Ishiura et al. 2019; Sone et al. 2019; Tian et al. 2019; Deng et al., 2019
?	FXPOI	*FMR1*	5' UTR	CGG	5 - 50	55 - 200	Conway et al. 1998
?	OPML	*LOC642361*	LncRNA	CGG	3 - 16	50 - 60	Ishiura et al. 2019
?	OPDM1	*LRP12*	5' UTR	CGG	13 - 45	80 - 130	Ishiura et al. 2019
?	OPDM2	*GIPC1*	5' UTR	CGG	12 - 32	70 - 120	Deng et al. 2020
?	OPDM3	*NOTCH2NLC*	5' UTR	CGG	7 - 60	60 - 200	Yu et al. 2021
RAN	ALS/FTD	*C9ORF72*	Intron	G4C2	3 - 25	> 30	Dejesus et al. 2011; Renton et al. 2011
RAN	SCA36	*NOP56*	Intron	G3C2T	5 - 14	650 - 2500	Kobayashi et al. 2011
RAN	SCA31	*BEAN1*	Intron	G2A2T	variable	110 - 760	Sato et al. 2009
?	CANVAS	*RFC1*	Intron	G3A2	variable	400 - 2000	Cortese et al. 2019; Rafehi et al. 2019
RNA	DM1	*DMPK*	3' UTR	CTG	5 - 37	50 - 10 000	Mahadevan et al. 1992; Brook et al., 1992; Fu et al., 1992
RNA	DM2	*CNBP*	Intron	CCTG	11 - 30	50 - 11 000	Liquori et al. 2001
RNA	FECD3	*TCF4*	Intron	CTG	5 - 31	> 50	Mootha et al. 2014
?	FAME1	*SAMD12*	Intron	TTTCA	absent	440 - 3680	Ishiura et al. 2018
?	FAME2	*STARD7*	Intron	TTTCA	absent	> 660 - 730	Corbett et al. 2019
?	FAME3	*MARCHF6*	Intron	TTTCA	absent	> 660 - 2800	Florian et al. 2019
?	FAME4	*YEATS2*	Intron	TTTCA	absent	> 500	Yeetong et al. 2019
?	FAME6	*TNRC6A*	Intron	TTTCA	absent	> 400	Ishiura et al. 2018
?	FAME7	*RAPGEF2*	Intron	TTTCA	absent	> 500	Ishiura et al. 2018
?	SCA10	*ATXN10*	Intron	TTCTA	10 - 32	280 - 4500	Matsuura et al. 2000
?	SCA37	*DAB1*	Intron	TTTCA	absent	31 -75	Seixas et al. 2018

Table 1. Repeat expansion diseases, sorted by their proposed pathogenic mechanism.
LOF, loss of function mechanism; polyAla, polyalanine; polyGly, polyglycine; polyQ, polyglutamine; RAN, repeat non-ATG translation; ALS, amyotrophic lateral sclerosis; BCCD, brachydactyly and cleidocranial dysplasia; BPES, blepharophimosis, ptosis and epicanthus inversus; BSS, Baratela-Scott syndrome; CANVAS, cerebellar ataxia, neuropathy and vestibular areflexia syndrome; CCHS, congenital central hypoventilation syndrome; DM1, myotonic dystrophy type 1; DM2, myotonic dystrophy type 2; DRPLA, dentatorubral-pallidoluysian atrophy; EIEE1, early infantile epileptic encephalopathy type 1; EPM1, progressive myoclonus epilepsy type 1; FAME, familial adult myoclonic epilepsy; FECD3, Fuchs endothelial corneal dystrophy type 3; FRAXE, fragile XE syndrome; FRDA, Friedreich ataxia; FTD, frontotemporal dementia/; FXPOI, Fragile X-associated premature ovarian infertility; FXS, fragile X syndrome; FXTAS, fragile X-associated tremor ataxia syndrome; GDPAG, global developmental delay, progressive ataxia and elevated glutamine; HD, Huntington disease; HDL2, Huntington disease-like 2; HFGS, hand-foot-genital syndrome; HPE5, holoprosencephaly type 5; MRGH, mental retardation with isolated growth hormone deficiency; NIID, neuronal intranuclear inclusion disease; OPDM, oculopharyngodistal myopathy type; OPMD, oculopharyngeal muscular dystrophy; OPML, oculopharyngeal myopathy with leukoencephalopathy; SBMA, spinal and bulbar muscular atrophy; SPD1, synpolydactyly type 1; SCA, spinocerebellar ataxia; XDP, X-linked dystonia parkinsonism.

polyglutamine-containing proteins that form inclusions and are toxic for neuronal cells, causing several neurodegenerative diseases with similar histopathological characteristic and related symptoms (Huntington's disease and Huntington disease-like 2, spinal-bulbar muscular atrophy, dentatorubral-pallidoluysian atrophy, spinocerebellar ataxia 1 to 3, 6 to 8 and 17; table 1).This example illustrates that repeat expansions with similar sequences, sizes and genetic localizations can be pathogenic by a common molecular mechanism, ultimately resulting in a group of diseases with related clinical symptoms and remarkably similar histopathological features. In that aspect, progress in long-read and whole human genome sequencing has recently unveiled a dozen of novel pathogenic repeat expansions with striking similarities in sequence and localization and that can be divided in two groups (review in Ishiura and Tsuji, 2020). First, six identical expansions of ATTTC repeats that are all localized within introns, but of various gene, are responsible of familial adult myoclonic epilepsy (FAME) 1 to 4, 6 and 7 (Ishiura et al., 2018; Corbett et al., 2019; Florian et al., 2019; Yeetong et al., 2019). Second and topic of this review, several identical CGG repeat expansions, embedded within the 5'UTR of different genes or in a long non coding RNA (lncRNA), were recently identified as the genetic causes of fragile X-associated tremor/ataxia syndrome (FXTAS), neuronal intranuclear inclusion disease (NIID), oculopharyngodistal myopathy type 1 to 3 (OPDM) and oculopharyngeal myopathy with leukoencephalopathy (OPML) (Hagerman et al., 2001; Ishiura et al., 2019; Sone et al., 2019; Deng et al., 2019; Tian et al., 2019; Deng et al., 2020; Xi et al., 2021; table 1 and figure 1). In addition to their similar genetic cause, FXTAS, NIID, OPDM and OPML present overlapping clinical manifestations and similar histopathological features, including the presence of characteristic eosinophilic ubiquitin-positive nuclear inclusions (NIs), suggesting that these diseases belong to a continuum of neuromuscular and neurodegenerative diseases potentially caused by a common molecular mechanism (figure 1).

Fragile X-associated tremor/ataxia syndrome

Fragile X-associated tremor/ataxia syndrome (FXTAS, OMIM: 300623) is a late-onset X-linked neurodegenerative disorder originally described in great-fathers of infants with fragile X (Hagerman et al., 2001). Due to its X-linkage, FXTAS mostly affect males, while females are protected from overt neurodegeneration by random X-chromosome inactivation of the mutant allele, but are nonetheless at risk of developing another disease, fragile X-associated primary ovarian insufficiency (FXPOI; Conway et al., 1998). FXTAS is characterized by variable progressive intention tremor, gait ataxia and dementia, frequently accompanied by progressive cognitive decline, parkinsonism, peripheral neuropathy and autonomic dysfunctions (Jacquemont et al., 2003). Neuroradiological features of FXTAS include mild brain atrophy and white matter lesions with T2 hyperintensities in the middle cerebellar peduncle, as well as in the splenium of the corpus callosum, in the pons, insula and periventricular white matter. Furthermore, FXTAS is characterized by the presence of large eosinophilic ubiquitin-, p62- and sumo-positive intranuclear inclusions in both neurons and astrocytes across the nervous system, as well as in non-nervous tissues (Greco et al., 2006; Hunsaker et al., 2011; Ariza et al., 2016). However, these inclusions proved to be negative for polyQ or any other common protein inclusion markers of neurodegenerative disorders (Synuclein, Tau, ß-amyloid, TDP43, FUS, etc.) and they origin remained mysterious until recently (Todd et al., 2013).

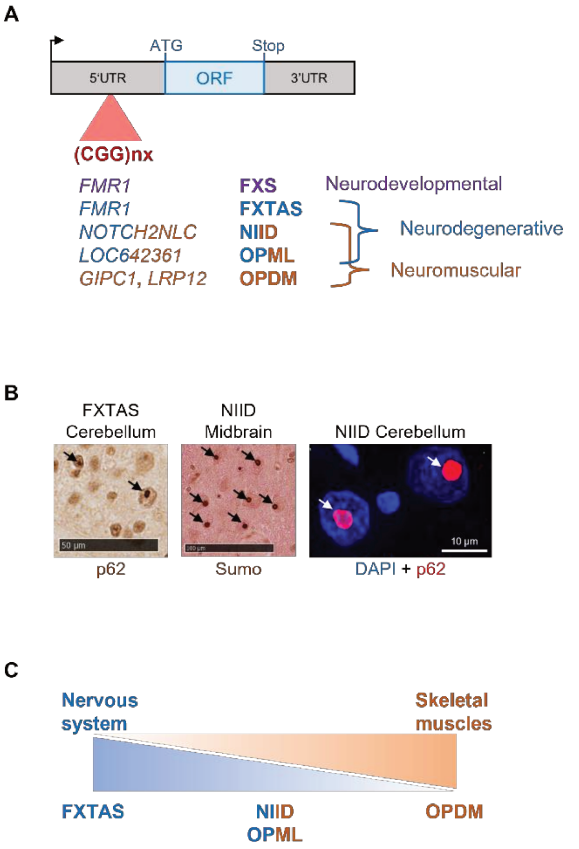


Figure 1. CGG repeat expansions cause a spectrum of disease.

(**A**) Identical CGG repeat expansions embedded within the 5'UTR of different genes cause various neurodevelopmental, neuromuscular and neurodegenerative disorders. (**B**) Brain sections of individuals with FXTAS or NIID show identical p62- or sumo-positive intranuclear inclusions. (**C**) FXTAS, NIID, OPML and OPDM may belong to a continuum of neuromuscular and neurodegenerative disorders.

FXTAS is caused by a limited expansion (called premutation) of 60-70 to ~200 CGG repeats localized in the 5'UTR of the *FMR1* gene (Hagerman et al., 2001). Interestingly, an identical but longer expansion (>200 CGG repeats, named full mutation) causes the neurodevelopmental fragile X mental retardation syndrome, while control individuals carry less than 50 CGG repeats, with 30 repeats being the most common allele. Importantly, the instable nature of this genetic mutation, notably during meiosis, explains the anticipation (also known as the Sherman paradox) observed in fragile X families where the clinical manifestations occur at progressively earlier age and/or with increasing severity in successive generations (Sherman et al., 1984; Sherman et al., 1985; Fu et al., 1991). At the molecular level, in fragile X syndrome, the

full mutation promotes DNA epigenetic changes leading to transcriptional silencing of the *FMR1* promoter and thus, loss of expression of the *FMR1*-encoded protein, FMRP, ultimately resulting in autism and intellectual disability in male. In female, FMRP haploinsufficiency is less marked due to expression by the second unmethylated *FMR1* allele. In contrast to the fragile X situation, in FXTAS the premutation allele, carrying 60 to 200 CGG repeats, enhances transcription, resulting in 2 to 8-fold increased levels of *FMR1* mRNA and near-normal to slightly reduced expression of FMRP (Tassone et al., 2000; Kenneson et al., 2001; Tassone et al., 2007). Importantly, the sole expression of expanded CGG repeats, embedded in the 5'UTR of *FMR1*, in cell or animal models is sufficient to cause neuronal cell dysfunctions and to induce formation of the typical FXTAS intranuclear inclusions (Willemsen et al., 2003; Jin et al., 2003; Arocena et al., 2005; Entezam et al., 2007; Hashem et al. 2009; Todd et al., 2013; Hukema et al., 2015). These pioneering studies indicate that FXTAS, in contrary to the fragile X syndrome, is not due to a reduce expression of FMRP and to a loss of function mechanism, but is more likely caused by an RNA and/or protein gain of function mechanism.

About a toxic RNA gain of function mechanism in FXTAS*, in vitro* assays revealed that multiple RNA binding proteins bind to CGG RNA repeats, including Pura, hnRNP A2, SAM68, CUGBP1, TDP43 and Drosha/DGCR8, among others (Iwahashi et al., 2006; Sofola et al., 2007; Jin et al., 2007; Sellier et al., 2010; Sellier et al., 2013; Cid-Samper et al., 2018). However, whether CGG expansions are toxic at the RNA level in FXTAS is unclear. It notably remains to clarify whether these CGG RNA repeats deplete a sufficient amount of these RNA binding proteins, so that they loss their mobility and functions below a pathological threshold. As a further note of caution, many studies, including ours, were performed with CGG repeats deleted of any natural *FMR1* sequence and that consequently accumulate artificially in nuclei in some cell models, potentially titrating RNA binding proteins in nuclear RNA foci (Sellier et al., 2010). However, expanded CGG repeats embedded in their natural *FMR1* 5'UTR are exported into the cytoplasm and thus may not accumulate sufficiently into the nucleus to drive an RNA gain of function toxicity mechanism (Sellier et al., 2017). Interestingly and notwithstanding their level of titration, these RNA binding proteins may also chaperone CGG RNA repeats to modulate their stability, localization and translation, resulting in diminished toxicity, thus illuminating potential therapeutic strategies for FXTAS (Sofola et al., 2007; Jin et al., 2007; Sellier et al., 2010; Qurashi et al., 2011; He et al., 2014; Malik et al., 2021).

Concerning a pathogenic protein gain of function in FXTAS, studies of cell and animal models coupled to mass spectrometry analyses revealed that the *FMR1* CGG expansion is translated into a novel protein, where each GGC triplet encodes for a glycine, resulting in a small polyglycine-containing protein (polyG), which was named FMRpolyG (Todd et al., 2013). Of interest, translation of the *FMR1* CGG repeats occurs in one principal frame (glycine) among the three possible (alanine, glycine and arginine), but in absence of any ATG start codon. Instead, translation initiation takes place at an ACG near-cognate codon located 36 nucleotides upstream of the CGG repeats and in the glycine frame (Kearse et al., 2016; Sellier et al., 2017, figure 2). Near-cognate initiation codons are codons differing from the cognate AUG start codon by one nucleotide, but that can still initiate translation through mispairing with the initiator methionine-tRNA. *In vitro* experiments and large-scale ribosome profiling revealed that predominantly four near-cognate start codons (CUG, GUG, UUG and ACG) are tolerated and can initiate translation, despite fidelity control, notably by the eIF1 protein (Clements et al., 1988; Kozac, 1989; Peabody, 1989, Ingolia et al., 2011; Lee et al., 2012; review in Kearse and wilusz, 2017). However, translation initiation at near-cognate codons is 5 to 10-fold less efficient compared to initiation at an AUG start codon, and is decisively dependent of surrounding secondary RNA structures and sequences, notably the Kozac sequence (G/ACCAUGG) that borders the start codon (Kozac et al., 1981; Kozac et al., 1990). Further experiments revealed that FMRpolyG translation follows a classical m7G cap-dependent ribosome scanning mechanism, and that the CGG repeat expansion is located within a small upstream ORF overlapping and negatively-regulating translation of the main *FMR1* ORF, FMRP (Kearse et al., 2016; Sellier et al., 2017; Rodriguez et al., 2020; figure 2). Upstream ORFs (uORFs) are short open reading frames located upstream of the main protein-encoding ORF, and are thus embedded in the 5'UTR sequence of gene. Whole genome analyses revealed that most mammalian genes contain uORFs, with their majority (>75%) initiating at near-cognate codons (Ingolia et al., 2011; Lee et al., 2012; Fields et al., 2015; Johnstone et al., 2016). As a consequence of the 5' to 3' ribosome scanning mechanism, a ribosome translating an uORF and dissociating at its stop codon may not re-initiate to a nearby downstream ORF, even initiated by a canonical AUG start codon. Thus, presence of upstream ORFs may negatively regulate, yet not systematically, translation of downstream ORFs. However, due to the low efficiency of translation initiation at near-cognate start codons, ribosome may bypass an uORF and instead initiate at the main downstream ORF start codon, a process known as leaky scanning translation (review in Hinnebusch et al., 2016; Kearse et al., 2017; Gao et al., 2017). As uORFs act predominantly as translation negative cis-regulatory RNA elements, they are rarely expressing functional proteins and the vast majority of uORFs encode for small and instable peptides that are hardly detectable. Consistent with this notion, in control condition with no repeat expansions (~30 CGG) the *FMR1* uORF encodes for a small protein of 83 amino acids (~7 kDa), hardly detectable without inhibition of the proteasomal degradation pathway (Sellier et al., 2017). In contrast, expansions over 60-70 CGG repeats in FXTAS lengthen this uORF, resulting in expression of a stable FMRpolyG protein (Todd et al., 2013; Kearse et al., 2016; Sellier et al., 2017, figure 2). Of interest, FMRpolyG with expansion over ~30 glycine repeats is prone to aggregation and expression of FMRpolyG in various cell and animal models leads to the formation of cytoplasmic and intranuclear aggregates, positives for ubiquitin and p62 (Todd et al., 2013; Hukema et al., 2015; Sellier et al., 2017; Derbis et al., 2018; Wenzel et al., 2019; Hoem et al., 2019). Furthermore, various mouse or rabbit antibodies

independently developed and directed against FMRpolyG revealed presence of this protein within the typical eosinophilic ubiquitin-positive intranuclear inclusions in cell and tissue sections of individuals with FXTAS (Todd et al., 2013; Buijsen et al., 2014; Sellier et al., 2017; Krans et al., 2019; Bonapace et al., 2019; Dijkstra et al., 2021). Of technical interest these antibodies are directed against the N- or C-terminal parts of FMRpolyG, but none are directly targeting the polyglycine stretch. Furthermore, due to its high propensity to aggregates, very little FMRpolyG is present in soluble brain extract or in fluids of individuals with FXTAS, tempering its relevance as an easily quantifiable biomarker for FXTAS (Ma et al., 2019; Holm et al., 2021). Finally, expression of FMRpolyG in cell and animal models is toxic, and drosophila or mice expressing FMRpolyG show locomotor alterations, neuronal cell loss and ultimately, premature death of these animals (Todd et al., 2013; Sellier et al., 2017). However, the mechanism of FMRpolyG toxicity is currently unclear (review in Boivin et al., 2018; Glineburg et al., 2018), with proposed alteration of the proteasome and degradation pathway (Oh et al., 2015), of the nuclear architecture and nucleocytoplasmic transport (Sellier et al., 2017), of mitochondrial functions (Gohel et al., 2019) and/or through binding to *FMR1* CGG RNA repeats (Asamitsu et al., 2021). Furthermore, it is unclear whether FMRpolyG toxicity is related to its propensity to form large cellular aggregates and/or to its localization and notably whether its import into nuclei is required for its pathogenicity. Similarly, whether FMRpolyG toxicity is mainly triggered by its polyglycine stretch, or whether bordering N- or C-terminal amino acid sequences contribute to FMRpolyG pathogenicity is currently ill-defined (Sellier et al., 2017).

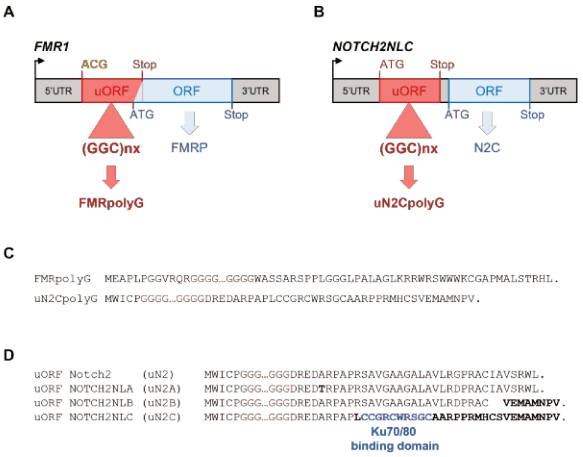


Figure 2. CGG repeat expansions are translated into polyG proteins in FXTAS and NIID.
(**A**) Scheme of *FMR1* indicating FMRpolyG upstream ORF and FMRP main ORF localization. (**A**) Scheme of *NOTCH2NLC* indicating uN2CpolyG upstream ORF and NOTCH2NLC (abbreviated N2C) main ORF localization. Initiation codons and stop codons are indicated in red for uORFs and in blue for the main ORFs. Near-cognate initiations codons are indicated in bold yellow. (**C**) Protein sequences of FMRpolyG and uN2CpolyG show no similitude beyond their polyglycine stretch. (D) Sequence of the putative uORFs embedded within *NOTCH2*, *NOTCHNLA*, B and C 5'UTRs. Variant amino acids are indicated in bold. The sequence required for uN2C to interact with KU70/KU80 is indicated in blue.

In parallel to the translation initiation of a CGG repeats-containing uORF at a near-cognate start codon resulting in expression of the toxic FMRpolyG protein, RAN translation where initiation occurs directly within the CGG expansion and in all three frames was proposed as a leading pathogenic mechanism in FXTAS (Todd et al., 2013). While modest to no translation of CGG repeats was found in the arginine frame, which is the frame of the downstream FMRP ORF, assays using highly-sensitive nanoluciferase detection methods revealed some translation of the expanded CGG repeats into the alanine frame, resulting in expression of a novel polyalanine containing protein that was named FMRpolyA (Todd et al., 2013; Oh et al., 2015; Kearse et al., 2016). However, independently developed mouse and rabbit antibodies directed against FMRpolyA failed to detect significant accumulation of this protein in cells and tissues of individuals with FXTAS, suggesting that FMRpolyA is not a major component of the intranuclear inclusions typical of that disease (Sellier et al., 2017; Krans et al., 2019). Furthermore, a mouse model expressing an expansion of CGG repeats but deleted of the *FMR1* 5'UTR sequence that contains the near-cognate start codon initiating FMRpolyG translation shows no inclusions and no locomotor or neurodegenerative phenotype. In contrast, mice expressing the same CGG expansion, but with its upstream natural *FMR1* near-cognate codon, show numerous FMRpolyG inclusions and neurodegeneration (Sellier et al., 2017). These animal models suggest that RAN translation initiating inside the CGG repeat track is not sufficiently efficient to drive inclusion formation and neuronal cell death, at least in the limited time frame of mouse longevity. Similarly, these models suggest that the sole expression of CGG repeats at the RNA level, and thus an RNA gain of function mechanism, is not sufficient to drive neuropathogenicity in these animals.

Overall, these results may explain how in a rare inherited neurodegenerative disease a microsatellite expansion, first annotated as "non-coding", is in reality translated into a novel protein, which contains a stetch of polyglycine, forms intranuclear inclusions and is toxic for neurons. These data clarify the origin of ubiquitin-positive intranuclear inclusions, a key histological feature observed in FXTAS, but they also support a novel model of pathogenicity in human diseases, where repeat expansions located in ill-defined upstream ORFs can be decoded into novel and toxic proteins through translation initiation at near-cognate start codons (Todd et al., 2013; Sellier et al., 2017). Interestingly, a similar mechanism was recently identified in amyotrophic lateral sclerosis and frontotemporal dementia, where an expansion of G4C2 repeats located in the first intron of the *C9ORF72* gene is translated into a polyglycine-alanine dipeptide repeat (polyGA DPR) protein through translation initiation at a CUG near-cognate start codon located 24 nucleotides upstream of the repeats (Green et al., 2017; Tabet et al., 2018; Sonobe et al., 2018; Almeida et al., 2019; Boivin et al., 2020; Sonobe et al., 2021; figure 3). It remains to determine whether these examples of translation initiation at near-cognate start codons are two isolated cases or reflect a more global mechanism of disease.

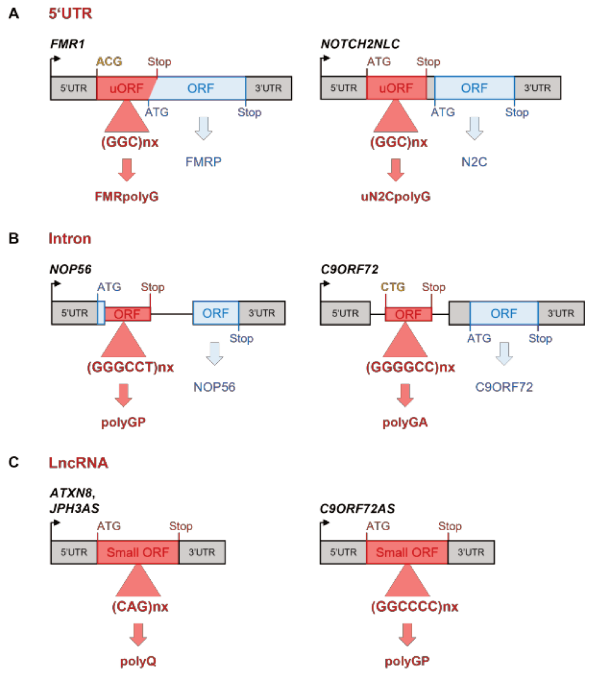


Figure 3. Repeat expansions located in "non-coding" regions are nonetheless translated.
(**A**) G2C repeat expansions embedded in the 5'-"untranslated" regions of the *FMR1* and *NOTCH2NLC* genes are translated into polyglycine-containing proteins in NIID and FXTAS. (**B**) G3C2T and G4C2 repeats embedded in the first and retained introns of the *NOP56* and *C9ORF72* genes are translated into poly(glycine-proline) and poly(glycine-alanine)-containing proteins in SCA36 and ALS/FTD, respectively. (**C**) CAG repeats and G2C4 repeats embedded in *ATXN8*, *JPH3AS* and *C9ORF72AS* long "non-coding" RNAs are translated into polyglutamine- or poly(glycine-proline)-containing proteins in SCA8, HDL2 and ALS/FTD, respectively. Initiation codons and stop codons of the ORFs containing the pathogenic expansion are indicated in red, while they are indicated in blue for the main ORFs. Near-cognate initiations codons are indicated in bold yellow.

Neuronal intranuclear inclusion disease

Neuronal intranuclear inclusion disease (NIID, OMIM: 603472), also known as intranuclear inclusion body disease (INIBD) or neuronal intranuclear hyaline inclusion disease (NIHID), is a rare autosomal dominant genetic disorder with variable age of onset, from infant to late adult (Lindenberg et al., 1968; Michaud and Gilbert, 1981; Munoz-Garcia and Ludwin, 1986; Takahashi-Fujigasaki, 2003; Sone et al., 2016). The clinical manifestations of NIID are highly heterogenous and generally comprise variable skeletal muscle weakness associated with variable dysfunctions of the central and peripheral nervous systems, including cerebellar ataxia, parkinsonism, cognitive decline, peripheral neuropathy and autonomic dysfunctions. Moreover, atypical NIID clinical manifestations are increasingly described, including leukoencephalopathy, essential tremor, multiple system atrophy, retinal changes and retinopathy, amyotrophic lateral sclerosis, Alzheimer or Parkinson disease as well as with various acute symptoms, including stroke-like episodes, migraine-like attack, epileptic seizures and/or encephalitic episodes (Takahashi-Fujigasaki, 2003; Sone et al., 2016; Takahashi-Fujigasaki et al., 2016; Okubo et al., 2019; Nakamura et al., 2020; Ng et al., 2020; Fang et al., 2020; Ma et al., 2020; Shi et al., 2020; Li et al., 2020; Yuan et al., 2020; Yang et al., 2021; Cao et al., 2021; review in Cao et al., 2021; Fan et al., 2021). Interestingly, brain MRI of individuals with NIID reveal abnormalities reminiscent of FXTAS with mild brain atrophy and white matter lesions, including T2-weighted hyperintensities in the middle cerebellar peduncle and high-intensity signals in the corticomedullary junction (Gelpi et al., 2017; Sugiyama et al., 2017; Ng et al., 2020). Furthermore, and as its name imply, NIID is characterized by the presence of large eosinophilic ubiquitin-, p62- and sumo-positive intranuclear inclusions. These aggregates are present in both neurons and glial cells in the central and peripheral nervous systems, as well as in various other tissues, including the skin, which observation can be determinant to confirm diagnosis of this multifaced syndrome (Patel et al., 1985; Kimber et al., 1998; Pountney et al., 2003; Mori et al., 2012; Liu et al., 2008; Sone et al., 2011; Chen et al., 2020; Zhang et al., 2021). Importantly, these intranuclear inclusions are virtually identical and indistinguishable to those observed in FXTAS (Gelpi et al., 2017; Lim et al., 2020; Toko et al., 2021).

The genetic cause of NIID was recently identified as a limited expansion of ~60 to 200-300 CGG repeats in the 5'UTR of the *NOTCH2NLC* gene (Sone et al., 2019; Ishiura et al., 2019; Deng et al., 2019; Tian et al., 2019). Of interest, an identical expansion of CGG repeats in *NOTCH2NLC* was also recently identified as the genetic cause of oculopharyngodistal myopathy type 3 (OPDM3) with variable neurological manifestations (Ogasawara et al., 2020; Yu et al., 2021). These studies highlight the diversity of NIID and OPDM clinical manifestations and the ensuing complexity to diagnose these syndromes, as well as the potential overlap between OPDM and NIID. Furthermore, *NOTCH2NLC* CGG expansion is mostly reported in individuals of Asian origin, but rarely observed in European NIID cases, despite clinical similitudes and the presence of identical intranuclear inclusions (Chen et al., 2020), suggesting a founder effect for this mutation, but also that NIID is genetically heterogenous with other mutations awaiting to be identified. The Notch 2 N-terminal like (*NOTCH2NL*) A, B and C genes are hominid-specific genes located on chromosome 1q21.1, which result from the duplication of the promoter region and N-terminal part (exons 1 to 5) of *Notch2*. As a consequence of this partial duplication, the NOTCH2NL proteins contain six epidermal growth factor (EGF)-like domains, but are deleted of the middle transmembrane and C-terminal cytoplasmic domains of Notch2. Moreover, these proteins also lack the N-terminal Notch2 peptide signal due to an intronic nucleotide variation changing *NOTCH2NLB* intron 1 splice acceptor by 8 nucleotides and modified ATG start site localization in *NOTCH2NLA* and *C*. Important to hominid brain size evolution, NOTCH2NL proteins regulate Notch signaling and expand human cortical progenitors. Furthermore, 1q21.1 genomic deletions or duplications potentially encompassing the *NOTCH2NLA* and B genes are associated with neurodevelopmental syndromes with microcephaly or macrocephaly, respectively (Fiddes et al., 2018; Suzuki et al., 2018). These seminal studies highlight the

importance of the NOTCH2NL proteins for neuronal progenitor proliferation and brain size expansion in human evolution, but question how a CGG repeat expansion embedded within the 5'UTR of *NOTCHNLC* can be pathogenic. First, a loss of function of the NOTCH2NLC (abbreviated N2C) protein is unlikely as the CGG repeats are located 140 nucleotides ahead of the ATG start site of this protein, and as *NOTCH2NLC* mRNA levels are found unaltered, or even increased, in blood, brain, fibroblasts and muscle samples of individuals with NIID or OPDM3 (Sone et al., 2019; Ishiura et al., 2019; Tian et al., 2019; Yu et al., 2021). Moreover, the NOTCHNLC protein is most likely expressed at very low level due to the presence of six ATG-driven upstream ORFs inhibiting its translation initiation, questioning the physiological expression and importance of this protein (Zhong et al., 2021). Second, an RNA gain of function mechanism has been considered for NIID, with some evidence of CGG RNA accumulation in RNA foci, however, mostly upon overexpression of CGG repeats in cell models and with unclear pathological consequences (Yu et al., 2021; Deng et al., 2021; Zhong et al., 2021). Third, translation of these CGG repeats into a potentially toxic protein has been recently observed in two independent reports (Boivin et al., 2021; Zhong et al., 2021; figure 2). Importantly, these studies revealed that the *NOTCH2NLC* CGG repeats are embedded into a small upstream ORF, which is translated from a canonical ATG start site located 15 nucleotides upstream of the CGG repeats. This uORF was hence named uN2C for upstream of the NOTCH2NLC ORF. As a results of this ATG translation initiation, each GGC triplet encodes a glycine, resulting in expression of a polyglycine-containing protein, which was named either uN2CpolyG (Boivin et al., 2021) or N2NLCpolyG (Zhong et al., 2021). Despite different names and independent identification, these two proteins are strictly identical (figure 2). As with the majority of upstream ORFs, in control condition the N2C uORF with no expansion (~12 CGG) encodes for a small protein (56 amino acids, ~6 kDa), which is unstable and hardly detectable without inhibiting degradation pathways (Boivin et al., 2021). In contrast, CGG repeat expansion increases the length of this uORF, resulting in expression of a stable polyglycine-containing protein in NIID. Importantly, mouse monoclonal antibodies directed against two different amino acid sequences, both located in the C-terminal part, of uN2CpolyG revealed presence of this protein in the typical ubiquitin- and p62-positive intranuclear inclusions in skin and brain sections of individuals with NIID (Boivin et al., 2021). These results were recently confirmed using an independently-developed rabbit polyclonal antibody directed against a stretch of 12 glycine followed by 11 amino acids from the uN2C sequence (Zhong et al., 2021). Furthermore, expression of this polyG protein in cell and animal models drives the formation of cytoplasmic and intranuclear inclusions, positive for ubiquitin, sumo and p62. Finally, expression of uN2CpolyG in mice is toxic, resulting in locomotor alterations, neuronal cell loss and reduced lifespan (Boivin et al., 2021). However, by which mechanism this polyG protein is toxic is unclear. Immunoprecipitation followed by mass spectrometry analyze indicate that the normal uN2C protein with a normal stretch of polyglycine (~12 repeats) interacts with the Ku70 and Ku80 proteins, which form a scaffold for the non-homologous end joining (NHEJ) repair of DNA double-strand breaks. Interestingly, expression of uN2C (with a control number of CGG repeats) boosts the repair of DNA double strand breaks (Boivin et al., 2021). These results indicate that the small uN2C protein is potentially a novel regulator of DNA damage response, suggesting that the *NOTCH2NL* genes may encode different proteins with complementary functions for human brain development. In this model, the *NOTCH2NLB* gene, encoding the NOTCH2NLB protein, would regulate Notch signaling to promote extensive neural progenitor proliferation (Fiddes et al., 2018; Suzuki et al., 2018), but at the cost of potentially generating multiple DNA damages. In contrast, the *NOTCH2NLC* gene contains six ATG-driven upstream ORFs that reduce expression of the NOTCH2NLC protein (Zhong et al., 2021), but the first of these uORFs encodes for a small uN2C protein that may protect neuronal cells from DNA damages by stimulating the NHEJ repair mechanism. Of interest, *NOTCH2*, *NOTCH2NLA* and *B* also contains 1, 5 and 2 potential ATG-driven uORFs, respectively, but with protein sequences different from uN2C and are thus unable to interacts with the Ku70 and 80 proteins (Boivin et al., 2021; figure 2). Note that alike *NOTCH2NLC*, *NOTCH2NLA* contains multiple ATG-driven uORFs that may reduce expression of the downstream NOTCH2NLA ORF, questioning the expression and relevance of this protein in physiological conditions.

In NIID, the polyglycine expansion alters the interaction of uN2C with the Ku proteins and reduces its DNA repair activity. However, this reduced function is likely compensated by the second *NOTCH2NLC* allele and is not sufficient to drive overt DNA repair alterations, as no clear evidence of DNA damage accumulation was observed in mice expressing the uN2CpolyG protein or in in brain sections of NIID individuals (Boivin et al., 2021). Interestingly, expression of this polyG protein promotes alterations of the nucleocytoplasmic transport (Zhong et al., 2021), which is reminiscent of the pathogenicity described for FMRpolyG in FXTAS (Sellier et al., 2017), as well as for DPR proteins expressed from the G4C2 repeats in the *C9ORF72* gene in ALS/FTD (Jovičić et al., 2016; Hayes et al., 2020). In this aspect, the molecular mechanisms by which the uN2CpolyG and FMRpolyG proteins may dysregulate nucleocytoplasmic traffic remain to be investigated. Furthermore, it also remains to determine whether RAN translation of the *NOTCH2NLC* repeats, with initiation within the CGG expansion and protein expression in all three frames, may occur and contribute to the pathogenicity in NIID.

Overall, these studies suggest the existence of a novel class of human genetic disorder, the polyG diseases, where expanded CGG repeats are embedded in small upstream ORFs and consequently, are translated into novel polyglycine-containing proteins that are toxic and forms ubiquitin-positive inclusions (Todd et al., 2013; Sellier et al., 2017; Boivin et al., 2021; Zhong et al., 2021). However, it remains to determine whether this group of disease is limited to NIID and FXTAS, or whether it may include other pathologies with related symptoms and/or similar histopathological features. Candidate pathologies may include mutations, yet to be

identified, in NIID cases negative for the *NOTCH2NLC* mutation, but that nevertheless present the typical clinical and histopathological features of that syndrome (Chen et al., 2020). Similarly, it remains to explore whether translation of expanded CGG repeats, yet to be identified, may underlie the presence of nuclear inclusions (NIs) of unknown origins (review in JM Woulfe, 2007), notably Marinesco bodies, which are eosinophilic ubiquitin-positive intranuclear inclusions found in pigmented neurons of the substantia nigra and locus ceruleus in aged human brain (Yuen and Baxter 1963; Odagiri et al., 2011). Finally, the recent identification of CGG repeat expansions in the OPDM and OPML neuromuscular disorders, which are characterized by the presence of eosinophilic ubiquitin-positive intranuclear inclusions, may potentially represent novel examples of polyG diseases.

Oculopharyngodistal myopathies

Oculopharyngodistal myopathy (OPDM, OMIM: 164310) is a rare adult-onset and slowly progressive autosomal dominant disease characterized by ptosis, external ophthalmoplegia, dysphagia and dysarthria, associated to facial and distal limb muscle weakness (Satoyoshi and Kinoshita 1977; Uyama et al., 1998; van der Sluijs et al., 2004; Lu et al., 2008; Durmus et al., 2011; Minami et al., 2011; Zhao et al., 2015). Serum creatine kinase levels are usually normal or mildly increased in individuals with OPDM, and skeletal muscles shows moderate fibrosis and small angular fibers. Besides these non-specific myopathic changes, OPDM histopathology is characterized by the presence of cytoplasmic rimmed vacuoles (RVs) and typical eosinophilic nuclear inclusions (NIs), which are both p62- and ubiquitin-positive (Zhao et al., 2015; Deng et al., 2020; Saito et al., 2020; Xi et al., 2021; Matsubara et al., 2021; Kumutpongpanich et al., 2021). Of interest, these intranuclear inclusions are also observed in skin sections of individuals with OPDM cases (Ogasawara et al., 2021), and are reminiscent of the typical eosinophilic ubiquitin-positive inclusions observed in FXTAS and NIID. Importantly, the mutation causing OPDM was recently identified as an expansion of ~70 to 200-300 CGG repeats, but located within the 5'UTR of two different genes, *LRP12* and *GIPC1* (Ishiura et al., 2019; Deng et al., 2020; Kumutpongpanich et al., 2021; Xi et al., 2021). Furthermore, an expansion of CGG repeats in the 5'UTR of the *NOTCH2NLC* gene was identified in OPDM cases with variable neurological manifestations (Ogasawara et al., 2020; Yu et al., 2021), and an expansion of CGG repeats within the 5'UTR of the *RILPL1* gene was recently reported in medRxiv in affected individuals of a large Chinese family with OPDM (Yang et al., 2021). Thus, OPDM is now distinguished by its genetic cause with oculopharyngodistal myopathy type 1 (OPDM1) caused by CGG repeat expansions in *LRP12*, OPDM type 2 associated with CGG expansions in *GIPC1*, OPDM3/NIID1 caused by CGG repeats in *NOTCH2NLC* and OPDM4 potentially associated to a CGG repeat expansion in *RILPL1*. Of further interest, CGG repeats in *LRP12* are mostly found in Japan cases, while CGG expansions in *GIPC1* are mostly reported in Chinese individuals, suggesting a founder effect for these mutations.

The mechanism of toxicity in OPDM is yet to be identified but a loss of function is unlikely as the CGG repeat expansions are located ahead of the ATG start sites of the GIPC1, LRP12, NOTCH2NLC and RILPL1 proteins. Furthermore, expression of the GIPC1 protein is unaltered in tissue samples from individuals with OPDM2 (Deng et al., 2020). A mechanism of CGG RNA gain of function and/or of translation into pathogenic proteins remains to be explored, but it is noteworthy that oculopharyngeal muscular dystrophy, a neuromuscular disease with clinical manifestations related to OPDM, is also caused by a GC-rich repeat expansion. Oculopharyngeal muscular dystrophy (OPMD, OMIM: 164300) is a late-onset and slowly progressive autosomal dominant neuromuscular disease characterized by proximal muscle weakness, ptosis, and swallowing difficulty. OPMD is caused by extension of 1 to 8 repeats of a GCG trinucleotide stretch located in the main ORF of the *PABPN1* gene, which encodes for the polyadenylate-binding protein 2 (PABP2). The N-terminal part of this protein contains a short polyalanine tract (Met-(Ala)10x-Gly-(Ala)2x…), and in OPMD a small GCG expansion adding 1 to 8 alanine, or a missense mutation of the +12 glycine into an alanine resulting in a contiguous stretch of 13 alanine, leads to a modified PABP2 protein that interfere with the cellular traffic of polyadenylated RNA (Brais et al., 1998; Robinson et al., 2006). This is reminiscent of other polyalanine diseases (table 1), which are systematically caused by GCN repeat expansions of relatively small size (<35 repeats) that are proposed to cause loss of function and/or dominant negative effect of the transcription factors hosting these elongated polyalanine stretches, ultimately leading to developmental disorders. Of interest, similar developmental aetiologies are observed in individuals with deletion and/or loss of function mutations in the genes coding for these transcription factors, strengthening the hypothesis that these polyalanine tracts results in a loss of function mechanism (review in Albrecht and Mundlos, 2005; Shoubridge and Geez, 2012). In contrast, the CGG repeat expansions in OPDM1 to 4 are longer and localized in the 5'UTR, and not in the coding sequences of the GIPC1, LRP12, NOTCH2NLC, and RILPL1 proteins. Thus, expansions of ~70 to 200-300 CGG repeats in OPDM are likely to be pathogenic by a different mechanism than the small (11 to 18) polyalanine tract that alters PABP2 protein functions in OPMD.

Oculopharyngodistal myopathy with leukoencephalopathy: the missing piece of a disease continuum?

Overall, it is striking to note that the OPDM, FXTAS and NIID diseases share overlapping clinical manifestations and identical histopathological features, with the presence of characteristic eosinophilic ubiquitin-positive nuclear inclusions (NIs). Furthermore, these diseases are caused by similar genetic causes, namely microsatellite expansions of identical sequence and size (60-70 to 200-300 CGG repeats). These data suggest that these disorders may belong to a continuum of neuromuscular and neurodegenerative diseases (figure 1). Importantly, this hypothesis is reinforced by the recent discovery that both NIID and OPDM3 can be caused by an identical CGG repeat expansion in

NOTCH2NLC (Sone et al., 2019; Ishiura et al., 2019; Deng et al., 2019; Ogasawara et al., 2020; Yu et al., 2021), as well as by the identification of a novel and rare clinical entity with clinical features of both NIID and OPDM, and that was consequently named oculopharyngodistal myopathy with leukoencephalopathy (OPML, OMIM: 618637) (Ishiura et al., 2019). OPML is characterized by muscle dystrophy and weakness, especially of the ocular, pharyngeal, facial and distal limb muscles, associated with leukoencephalopathy, brain atrophy and T2 hyperintensity signals in the white matter and sign of tremor and ataxia in some individuals. Interestingly, OPML is caused by an expansion of CGG repeats located within the ill-defined *LOC642361/NUTM2B-AS1* genetic region, which is transcribed but predicted as non-coding (Ishiura et al., 2019). Thus OPML, alike FXTAS, OPDM and NIID, is caused by a CGG repeat expansion embedded in a genetic region predicted as non-coding, questioning the pathogenic mechanisms underlying these diseases. In that aspect, it was recently unveiled that CGG repeat expansions located in the 5'UTR of *FMR1* or *NOTCH2NLC*, or G4C2 and G3C2T repeat expansions localized within the first intron of the *C9ORF72* and *NOP56* genes, respectively, are translated into toxic proteins through canonical translation initiation at canonical AUG, or near-cognate, start codons (Todd et al., 2013; Sellier et al., 2017; Green et al., 2017; Tabet et al., 2018; Sonobe et al., 2018; Almeida et al., 2019; Boivin et al., 2020; McEachin et al., 2020; Sonobe et al., 2021; Boivin et al., 2021; Zhong et al., 2021; figure 3). These results are reminiscent of pioneering studies showing that CAG expanded repeats located within the *ATXN8* and *JPH3AS* RNAs, which were initially considered as long non-coding RNAs but were later on found to be carrying small ORFs, are translated through canonical ATG-initiated translation into toxic polyglutamine proteins in SCA8 and HDL2 (Moseley et al., 2006; Wilburn et al., 2011; figure 3). Finally, it is consistently reported that various repeat expansions can be RAN translated into toxic proteins through initiation directly within the repeats and in all three frames (Zu et al., 2011; review in Cleary et al., 2018; Nguyen et al., 2019). Thus, whether the CGG repeat expansions causing OPML and OPDM pathologies and located in genetic sequences predicted as non-coding are nonetheless translated into potentially toxic proteins, which would accumulate into ubiquitin-positive inclusions, is an exciting question that remains to be addressed.

Expansion without anticipation in NIID and OPDM?

In contrast to familial anticipation and correlation between expansion lengths and disease severity and/or age of onset observed in various microsatellite diseases (FXS, polyQ disorders, myotonic dystrophy type 1, etc.), there are only few examples of anticipation reported in NIID and OPDM families, and a limited correlation between the size of CGG expansions and the age of onset in OPDM1 and OPDM2, with a r^2 of 0.188 and 0.158, respectively (Kumutpongpanich et al., 2021; Xi et al., 2021). On the contrary, there are increasing reports of individuals carrying large CGG expansions (> 200-300 repeats) in the *NOTCH2NLC*, *GIPC1* and *LRP12* genes and who are free of overt clinical manifestations (Deng et al., 2020; Ogasawara et al., 2020; Deng et al., 2021; Kumutpongpanich et al., 2021; Fukuda et al., 2021; Yu et al., 2021). Molecular analyses indicate that these individuals present DNA hypermethylation associated to promoter silencing and loss of expression of the allele carrying these large CGG expansions. These results have important consequences. First, they indicate that a partial reduction in expression and/or haploinsufficiency of the NOTCH2NLC, LRP12 and GIPC1 proteins are not associated with overt clinical manifestations, dismissing the hypothesis of a loss of function mechanism in OPDM and NIID. Second, these data also provide some molecular bases to the complex pedigree observed in these diseases, with increasing reports of asymptomatic ascendants who nevertheless carry the causative mutation, but of longer size (over 200-300 CGG repeats). Interestingly, of the ~10 currently reported examples of asymptomatic individuals with large expansions and who have transmitted shorter repeats to their NIID or OPDM-affected descendants, all were father to infant transmission cases, suggesting a bias of expansions contraction in male germinal cells (Deng et al., 2020; Deng et al., 2021; Kumutpongpanich et al., 2021; Fukuda et al., 2021; Yu et al., 2021). Of interest, a similar bias of transmission was observed in carriers of the *FMR1* CGG repeat premutation where contraction occurred more often in paternal than in maternal transmission (Nolin et al., 2019). Moreover, it is noteworthy that a difference in parental gender transmission is observed in other microsatellite diseases, such as myotonic dystrophy type 1 where large expansions of CTG repeats occurs mainly through maternal transmission (review in Lanni and Pearson, 2019), while in Huntington disease, maternal transmission mainly results in CAG repeat contraction, whereas paternal transmission is associated with CAG repeat expansion (Aziz et al., 2011).

Overall, these studies suggest that in NIID and OPDM, the causing CGG expansions are pathogenic in a narrow range of size, with a lower limit of 60 to 70 repeats, below which they are likely not long enough to be expressed in stable and toxic proteins, and a "protective" threshold around 200 to 300 CGG repeats, over which carriers of large expansions benefit from the silencing of the toxic allele (figure 4). However, these intriguing observations are currently limited to a small number of clinical cases, and larger studies will be required to refine the correlation between CGG expansion sizes and age of onset and/or clinical severity in these disorders.

Putative mechanisms of polyG proteins toxicity

FXTAS and NIID are characterized by the translation of CGG repeats into pathogenic FMRpolyG and uN2CpolyG proteins, respectively. These two proteins form protein inclusions, are toxic in cell and animal models and possess an identical polyglycine stretch, suggesting a common mechanism of toxicity. However, this raises numerous questions, notably it is unclear whether these polyG proteins are toxic under their aggregated or soluble forms. Similarly, it remains to explore whether their cellular localization, notably their

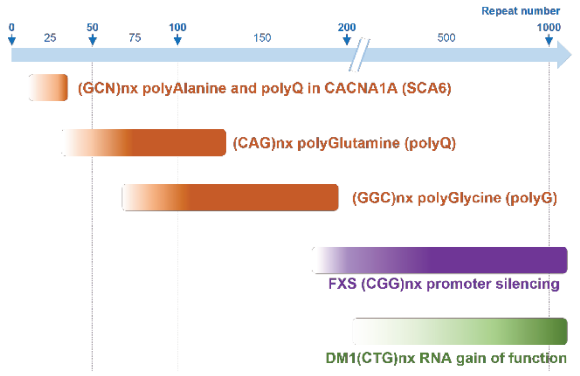


Figure 4. Repeat expansion range of toxicity in different microsatellite diseases.
We propose that CGG expansions are pathogenic between ~60-70 to 200-300 repeats when they are expressed into toxic polyglycine-containing proteins, but over 200 repeats when they promote DNA epigenetic changes, promoter silencing and a loss of function mechanism, such as in FXS. This range of protein toxicity is to be compared to other translated microsatellite diseases such as polyalanine-containing proteins that are generally toxic with 10 to 35 GCN repeats, while polyQ proteins are generally pathogenic with longer CAG expansions (~40 to 80-200 repeats), with the exception of short polyglutamine stretch altering functions of the calcium voltage-gated channel subunit alpha1A (CACNA1A) in SCA6. Of interest, DNA epigenetic changes and promoter silencing in a loss of function mechanism, or an RNA gain of function mechanism such as titration of the MBNL RNA binding proteins in DM1, require much longer repeat expansions.

import into the nucleus, is required for their pathogenicity. In that aspect, overexpression of FMRpolyG or uN2CpolyG in transformed cell lines mainly results in their accumulation in cytoplasmic aggregates, while these proteins are mainly found in intranuclear inclusions in patient tissues. Thus, it remains to determine the mechanism responsible of polyglycine-containing proteins nuclear import in these diseases. Furthermore, FXTAS and NIID are neurodegenerative diseases, but *FMR1* and *NOTCHNLC* expression are not restricted to neurons, and intranuclear inclusions are widely observed outside of the CNS in individuals with FXTAS and NIID (Greco et al., 2006; Hunsaker et al., 2011; Yamaguchi et al., 2018; Chen et al., 2020). However, presences of these inclusions are associated with only limited tissue dysfunctions and/or sub-clinical observations. In consequence, the mechanisms underlying the cellular specificity of FMRpolyG or uN2polyG pathogenicity remain to be determined. Finally, whether their toxicity is uniquely caused by their polyglycine stretches, or whether there is a toxic contribution of their bordering N- or C-terminal amino acid sequences is ill-defined. In that aspect, expression of a polyglycine protein deleted of any *FMR1* or *NOTCH2NLC* bordering sequence is toxic in cell models, however not as much as the full-length FMRpolyG or uN2polyG proteins (Sellier et al., 2017; Boivin et al., 2021). Furthermore, C-terminal tagging of the FMRpolyG protein by the cherry fluorescent protein prevents its aggregation, while fusion of FMRpolyG with the GFP promotes its stability and increases its expression (Sellier et al., 2017; Derbis et al., 2018). These data suggest that the sequences bordering the polyglycine stretch of FMRpolyG or uN2polyG may contribute to their pathogenicity. However, this remains to be confirmed, notably in animal models.

Conclusion

Translation of GGC repeat expansions into simar polyglycine-containing (polyG) proteins in two neurodegenerative diseases with overlapping clinical manifestations and nearly identical histopathological features, FXTAS and NIID, suggest the existence of a novel class of human genetic disorders, the polyG diseases. This model is inspired by the translation into toxic polyalanine- or polyglutamine-containing proteins of GCN or CAG repeat expansions embedded in the ORFs of diverse genes, resulting in the polyAla or polyQ diseases, respectively. Furthermore, these data confirm that expanded repeats localized in human genomic regions predicted as “non-coding” can nevertheless be translated into novel pathogenic proteins, notably through classical initiation at canonical AUG, or near-cognate, start codons. In that aspect, the similarities of clinical manifestations and histopathological characteristic between FXTAS, NIID, OPDM, and OPML question whether CGG repeats also located in sequences predicted as non-coding, such as the *LOC642361/NUTM2B-AS1* locus or the 5’UTR of the *LRP12* and *GIPC1* genes, are similarly translated into novel and potentially pathogenic polyG proteins. Importantly, a similar mechanism of toxicity for NIID, FXTAS and potentially other diseases, suggests that therapeutic strategies targeting the common CGG repeats/polyglycine-stretch could be of interest for the whole group of polyG diseases. In that aspect, recent pre-clinical development of small molecules and/or antisense oligonucleotides targeting the *FMR1* CGG repeats and that efficiently decrease expression of the toxic FMRpolyG in cell and animal models provide exciting therapeutic hope for these disorders (Disney et al., 2012; Qurashi et al., 2012; Tran et al., 2014; Yang et al., 2016; Green et al., 2019; Rodriguez et al., 2020; Verma et al., 2020; Konieczny 2021; Haify 2021; Derbis et al., 2021; Malik et al., 2021; Asamitsu et al., 2021).

Acknowledgments

This work was supported by the following grants: ERC-2012-StG 310659, ANR-18-CE16-0019 and FRM EQU202103012936 (NCB); ANR-10-LABX-0030-INRT and ANR-10-IDEX-0002-02 (IGBMC).

Author contributions

The manuscript was written and edited by M.B. and N.C-B.

Declaration of interests

The authors declare no conflicts of interest.

References

- Akarsu AN, Stoilov I, Yilmaz E, Sayli BS, Sarfarazi M. Genomic structure of HOXD13 gene: a nine polyalanine duplication causes synpolydactyly in two unrelated families. Hum Mol Genet. 1996 Jul;5(7):945-52.

- Albrecht A, Mundlos S. The other trinucleotide repeat: polyalanine expansion disorders. Curr Opin Genet Dev. 2005 Jun;15(3):285-93.

- Almeida S, Krishnan G, Rushe M, Gu Y, Kankel MW, Gao FB. Production of poly(GA) in C9ORF72 patient motor neurons derived from induced pluripotent stem cells. Acta Neuropathol. 2019 Dec;138(6):1099-1101.

- Amiel J, Laudier B, Attié-Bitach T, Trang H, de Pontual L, Gener B, Trochet D, Etchevers H, Ray P, Simonneau M, Vekemans M, Munnich A, Gaultier C, Lyonnet S. Polyalanine expansion and frameshift mutations of the paired-like homeobox gene PHOX2B in congenital central hypoventilation syndrome. Nat Genet. 2003 Apr;33(4):459-61.

- Asamitsu S, Yabuki Y, Ikenoshita S, Kawakubo K, Kawasaki M, Usuki S, Nakayama Y, Adachi K, Kugoh H, Ishii K, Matsuura T, Nanba E, Sugiyama H, Fukunaga K, Shioda N. CGG repeat RNA G-quadruplexes interact with FMRpolyG to cause neuronal dysfunction in fragile X-related tremor/ataxia syndrome. Sci Adv. 2021 Jan 13;7(3):eabd9440.

- Ariza J, Rogers H, Monterrubio A, Reyes-Miranda A, Hagerman PJ, Martínez-Cerdeño V. A Majority of FXTAS Cases Present with Intranuclear Inclusions Within Purkinje Cells. Cerebellum. 2016 Oct;15(5):546-51.

- Arocena DG, Iwahashi CK, Won N, Beilina A, Ludwig AL, Tassone F, Schwartz PH, Hagerman PJ. Induction of inclusion formation and disruption of lamin A/C structure by premutation CGG-repeat RNA in human cultured neural cells. Hum Mol Genet. 14: 3661-71.

- Aziz NA, van Belzen MJ, Coops ID, Belfroid RD, Roos RA. Parent-of-origin differences of mutant HTT CAG repeat instability in Huntington's disease. Eur J Med Genet. 2011 Jul-Aug;54(4):e413-8.

- Boivin M, Willemsen R, Hukema RK, Sellier C. Potential pathogenic mechanisms underlying Fragile X Tremor Ataxia Syndrome: RAN translation and/or RNA gain-of-function? Eur J Med Genet. 2018 Nov;61(11):674-679.

- Boivin M, Pfister V, Gaucherot A, Ruffenach F, Negroni L, Sellier C, Charlet-Berguerand N. Reduced autophagy upon C9ORF72 loss synergizes with dipeptide repeat protein toxicity in G4C2 repeat expansion disorders. EMBO J. 2020 Feb 17;39(4):e100574.

- Boivin M, Deng J, Pfister V, Grandgirard E, Oulad-Abdelghani M, Morlet B, Ruffenach F, Negroni L, Koebel P, Jacob H, Riet F, Dijkstra AA, McFadden K, Clayton WA, Hong D, Miyahara H, Iwasaki Y, Sone J, Wang Z, Charlet-Berguerand N. Translation of GGC repeat expansions into a toxic polyglycine protein in NIID defines a novel class of human genetic disorders: The polyG diseases. Neuron. 2021 Jun 2;109(11):1825-1835.e5.

- Bonapace G, Gullace R, Concolino D, Iannello G, Procopio R, Gagliardi M, Arabia G, Barbagallo G, Lupo A, Manfredini LI, Annesi G, Quattrone A. Intracellular FMRpolyG-Hsp70 complex in fibroblast cells from a patient affected by fragile X tremor ataxia syndrome. Heliyon. 2019 Jun 20;5(6):e01954.

- Bragg DC, Mangkalaphiban K, Vaine CA, Kulkarni NJ, Shin D, Yadav R, Dhakal J, Ton ML, Cheng A, Russo CT, Ang M, Acuña P, Go C, Franceour TN, Multhaupt-Buell T, Ito N, Müller U, Hendriks WT, Breakefield XO, Sharma N, Ozelius LJ. Disease onset in X-linked dystonia-parkinsonism correlates with expansion of a hexameric repeat within an SVA retrotransposon in TAF1. Proc Natl Acad Sci U S A. 2017 Dec 19;114(51):E11020-E11028.

- Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, Hunter K, Stanton VP, Thirion JP, Hudson T, et al. Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. Cell. 1992 Feb 21;68(4):799-808.

- Brown LY, Odent S, David V, Blayau M, Dubourg C, Apacik C, Delgado MA, Hall BD, Reynolds JF, Sommer A, Wieczorek D, Brown SA, Muenke M. Holoprosencephaly due to mutations in ZIC2: alanine tract expansion mutations may be caused by parental somatic recombination. Hum Mol Genet. 2001 Apr 1;10(8):791-6.

- Buijsen RA, Sellier C, Severijnen LA, Oulad-Abdelghani M, Verhagen RF, Berman RF, Charlet-Berguerand N, Willemsen R, Hukema RK. FMRpolyG-positive inclusions in CNS and non-CNS organs of a fragile X premutation carrier with fragile X-associated tremor/ataxia syndrome. Acta Neuropathol Commun. 2014 Nov 26;2:162.

- Brais B, Bouchard JP, Xie YG, Rochefort DL, Chrétien N, Tomé FM, Lafrenière RG, Rommens JM, Uyama E, Nohira O, Blumen S, Korczyn AD, Heutink P, Mathieu J, Duranceau A, Codère F, Fardeau M, Rouleau GA. Short GCG expansions in the PABP2 gene cause oculopharyngeal muscular dystrophy. Nat Genet. 1998 Feb;18(2):164-7.

- Campuzano V, Montermini L, Moltò MD, Pianese L, Cossée M, Cavalcanti F, Monros E, Rodius F, Duclos F, Monticelli A, Zara F, Cañizares J, Koutnikova H, Bidichandani SI, Gellera C, Brice A, Trouillas P, De Michele G, Filla A, De Frutos R, Palau F, Patel PI, Di Donato S, Mandel JL, Cocozza S, Koenig M, Pandolfo M. Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science. 1996 Mar 8;271(5254):1423-7.

- Cao Y, Wu J, Yue Y, Zhang C, Liu S, Zhong P, Wang S, Huang X, Deng W, Pan J, Zheng L, Liu Q, Shang L, Zhang B, Yang J, Chen G, Chen S, Cao L, Luan X. Expanding the clinical spectrum of adult-onset neuronal intranuclear inclusion disease. Acta Neurol Belg. 2021 Feb 24.

- Cao L, Yan Y, Zhao G. NOTCH2NLC-related repeat expansion disorders: an expanding group of neurodegenerative disorders. Neurol Sci. 2021 Oct;42(10):4055-4062.

- Chen H, Lu L, Wang B, Cui G, Wang X, Wang Y, Raza HK, Min Y, Li K, Cui Y, Miao Z, Wan B, Sun M, Xu X. Re-defining the clinicopathological spectrum of neuronal intranuclear inclusion disease. Ann Clin Transl Neurol. 2020 Oct;7(10):1930-1941.

- Chen Z, Yan Yau W, Jaunmuktane Z, Tucci A, Sivakumar P, Gagliano Taliun SA, Turner C, Efthymiou S, Ibáñez K, Sullivan R, Bibi F, Athanasiou-Fragkouli A, Bourinaris T, Zhang D, Revesz T, Lashley T, DeTure M, Dickson DW, Josephs KA, Gelpi E, Kovacs GG, Halliday G, Rowe DB, Blair I, Tienari PJ, Suomalainen A, Fox NC, Wood NW, Lees AJ, Haltia MJ; Genomics England Research Consortium, Hardy J, Ryten M, Vandrovcova J, Houlden H. Neuronal intranuclear inclusion disease is genetically heterogeneous. Ann Clin Transl Neurol. 2020 Sep;7(9):1716-1725.

- Cid-Samper F, Gelabert-Baldrich M, Lang B, Lorenzo-Gotor N, Ponti RD, Severijnen LWFM, Bolognesi B, Gelpi E, Hukema RK, Botta-Orfila T, Tartaglia GG. An Integrative Study of Protein-RNA Condensates Identifies Scaffolding RNAs and Reveals Players in Fragile X-Associated Tremor/Ataxia Syndrome. Cell Rep. 2018 Dec 18;25(12):3422-3434.e7.

- Cleary JD, Pattamatta A, Ranum LPW. Repeat-associated non-ATG (RAN) translation. J Biol Chem. 2018 Oct 19;293(42):16127-16141.

- Clements JM, Laz TM, Sherman F. Efficiency of translation initiation by non-AUG codons in Saccharomyces cerevisiae. Mol Cell Biol. 1988 Oct;8(10):4533-6.

- Conway GS, Payne NN, Webb J, Murray A, Jacobs PA. Fragile X premutation screening in women with premature ovarian failure. Hum Reprod. 1998; 13: 1184-1187.

- Corbett MA, Kroes T, Veneziano L, Bennett MF, Florian R, Schneider AL, Coppola A, Licchetta L, Franceschetti S, Suppa A, Wenger A, Mei D, Pendziwiat M, Kaya S, Delledonne M, Straussberg R, Xumerle L, Regan B, Crompton D, van

Rootselaar AF, Correll A, Catford R, Bisulli F, Chakraborty S, Baldassari S, Tinuper P, Barton K, Carswell S, Smith M, Berardelli A, Carroll R, Gardner A, Friend KL, Blatt I, Iacomino M, Di Bonaventura C, Striano S, Buratti J, Keren B, Nava C, Forlani S, Rudolf G, Hirsch E, Leguern E, Labauge P, Balestrini S, Sander JW, Afawi Z, Helbig I, Ishiura H, Tsuji S, Sisodiya SM, Casari G, Sadleir LG, van Coller R, Tijssen MAJ, Klein KM, van den Maagdenberg AMJM, Zara F, Guerrini R, Berkovic SF, Pippucci T, Canafoglia L, Bahlo M, Striano P, Scheffer IE, Brancati F, Depienne C, Gecz J. Intronic ATTTC repeat expansions in STARD7 in familial adult myoclonic epilepsy linked to chromosome 2. Nat Commun. 2019 Oct 29;10(1):4920.

- Cortese A, Simone R, Sullivan R, Vandrovcova J, Tariq H, Yau WY, Humphrey J, Jaunmuktane Z, Sivakumar P, Polke J, Ilyas M, Tribollet E, Tomaselli PJ, Devigili G, Callegari I, Versino M, Salpietro V, Efthymiou S, Kaski D, Wood NW, Andrade NS, Buglo E, Rebelo A, Rossor AM, Bronstein A, Fratta P, Marques WJ, Züchner S, Reilly MM, Houlden H. Biallelic expansion of an intronic repeat in RFC1 is a common cause of late-onset ataxia. Nat Genet. 2019 Apr;51(4):649-658.

- De Baere E, Dixon MJ, Small KW, Jabs EW, Leroy BP, Devriendt K, Gillerot Y, Mortier G, Meire F, Van Maldergem L, Courtens W, Hjalgrim H, Huang S, Liebaers I, Van Regemorter N, Touraine P, Praphanphoj V, Verloes A, Udar N, Yellore V, Chalukya M, Yelchits S, De Paepe A, Kuttenn F, Fellous M, Veitia R, Messiaen L. Spectrum of FOXL2 gene mutations in blepharophimosis-ptosis-epicanthus inversus (BPES) families demonstrates a genotype--phenotype correlation. Hum Mol Genet. 2001 Jul 15;10(15):1591-600.

- DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford NJ, Nicholson AM, Finch NA, Flynn H, Adamson J, Kouri N, Wojtas A, Sengdy P, Hsiung GY, Karydas A, Seeley WW, Josephs KA, Coppola G, Geschwind DH, Wszolek ZK, Feldman H, Knopman DS, Petersen RC, Miller BL, Dickson DW, Boylan KB, Graff-Radford NR, Rademakers R. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron. 2011 Oct 20;72(2):245-56.

- Deng J, Gu M, Miao Y, Yao S, Zhu M, Fang P, Yu X, Li P, Su Y, Huang J, Zhang J, Yu J, Li F, Bai J, Sun W, Huang Y, Yuan Y, Hong D, Wang Z. Long-read sequencing identified repeat expansions in the 5'UTR of the NOTCH2NLC gene from Chinese patients with neuronal intranuclear inclusion disease. J Med Genet. 2019 Nov;56(11):758-764.

- Deng J, Yu J, Li P, Luan X, Cao L, Zhao J, Yu M, Zhang W, Lv H, Xie Z, Meng L, Zheng Y, Zhao Y, Gang Q, Wang Q, Liu J, Zhu M, Guo X, Su Y, Liang Y, Liang F, Hayashi T, Maeda MH, Sato T, Ura S, Oya Y, Ogasawara M, Iida A, Nishino I, Zhou C, Yan C, Yuan Y, Hong D, Wang Z. Expansion of GGC Repeat in GIPC1 Is Associated with Oculopharyngodistal Myopathy. Am J Hum Genet. 2020 Jun 4;106(6):793-804.

- Deng J, Zhou B, Yu J, Han X, Fu J, Li X, Xie X, Zhu M, Zheng Y, Guo X, Li P, Wang Q, Liu J, Zhang W, Yuan Y, Yao S, Wang Z, Hong D. Genetic origin of sporadic cases and RNA toxicity in neuronal intranuclear inclusion disease. J Med Genet. 2021 Mar 25:jmedgenet-2020-107649.

- Depienne C, Mandel JL. 30 years of repeat expansion disorders: What have we learned and what are the remaining challenges? Am J Hum Genet. 2021 May 6;108(5):764-785.

- Derbis M, Konieczny P, Walczak A, Sekrecki M, Sobczak K. Quantitative Evaluation of Toxic Polyglycine Biosynthesis and Aggregation in Cell Models Expressing Expanded CGG Repeats. Front Genet. 2018 Jun 19;9:216.

- Derbis M, Kul E, Niewiadomska D, Sekrecki M, Piasecka A, Taylor K, Hukema RK, Stork O, Sobczak K. Short antisense oligonucleotides alleviate the pleiotropic toxicity of RNA harboring expanded CGG repeats. Nat Commun. 2021 Feb 24;12(1):1265.

- Dijkstra AA, Haify SN, Verwey NA, Prins ND, van der Toorn EC, Rozemuller AJM, Bugiani M, den Dunnen WFA, Todd PK, Charlet-Berguerand N, Willemsen R, Hukema RK, Hoozemans JJM. Neuropathology of FMR1-premutation carriers presenting with dementia and neuropsychiatric symptoms. Brain Commun. 2021 Jan 27;3(1):fcab007.

- Disney MD, Liu B, Yang WY, Sellier C, Tran T, Charlet-Berguerand N, Childs-Disney JL. A small molecule that targets r(CGG)(exp) and improves defects in fragile X-associated tremor ataxia syndrome. ACS Chem Biol. 2012 Oct 19;7(10):1711-8.

- Durmus H, Laval SH, Deymeer F, Parman Y, Kiyan E, Gokyigiti M, Ertekin C, Ercan I, Solakoglu S, Karcagi V, Straub V, Bushby K, Lochmüller H, Serdaroglu-Oflazer P. Oculopharyngodistal myopathy is a distinct entity: clinical and genetic features of 47 patients. Neurology. 2011 Jan 18;76(3):227-35.

- Entezam A, Biacsi R, Orrison B, Saha T, Hoffman GE, Grabczyk E, Nussbaum RL, Usdin K. Regional FMRP deficits and large repeat expansions into the full mutation range in a new Fragile X premutation mouse model. Gene. 395:125-34.

- Fan Y, Xu Y, Shi C. NOTCH2NLC-related disorders: the widening spectrum and genotype-phenotype correlation. J Med Genet. 2022 Jan;59(1):1-9.

- Fang P, Yu Y, Yao S, Chen S, Zhu M, Chen Y, Zou K, Wang L, Wang H, Xin L, Hong T, Hong D. Repeat expansion scanning of the NOTCH2NLC gene in patients with multiple system atrophy. Ann Clin Transl Neurol. 2020 Apr;7(4):517-526.

- Fardaei M, Rogers MT, Thorpe HM, Larkin K, Hamshere MG, Harper PS, Brook JD. Three proteins, MBNL, MBLL and MBXL, co-localize in vivo with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells. Hum Mol Genet. 2002 Apr 1;11(7):805-14.

- Fiddes IT, Lodewijk GA, Mooring M, Bosworth CM, Ewing AD, Mantalas GL, Novak AM, van den Bout A, Bishara A, Rosenkrantz JL, Lorig-Roach R, Field AR, Haeussler M, Russo L, Bhaduri A, Nowakowski TJ, Pollen AA, Dougherty ML, Nuttle X, Addor MC, Zwolinski S, Katzman S, Kriegstein A, Eichler EE, Salama SR, Jacobs FMJ, Haussler D. Human-Specific NOTCH2NL Genes Affect Notch Signaling and Cortical Neurogenesis. Cell. 2018 May 31;173(6):1356-1369.e22.

- Fields AP, Rodriguez EH, Jovanovic M, Stern-Ginossar N, Haas BJ, Mertins P, Raychowdhury R, Hacohen N, Carr SA, Ingolia NT, Regev A, Weissman JS. A Regression-Based Analysis of Ribosome-Profiling Data Reveals a Conserved Complexity to Mammalian Translation. Mol Cell. 2015 Dec 3;60(5):816-827.

- Florian RT, Kraft F, Leitão E, Kaya S, Klebe S, Magnin E, van Rootselaar AF, Buratti J, Kühnel T, Schröder C, Giesselmann S, Tschernoster N, Altmueller J, Lamiral A, Keren B, Nava C, Bouteiller D, Forlani S, Jornea L, Kubica R, Ye T, Plassard D, Jost B, Meyer V, Deleuze JF, Delpu Y, Avarello MDM, Vijfhuizen LS, Rudolf G, Hirsch E, Kroes T, Reif PS, Rosenow F, Ganos C, Vidailhet M, Thivard L, Mathieu A, Bourgeron T, Kurth I, Rafehi H, Steenpass L, Horsthemke B; FAME consortium, LeGuern E, Klein KM, Labauge P, Bennett MF, Bahlo M, Gecz J, Corbett MA, Tijssen MAJ, van den Maagdenberg AMJM, Depienne C. Unstable TTTTA/TTTCA expansions in MARCH6 are associated with Familial Adult Myoclonic Epilepsy type 3. Nat Commun. 2019 Oct 29;10(1):4919.

- Fu YH, Kuhl DP, Pizzuti A, Pieretti M, Sutcliffe JS, Richards S, Verkerk AJ, Holden JJ, Fenwick RG Jr, Warren ST, et al. Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. Cell. 1991 Dec 20;67(6):1047-58.

- Fu YH, Pizzuti A, Fenwick RG Jr, King J, Rajnarayan S, Dunne PW, Dubel J, Nasser GA, Ashizawa T, de Jong P, et al. An unstable triplet repeat in a gene related to myotonic muscular dystrophy. Science. 1992 Mar 6;255(5049):1256-8.

- Fukuda H, Yamaguchi D, Nyquist K, Yabuki Y, Miyatake S, Uchiyama Y, Hamanaka K, Saida K, Koshimizu E, Tsuchida N, Fujita A, Mitsuhashi S, Ohbo K, Satake Y, Sone J, Doi H,

Morihara K, Okamoto T, Takahashi Y, Wenger AM, Shioda N, Tanaka F, Matsumoto N, Mizuguchi T. Father-to-offspring transmission of extremely long NOTCH2NLC repeat expansions with contractions: genetic and epigenetic profiling with long-read sequencing. Clin Epigenetics. 2021 Nov 13;13(1):204.

- Gao FB, Richter JD, Cleveland DW. Rethinking Unconventional Translation in Neurodegeneration. Cell. 2017 Nov 16;171(5):994-1000.

- Gelpi E, Botta-Orfila T, Bodi L, Marti S, Kovacs G, Grau-Rivera O, Lozano M, Sánchez-Valle R, Muñoz E, Valldeoriola F, Pagonabarraga J, Tartaglia GG, Milà M. Neuronal intranuclear (hyaline) inclusion disease and fragile X-associated tremor/ataxia syndrome: a morphological and molecular dilemma. Brain. 2017 Aug 1;140(8):e51.

- Glineburg MR, Todd PK, Charlet-Berguerand N, Sellier C. Repeat-associated non-AUG (RAN) translation and other molecular mechanisms in Fragile X Tremor Ataxia Syndrome. Brain Res. 2018 Aug 15;1693(Pt A):43-54.

- Gohel D, Sripada L, Prajapati P, Singh K, Roy M, Kotadia D, Tassone F, Charlet-Berguerand N, Singh R. FMRpolyG alters mitochondrial transcripts level and respiratory chain complex assembly in Fragile X associated tremor/ataxia syndrome [FXTAS]. Biochim Biophys Acta Mol Basis Dis. 2019 Jun 1;1865(6):1379-1388.

- Goodman FR, Bacchelli C, Brady AF, Brueton LA, Fryns JP, Mortlock DP, Innis JW, Holmes LB, Donnenfeld AE, Feingold M, Beemer FA, Hennekam RC, Scambler PJ. Novel HOXA13 mutations and the phenotypic spectrum of hand-foot-genital syndrome. Am J Hum Genet. 2000 Jul;67(1):197-202.

- Greco CM, Berman RF, Martin RM, Tassone F, Schwartz PH, Chang A, Trapp BD, Iwahashi C, Brunberg J, Grigsby J, et al. (2006) Neuropathology of fragile X-associated tremor/ataxia syndrome (FXTAS). Brain. 1:243-55.

- Green KM, Glineburg MR, Kearse MG, Flores BN, Linsalata AE, Fedak SJ, Goldstrohm AC, Barmada SJ, Todd PK. RAN translation at C9orf72-associated repeat expansions is selectively enhanced by the integrated stress response. Nat Commun. 2017 Dec 8;8(1):2005.

- Green KM, Sheth UJ, Flores BN, Wright SE, Sutter AB, Kearse MG, Barmada SJ, Ivanova MI, Todd PK.High-throughput screening yields several small-molecule inhibitors of repeat-associated non-AUG translation. J Biol Chem. 2019 Dec 6;294(49):18624-18638.

- Hagerman RJ, Leehey M, Heinrichs W, Tassone F, Wilson R, Hills J, Grigsby J, Gage B, Hagerman PJ. (2001). Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. Neurology. 1:127-30.

- Haify SN, Buijsen RAM, Verwegen L, Severijnen LWFM, de Boer H, Boumeester V, Monshouwer R, Yang WY, Cameron MD, Willemsen R, Disney MD, Hukema RK. Small molecule 1a reduces FMRpolyG-mediated toxicity in in vitro and in vivo models for FMR1 premutation. Hum Mol Genet. 2021 Aug 12;30(17):1632-1648.

- Harley HG, Brook JD, Rundle SA, Crow S, Reardon W, Buckler AJ, Harper PS, Housman DE, Shaw DJ. Expansion of an unstable DNA region and phenotypic variation in myotonic dystrophy. Nature. 1992 Feb 6;355(6360):545-6.

- Hashem V, Galloway JN, Mori M, Willemsen R, Oostra BA, Paylor R, Nelson DL. (2009). Ectopic expression of CGG containing mRNA is neurotoxic in mammals. Hum Mol Genet. 18: 2443-51.

- Hayes LR, Duan L, Bowen K, Kalab P, Rothstein JD. C9orf72 arginine-rich dipeptide repeat proteins disrupt karyopherin-mediated nuclear import. Elife. 2020 Mar 2;9:e51685.

- He F, Krans A, Freibaum BD, Taylor JP, Todd PK. TDP-43 suppresses CGG repeat-induced neurotoxicity through interactions with HnRNP A2/B1. Hum Mol Genet. 2014 Oct 1;23(19):5036-51.

- Hinnebusch AG, Ivanov IP, Sonenberg N. Translational control by 5'-untranslated regions of eukaryotic mRNAs. Science. 2016 Jun 17;352(6292):1413-6.

- Hoem G, Bowitz Larsen K, Øvervatn A, Brech A, Lamark T, Sjøttem E, Johansen T. The FMRpolyGlycine Protein Mediates Aggregate Formation and Toxicity Independent of the CGG mRNA Hairpin in a Cellular Model for FXTAS. Front Genet. 2019 Mar 28;10:249.

- Holm KN, Herren AW, Taylor SL, Randol JL, Kim K, Espinal G, Martiínez-Cerdeño V, Pessah IN, Hagerman RJ, Hagerman PJ. Human Cerebral Cortex Proteome of Fragile X-Associated Tremor/Ataxia Syndrome. Front Mol Biosci. 2021 Jan 29;7:600840.

- Holmes SE, O'Hearn EE, McInnis MG, Gorelick-Feldman DA, Kleiderlein JJ, Callahan C, Kwak NG, Ingersoll-Ashworth RG, Sherr M, Sumner AJ, Sharp AH, Ananth U, Seltzer WK, Boss MA, Vieria-Saecker AM, Epplen JT, Riess O, Ross CA, Margolis RL. Expansion of a novel CAG trinucleotide repeat in the 5' region of PPP2R2B is associated with SCA12. Nat Genet. 1999 Dec;23(4):391-2.

- Hukema RK, Buijsen RA, Schonewille M, Raske C, Severijnen LA, Nieuwenhuizen-Bakker I, Verhagen RF, van Dessel L, Maas A, Charlet-Berguerand N, et al. Reversibility of neuropathology and motor deficits in an inducible mouse model for FXTAS. Hum Mol Genet. 24(17):4948-57.

- Hunsaker MR, Greco CM, Spath MA, Smits AP, Navarro CS, Tassone F, Kros JM, Severijnen LA, Berry-Kravis EM, Berman RF, Hagerman PJ, Willemsen R, Hagerman RJ, Hukema RK. Widespread non-central nervous system organ pathology in fragile X premutation carriers with fragile X-associated tremor/ataxia syndrome and CGG knock-in mice. Acta Neuropathol. 2011 Oct;122(4):467-79.

- Ingolia NT, Lareau LF, Weissman JS. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell. 2011 Nov 11;147(4):789-802.

- Ishiura H, Tsuji S. Advances in repeat expansion diseases and a new concept of repeat motif-phenotype correlation. Curr Opin Genet Dev. 2020 Dec;65:176-185.

- Ishiura H, Doi K, Mitsui J, Yoshimura J, Matsukawa MK, Fujiyama A, Toyoshima Y, Kakita A, Takahashi H, Suzuki Y, Sugano S, Qu W, Ichikawa K, Yurino H, Higasa K, Shibata S, Mitsue A, Tanaka M, Ichikawa Y, Takahashi Y, Date H, Matsukawa T, Kanda J, Nakamoto FK, Higashihara M, Abe K, Koike R, Sasagawa M, Kuroha Y, Hasegawa N, Kanesawa N, Kondo T, Hitomi T, Tada M, Takano H, Saito Y, Sanpei K, Onodera O, Nishizawa M, Nakamura M, Yasuda T, Sakiyama Y, Otsuka M, Ueki A, Kaida KI, Shimizu J, Hanajima R, Hayashi T, Terao Y, Inomata-Terada S, Hamada M, Shirota Y, Kubota A, Ugawa Y, Koh K, Takiyama Y, Ohsawa-Yoshida N, Ishiura S, Yamasaki R, Tamaoka A, Akiyama H, Otsuki T, Sano A, Ikeda A, Goto J, Morishita S, Tsuji S. Expansions of intronic TTTCA and TTTTA repeats in benign adult familial myoclonic epilepsy. Nat Genet. 2018 Apr;50(4):581-590.

- Ishiura H, Shibata S, Yoshimura J, Suzuki Y, Qu W, Doi K, Almansour MA, Kikuchi JK, Taira M, Mitsui J, Takahashi Y, Ichikawa Y, Mano T, Iwata A, Harigaya Y, Matsukawa MK, Matsukawa T, Tanaka M, Shirota Y, Ohtomo R, Kowa H, Date H, Mitsue A, Hatsuta H, Morimoto S, Murayama S, Shiio Y, Saito Y, Mitsutake A, Kawai M, Sasaki T, Sugiyama Y, Hamada M, Ohtomo G, Terao Y, Nakazato Y, Takeda A, Sakiyama Y, Umeda-Kameyama Y, Shinmi J, Ogata K, Kohno Y, Lim SY, Tan AH, Shimizu J, Goto J, Nishino I, Toda T, Morishita S, Tsuji S. Noncoding CGG repeat expansions in neuronal intranuclear inclusion disease, oculopharyngodistal myopathy and an overlapping disease. Nat Genet. 2019 Aug;51(8):1222-1232.

- Iwahashi CK, Yasui DH, An HJ, Greco CM, Tassone F, Nannen K, Babineau B, Lebrilla CB, Hagerman RJ, Hagerman PJ. (2006). Protein composition of the intranuclear inclusions of FXTAS. Brain 129: 256-271

- Jacquemont S, Hagerman RJ, Leehey M, Grigsby J, Zhang L, Brunberg JA, Greco C, Des Portes V, Jardini T, Levine R, et al. (2003). Fragile X premutation tremor/ataxia syndrome: molecular, clinical, and neuroimaging correlates. Am J Hum

Genet. 4:869-78.

- Jin P, Zarnescu DC, Zhang F, Pearson CE, Lucchesi JC, Moses K, Warren ST. (2003). RNA-mediated neurodegeneration caused by the fragile X premutation rCGG repeats in Drosophila. Neuron 39:739-747.

- Jin P, Duan R, Qurashi A, Qin Y, Tian D, Rosser TC, Liu H, Feng Y, Warren ST. (2007). Pur alpha binds to rCGG repeats and modulates repeat-mediated neurodegeneration in a Drosophila model of fragile X tremor/ataxia syndrome. Neuron 55: 556-564

- Johnstone TG, Bazzini AA, Giraldez AJ. Upstream ORFs are prevalent translational repressors in vertebrates. EMBO J. 2016 Apr 1;35(7):706-23.

- Jovičić A, Mertens J, Boeynaems S, Bogaert E, Chai N, Yamada SB, Paul JW 3rd, Sun S, Herdy JR, Bieri G, Kramer NJ, Gage FH, Van Den Bosch L, Robberecht W, Gitler AD. Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. Nat Neurosci. 2015 Sep;18(9):1226-9.

- Kawaguchi Y, Okamoto T, Taniwaki M, Aizawa M, Inoue M, Katayama S, Kawakami H, Nakamura S, Nishimura M, Akiguchi I, et al. CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. Nat Genet. 1994 Nov;8(3):221-8.

- Kearse MG, Green KM, Krans A, Rodriguez CM, Linsalata AE, Goldstrohm AC, Todd PK. CGG Repeat-Associated Non-AUG Translation Utilizes a Cap-Dependent Scanning Mechanism of Initiation to Produce Toxic Proteins. Mol Cell. 62(2):314-22.

- Kearse MG, Wilusz JE. Non-AUG translation: a new start for protein synthesis in eukaryotes. Genes Dev. 2017 Sep 1;31(17):1717-1731.

- Kenneson A, Zhang F, Hagedorn CH, Warren ST. Reduced FMRP and increased FMR1 transcription is proportionally associated with CGG repeat number in intermediate-length and premutation carriers. Hum Mol Genet. 2001 Jul 1;10(14):1449-54.

- Kimber TE, Blumbergs PC, Rice JP, Hallpike JF, Edis R, Thompson PD, Suthers G. Familial neuronal intranuclear inclusion disease with ubiquitin positive inclusions. J Neurol Sci. 1998 Sep 18;160(1):33-40.

- Knight SJ, Flannery AV, Hirst MC, Campbell L, Christodoulou Z, Phelps SR, Pointon J, Middleton-Price HR, Barnicoat A, Pembrey ME, et al. Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. Cell. 1993 Jul 16;74(1):127-34.

- Kobayashi H, Abe K, Matsuura T, Ikeda Y, Hitomi T, Akechi Y, Habu T, Liu W, Okuda H, Koizumi A. Expansion of intronic GGCCTG hexanucleotide repeat in NOP56 causes SCA36, a type of spinocerebellar ataxia accompanied by motor neuron involvement. Am J Hum Genet. 2011 Jul 15;89(1):121-30.

- Koide R, Ikeuchi T, Onodera O, Tanaka H, Igarashi S, Endo K, Takahashi H, Kondo R, Ishikawa A, Hayashi T, et al. Unstable expansion of CAG repeat in hereditary dentatorubral-pallidoluysian atrophy (DRPLA). Nat Genet. 1994 Jan;6(1):9-13.

- Koide R, Kobayashi S, Shimohata T, Ikeuchi T, Maruyama M, Saito M, Yamada M, Takahashi H, Tsuji S. A neurological disease caused by an expanded CAG trinucleotide repeat in the TATA-binding protein gene: a new polyglutamine disease? Hum Mol Genet. 1999 Oct;8(11):2047-53.

- Konieczny P, Mukherjee S, Stepniak-Konieczna E, Taylor K, Niewiadomska D, Piasecka A, Walczak A, Baud A, Dohno C, Nakatani K, Sobczak K. Cyclic mismatch binding ligands interact with disease-associated CGG trinucleotide repeats in RNA and suppress their translation. Nucleic Acids Res. 2021 Sep 20;49(16):9479-9495.

- Koob MD, Moseley ML, Schut LJ, Benzow KA, Bird TD, Day JW, Ranum LP. An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). Nat Genet. 1999 Apr;21(4):379-84.

- Kozak M. Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. Nucleic Acids Res. 1981 Oct 24;9(20):5233-52.

- Kozak M. Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems. Mol Cell Biol. 1989 Nov;9(11):5073-80.

- Kozak M. Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. Proc Natl Acad Sci U S A. 1990 Nov;87(21):8301-5.

- Krans A, Skariah G, Zhang Y, Bayly B, Todd PK. Neuropathology of RAN translation proteins in fragile X-associated tremor/ataxia syndrome. Acta Neuropathol Commun. 2019 Oct 30;7(1):152.

- Kremer EJ, Pritchard M, Lynch M, Yu S, Holman K, Baker E, Warren ST, Schlessinger D, Sutherland GR, Richards RI. Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. Science. 1991 Jun 21;252(5013):1711-4.

- Kumutpongpanich T, Ogasawara M, Ozaki A, Ishiura H, Tsuji S, Minami N, Hayashi S, Noguchi S, Iida A, Nishino I; OPDM_LRP12 Study Group, Mori-Yoshimura M, Oya Y, Ono K, Shimizu T, Kawata A, Shimohama S, Toyooka K, Endo K, Toru S, Sasaki O, Isahaya K, Takahashi MP, Iwasa K, Kira JI, Yamamoto T, Kawamoto M, Hamano T, Sugie K, Eura N, Shiota T, Koide M, Sekiya K, Kishi H, Hideyama T, Kawai S, Yanagimoto S, Sato H, Arahata H, Murayama S, Saito K, Hara H, Kanda T, Yaguchi H, Imai N, Kawagashira Y, Sanada M, Obara K, Kaido M, Furuta M, Kurashige T, Hara W, Kuzume D, Yamamoto M, Tsugawa J, Kishida H, Ishizuka N, Morimoto K, Tsuji Y, Tsuneyama A, Matsuno A, Sasaki R, Tamakoshi D, Abe E, Yamada S, Uzawa A. Clinicopathologic Features of Oculopharyngodistal Myopathy With LRP12 CGG Repeat Expansions Compared With Other Oculopharyngodistal Myopathy Subtypes. JAMA Neurol. 2021 Jul 1;78(7):853-863.

- LaCroix AJ, Stabley D, Sahraoui R, Adam MP, Mehaffey M, Kernan K, Myers CT, Fagerstrom C, Anadiotis G, Akkari YM, Robbins KM, Gripp KW, Baratela WAR, Bober MB, Duker AL, Doherty D, Dempsey JC, Miller DG, Kircher M, Bamshad MJ, Nickerson DA; University of Washington Center for Mendelian Genomics, Mefford HC, Sol-Church K. GGC Repeat Expansion and Exon 1 Methylation of XYLT1 Is a Common Pathogenic Variant in Baratela-Scott Syndrome. Am J Hum Genet. 2019 Jan 3;104(1):35-44.

- Lanni S, Pearson CE. Molecular genetics of congenital myotonic dystrophy. Neurobiol Dis. 2019 Dec;132:104533.

- La Spada AR, Wilson EM, Lubahn DB, Harding AE, Fischbeck KH. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. Nature. 1991 Jul 4;352(6330):77-9.

- Laumonnier F, Ronce N, Hamel BC, Thomas P, Lespinasse J, Raynaud M, Paringaux C, Van Bokhoven H, Kalscheuer V, Fryns JP, Chelly J, Moraine C, Briault S. Transcription factor SOX3 is involved in X-linked mental retardation with growth hormone deficiency. Am J Hum Genet. 2002 Dec;71(6):1450-5.

- Lee S, Liu B, Lee S, Huang SX, Shen B, Qian SB. Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. Proc Natl Acad Sci U S A. 2012 Sep 11;109(37):E2424-32.

- Li M, Li K, Li X, Tian Y, Shen L, Wu G, Zhang Z, Chen W. Multiple reversible encephalitic attacks: a rare manifestation of neuronal intranuclear inclusion disease. BMC Neurol. 2020 Apr 8;20(1):125.

- Lieberman AP, Shakkottai VG, Albin RL. Polyglutamine Repeats in Neurodegenerative Diseases. Annu Rev Pathol. 2019 Jan 24;14:1-27.

- Lim SY, Ishiura H, Ramli N, Shibata S, Almansour MA, Tan AH, Houlden H, Lang AE, Tsuji S. Adult-onset neuronal intranuclear inclusion disease mimicking Fragile X-associated tremor-ataxia syndrome in ethnic Chinese patients. Parkinsonism Relat Disord. 2020 May;74:25-27.

- Lindenberg, R., Rubinstein, L. J., Herman, M. M. & Haydon, G. B. A light and electron microscopy study of an unusual widespread nuclear inclusion body disease. A possible residuum of an old herpesvirus infection. Acta Neuropathol. 10, 54–73 (1968).

- Lindblad K, Savontaus ML, Stevanin G, Holmberg M, Digre K, Zander C, Ehrsson H, David G, Benomar A, Nikoskelainen E, Trottier Y, Holmgren G, Ptacek LJ, Anttinen A, Brice A, Schalling M. An expanded CAG repeat sequence in spinocerebellar ataxia type 7. Genome Res. 1996 Oct;6(10):965-71.

- Liquori CL, Ricker K, Moseley ML, Jacobsen JF, Kress W, Naylor SL, Day JW, Ranum LP. Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. Science. 2001 Aug 3;293(5531):864-7.

- Liu Y, Mimuro M, Yoshida M, Hashizume Y, Niwa H, Miyao S, Ujihira N, Akatsu H. Inclusion-positive cell types in adult-onset intranuclear inclusion body disease: implications for clinical diagnosis. Acta Neuropathol. 2008 Dec;116(6):615-23.

- Lu H, Luan X, Yuan Y, Dong M, Sun W, Yan C. The clinical and myopathological features of oculopharyngodistal myopathy in a Chinese family. Neuropathology. 2008 Dec;28(6):599-603.

- Ma L, Herren AW, Espinal G, Randol J, McLaughlin B, Martinez-Cerdeño V, Pessah IN, Hagerman RJ, Hagerman PJ. Composition of the Intranuclear Inclusions of Fragile X-associated Tremor/Ataxia Syndrome. Acta Neuropathol Commun. 2019 Sep 3;7(1):143.

- Ma D, Tan YJ, Ng ASL, Ong HL, Sim W, Lim WK, Teo JX, Ng EYL, Lim EC, Lim EW, Chan LL, Tan LCS, Yi Z, Tan EK. Association of NOTCH2NLC Repeat Expansions with Parkinson Disease. JAMA Neurol. 2020 Dec 1;77(12):1559-1563.

- Matsuura T, Yamagata T, Burgess DL, Rasmussen A, Grewal RP, Watase K, Khajavi M, McCall AE, Davis CF, Zu L, Achari M, Pulst SM, Alonso E, Noebels JL, Nelson DL, Zoghbi HY, Ashizawa T. Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10. Nat Genet. 2000 Oct;26(2):191-4.

- Margolis RL, O'Hearn E, Rosenblatt A, Willour V, Holmes SE, Franz ML, Callahan C, Hwang HS, Troncoso JC, Ross CA. A disorder similar to Huntington's disease is associated with a novel CAG repeat expansion. Ann Neurol. 2001 Dec;50(6):373-80

- McEachin ZT, Gendron TF, Raj N, García-Murias M, Banerjee A, Purcell RH, Ward PJ, Todd TW, Merritt-Garza ME, Jansen-West K, Hales CM, García-Sobrino T, Quintáns B, Holler CJ, Taylor G, San Millán B, Teijeira S, Yamashita T, Ohkubo R, Boulis NM, Xu C, Wen Z, Streichenberger N; Neuro–CEB Neuropathology Network, Fogel BL, Kukar T, Abe K, Dickson DW, Arias M, Glass JD, Jiang J, Tansey MG, Sobrido MJ, Petrucelli L, Rossoll W, Bassell GJ. Chimeric Peptide Species Contribute to Divergent Dipeptide Repeat Pathology in c9ALS/FTD and SCA36. Neuron. 2020 Jul 22;107(2):292-305.e6.

- Mahadevan M, Tsilfidis C, Sabourin L, Shutler G, Amemiya C, Jansen G, Neville C, Narang M, Barceló J, O'Hoy K, et al. Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. Science. 1992 Mar 6;255(5049):1253-5.

- Malik I, Tseng YJ, Wright SE, Zheng K, Ramaiyer P, Green KM, Todd PK. SRSF protein kinase 1 modulates RAN translation and suppresses CGG repeat toxicity. EMBO Mol Med. 2021 Nov 8;13(11):e14163.

- Malik I, Kelley CP, Wang ET, Todd PK. Molecular mechanisms underlying nucleotide repeat expansion disorders. Nat Rev Mol Cell Biol. 2021 Sep;22(9):589-607.

- Matsubara T, Saito Y, Kurashige T, Higashihara M, Hasegawa F, Ogasawara M, Iida A, Nishino I, Adachi T, Kubota A, Murayama S. Neuropathy/intranuclear inclusion bodies in oculopharyngodistal myopathy: A case report. eNeurologicalSci. 2021 Jun 4;24:100348.

- Michaud J, Gilbert JJ. Multiple system atrophy with neuronal intranuclear hyaline inclusions. Report of a new case with light and electron microscopic studies. Acta Neuropathol. 1981;54(2):113-9.

- Miller JW, Urbinati CR, Teng-Umnuay P, Stenberg MG, Byrne BJ, Thornton CA, Swanson MS. Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy. EMBO J. 2000 Sep 1;19(17):4439-48.

- Minami N, Ikezoe K, Kuroda H, Nakabayashi H, Satoyoshi E, Nonaka I. Oculopharyngodistal myopathy is genetically heterogeneous and most cases are distinct from oculopharyngeal muscular dystrophy. Neuromuscul Disord. 2001 Nov;11(8):699-702.

- Mootha VV, Gong X, Ku HC, Xing C. Association and familial segregation of CTG18.1 trinucleotide repeat expansion of TCF4 gene in Fuchs' endothelial corneal dystrophy. Invest Ophthalmol Vis Sci. 2014 Jan 2;55(1):33-42.

- Mori F, Tanji K, Odagiri S, Hattori M, Hoshikawa Y, Kono C, Yasui K, Yokoi S, Hasegawa Y, Kamitani T, Yoshida M, Wakabayashi K. Ubiquitin-related proteins in neuronal and glial intranuclear inclusions in intranuclear inclusion body disease. Pathol Int. 2012 Jun;62(6):407-11.

- Morriss GR, Cooper TA. Protein sequestration as a normal function of long noncoding RNAs and a pathogenic mechanism of RNAs containing nucleotide repeat expansions. Hum Genet. 2017 Sep;136(9):1247-1263.

- Moseley ML, Zu T, Ikeda Y, Gao W, Mosemiller AK, Daughters RS, Chen G, Weatherspoon MR, Clark HB, Ebner TJ, Day JW, Ranum LP. Bidirectional expression of CUG and CAG expansion transcripts and intranuclear polyglutamine inclusions in spinocerebellar ataxia type 8. Nat Genet. 2006 Jul;38(7):758-69.

- Mundlos S, Otto F, Mundlos C, Mulliken JB, Aylsworth AS, Albright S, Lindhout D, Cole WG, Henn W, Knoll JH, Owen MJ, Mertelsmann R, Zabel BU, Olsen BR. Mutations involving the transcription factor CBFA1 cause cleidocranial dysplasia. Cell. 1997 May 30;89(5):773-9.

- Munoz-Garcia D, Ludwin, SK. Adult-onset neuronal intranuclear hyaline inclusion disease. Neurology 36, 785–790 (1986).

- Nagafuchi S, Yanagisawa H, Sato K, Shirayama T, Ohsaki E, Bundo M, Takeda T, Tadokoro K, Kondo I, Murayama N, et al. Dentatorubral and pallidoluysian atrophy expansion of an unstable CAG trinucleotide on chromosome 12p. Nat Genet. 1994 Jan;6(1):14-8.

- Nakamura N, Tsunoda K, Mitsutake A, Shibata S, Mano T, Nagashima Y, Ishiura H, Iwata A, Toda T, Tsuji S, Sawamura H. Clinical Characteristics of Neuronal Intranuclear Inclusion Disease-Related Retinopathy With CGG Repeat Expansions in the NOTCH2NLC Gene. Invest Ophthalmol Vis Sci. 2020 Sep 1;61(11):27.

- Ng ASL, Xu Z, Chen Z, Tan YJ, Lim WK, Ting SKS, Yu WY, Cheng QH, Foo JN, Tan EK, Lim TCC. NOTCH2NLC-linked neuronal intranuclear inclusion body disease and fragile X-associated tremor/ataxia syndrome. Brain. 2020 Aug 1;143(8):e69.

- Ng ASL, Lim WK, Xu Z, Ong HL, Tan YJ, Sim WY, Ng EYL, Teo JX, Foo JN, Lim TCC, Yu WY, Chan LL, Lee HY, Chen Z, Lim EW, Ting SKS, Prakash KM, Tan LCS, Yi Z, Tan EK. NOTCH2NLC GGC Repeat Expansions Are Associated with Sporadic Essential Tremor: Variable Disease Expressivity on Long-Term Follow-up. Ann Neurol. 2020 Sep;88(3):614-618.

- Nguyen L, Cleary JD, Ranum LPW. Repeat-Associated Non-ATG Translation: Molecular Mechanisms and Contribution to Neurological Disease. Annu Rev Neurosci. 2019 Jul 8;42:227-247.

- Nolin SL, Glicksman A, Tortora N, Allen E, Macpherson J, Mila M, Vianna-Morgante AM, Sherman SL, Dobkin C, Latham GJ, Hadd AG. Am J Med Genet A. Expansions and contractions of the FMR1 CGG repeat in 5,508 transmissions

of normal, intermediate, and premutation alleles. 2019 Jul;179(7):1148-1156.

- Oberlé I, Rousseau F, Heitz D, Kretz C, Devys D, Hanauer A, Boué J, Bertheas MF, Mandel JL. Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. Science. 1991 May 24;252(5009):1097-102.

- Odagiri S, Tanji K, Mori F, Kakita A, Takahashi H, Kamitani T, Wakabayashi K. Immunohistochemical analysis of Marinesco bodies, using antibodies against proteins implicated in the ubiquitin-proteasome system, autophagy and aggresome formation. Neuropathology. 2012 Jun;32(3):261-6.

- Ogasawara M, Iida A, Kumutpongpanich T, Ozaki A, Oya Y, Konishi H, Nakamura A, Abe R, Takai H, Hanajima R, Doi H, Tanaka F, Nakamura H, Nonaka I, Wang Z, Hayashi S, Noguchi S, Nishino I. CGG expansion in NOTCH2NLC is associated with oculopharyngodistal myopathy with neurological manifestations. Acta Neuropathol Commun. 2020 Nov 25;8(1):204.

- Oh SY, He F, Krans A, Frazer M, Taylor JP, Paulson HL, Todd PK. RAN translation at CGG repeats induces ubiquitin proteasome system impairment in models of fragile X-associated tremor ataxia syndrome. Hum Mol Genet. 2015 Aug 1;24(15):4317-26.

- Okubo M, Doi H, Fukai R, Fujita A, Mitsuhashi S, Hashiguchi S, Kishida H, Ueda N, Morihara K, Ogasawara A, Kawamoto Y, Takahashi T, Takahashi K, Nakamura H, Kunii M, Tada M, Katsumoto A, Fukuda H, Mizuguchi T, Miyatake S, Miyake N, Suzuki J, Ito Y, Sone J, Sobue G, Takeuchi H, Matsumoto N, Tanaka F. GGC Repeat Expansion of NOTCH2NLC in Adult Patients with Leukoencephalopathy. Ann Neurol. 2019 Dec;86(6):962-968.

- Orr HT, Chung MY, Banfi S, Kwiatkowski TJ Jr, Servadio A, Beaudet AL, McCall AE, Duvick LA, Ranum LP, Zoghbi HY. Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. Nat Genet. 1993 Jul;4(3):221-6.

- Patel H, Norman MG, Perry TL, Berry KE. Multiple system atrophy with neuronal intranuclear hyaline inclusions. Report of a case and review of the literature. J Neurol Sci. 1985 Jan;67(1):57-65.

- Peabody DS. Translation initiation at non-AUG triplets in mammalian cells. J Biol Chem. 1989 Mar 25;264(9):5031-5.

- Pountney DL, Huang Y, Burns RJ, Haan E, Thompson PD, Blumbergs PC, Gai WP. SUMO-1 marks the nuclear inclusions in familial neuronal intranuclear inclusion disease. Exp Neurol. 2003 Nov;184(1):436-46.

- Pulst SM, Nechiporuk A, Nechiporuk T, Gispert S, Chen XN, Lopes-Cendes I, Pearlman S, Starkman S, Orozco-Diaz G, Lunkes A, DeJong P, Rouleau GA, Auburger G, Korenberg JR, Figueroa C, Sahba S. Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. Nat Genet. 1996 Nov;14(3):269-76.

- Qurashi A, Li W, Zhou JY, Peng J, Jin P. Nuclear accumulation of stress response mRNAs contributes to the neurodegeneration caused by Fragile X premutation rCGG repeats. PLoS Genet. 2011 Jun;7(6):e1002102.

- Qurashi A, Liu H, Ray L, Nelson DL, Duan R, Jin P. Chemical screen reveals small molecules suppressing fragile X premutation rCGG repeat-mediated neurodegeneration in Drosophila. Hum Mol Genet. 2012 May 1;21(9):2068-75.

- Rafehi H, Szmulewicz DJ, Bennett MF, Sobreira NLM, Pope K, Smith KR, Gillies G, Diakumis P, Dolzhenko E, Eberle MA, Barcina MG, Breen DP, Chancellor AM, Cremer PD, Delatycki MB, Fogel BL, Hackett A, Halmagyi GM, Kapetanovic S, Lang A, Mossman S, Mu W, Patrikios P, Perlman SL, Rosemergy I, Storey E, Watson SRD, Wilson MA, Zee DS, Valle D, Amor DJ, Bahlo M, Lockhart PJ. Bioinformatics-Based Identification of Expanded Repeats: A Non-reference Intronic Pentamer Expansion in RFC1 Causes CANVAS. Am J Hum Genet. 2019 Jul 3;105(1):151-165.

- Renton AE, Majounie E, Waite A, Simón-Sánchez J, Rollinson S, Gibbs JR, Schymick JC, Laaksovirta H, van Swieten JC, Myllykangas L, Kalimo H, Paetau A, Abramzon Y, Remes AM, Kaganovich A, Scholz SW, Duckworth J, Ding J, Harmer DW, Hernandez DG, Johnson JO, Mok K, Ryten M, Trabzuni D, Guerreiro RJ, Orrell RW, Neal J, Murray A, Pearson J, Jansen IE, Sondervan D, Seelaar H, Blake D, Young K, Halliwell N, Callister JB, Toulson G, Richardson A, Gerhard A, Snowden J, Mann D, Neary D, Nalls MA, Peuralinna T, Jansson L, Isoviita VM, Kaivorinne AL, Hölttä-Vuori M, Ikonen E, Sulkava R, Benatar M, Wuu J, Chiò A, Restagno G, Borghero G, Sabatelli M; ITALSGEN Consortium, Heckerman D, Rogaeva E, Zinman L, Rothstein JD, Sendtner M, Drepper C, Eichler EE, Alkan C, Abdullaev Z, Pack SD, Dutra A, Pak E, Hardy J, Singleton A, Williams NM, Heutink P, Pickering-Brown S, Morris HR, Tienari PJ, Traynor BJ. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron. 2011 Oct 20;72(2):257-68.

- Robinson DO, Wills AJ, Hammans SR, Read SP, Sillibourne J. Oculopharyngeal muscular dystrophy: a point mutation which mimics the effect of the PABPN1 gene triplet repeat expansion mutation. J Med Genet. 2006 May;43(5):e23.

- Rodriguez CM, Wright SE, Kearse MG, Haenfler JM, Flores BN, Liu Y, Ifrim MF, Glineburg MR, Krans A, Jafar-Nejad P, Sutton MA, Bassell GJ, Parent JM, Rigo F, Barmada SJ, Todd PK. A native function for RAN translation and CGG repeats in regulating fragile X protein synthesis. Nat Neurosci. 2020 Mar;23(3):386-397.

- Rong Z, Hu J, Corey DR, Mootha VV. Quantitative Studies of Muscleblind Proteins and Their Interaction with TCF4 RNA Foci Support Involvement in the Mechanism of Fuchs' Dystrophy. Invest Ophthalmol Vis Sci. 2019 Sep 3;60(12):3980-3991.

- Saito R, Shimizu H, Miura T, Hara N, Mezaki N, Higuchi Y, Miyashita A, Kawachi I, Sanpei K, Honma Y, Onodera O, Ikeuchi T, Kakita A. Oculopharyngodistal myopathy with coexisting histology of systemic neuronal intranuclear inclusion disease: Clinicopathologic features of an autopsied patient harboring CGG repeat expansions in LRP12. Acta Neuropathol Commun. 2020 Jun 3;8(1):75.

- Sato N, Amino T, Kobayashi K, Asakawa S, Ishiguro T, Tsunemi T, Takahashi M, Matsuura T, Flanigan KM, Iwasaki S, Ishino F, Saito Y, Murayama S, Yoshida M, Hashizume Y, Takahashi Y, Tsuji S, Shimizu N, Toda T, Ishikawa K, Mizusawa H. Spinocerebellar ataxia type 31 is associated with "inserted" penta-nucleotide repeats containing (TGGAA)n. Am J Hum Genet. 2009 Nov;85(5):544-57.

- Satoyoshi E, Kinoshita M. Oculopharyngodistal myopathy. Arch Neurol. 1977 Feb;34(2):89-92.

- Seixas AI, Loureiro JR, Costa C, Ordóñez-Ugalde A, Marcelino H, Oliveira CL, Loureiro JL, Dhingra A, Brandão E, Cruz VT, Timóteo A, Quintáns B, Rouleau GA, Rizzu P, Carracedo Á, Bessa J, Heutink P, Sequeiros J, Sobrido MJ, Coutinho P, Silveira I. A Pentanucleotide ATTTC Repeat Insertion in the Non-coding Region of DAB1, Mapping to SCA37, Causes Spinocerebellar Ataxia. Am J Hum Genet. 2017 Jul 6;101(1):87-103.

- Sellier C, Rau F, Liu Y, Tassone F, Hukema RK, Gattoni R, Schneider A, Richard S, Willemsen R, Elliott DJ, et al. Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients. EMBO J. 29: 1248-61.

- Sellier, C., Freyermuth, F., Tabet, R., Tran, T., He, F., Ruffenach, F., Alunni, V., Moine, H., Thibault, C., Page, A., et al. Sequestration of DROSHA and DGCR8 by expanded CGG RNA repeats alters microRNA processing in fragile X-associated tremor/ataxia syndrome. Cell Rep 3(3):869-80.

- Sellier C, Buijsen RAM, He F, Natla S, Jung L, Tropel P, Gaucherot A, Jacobs H, Meziane H, Vincent A, Champy MF, Sorg T, Pavlovic G, Wattenhofer-Donze M, Birling MC, Oulad-Abdelghani M, Eberling P, Ruffenach F, Joint M, Anheim M, Martinez-Cerdeno V, Tassone F, Willemsen R, Hukema RK, Viville S, Martinat C, Todd PK, Charlet-Berguerand N. Translation of Expanded CGG Repeats into FMRpolyG Is

Pathogenic and May Contribute to Fragile X Tremor Ataxia Syndrome. Neuron. 2017 Jan 18;93(2):331-347.

- Sherman SL, Morton NE, Jacobs PA, Turner G. The marker (X) syndrome: a cytogenetic and genetic analysis. Ann Hum Genet. 1984 Jan;48(1):21-37.

- Sherman SL, Jacobs PA, Morton NE, Froster-Iskenius U, Howard-Peebles PN, Nielsen KB, Partington MW, Sutherland GR, Turner G, Watson M. Further segregation analysis of the fragile X syndrome with special reference to transmitting males. Hum Genet. 1985;69(4):289-99.

- Shi CH, Fan Y, Yang J, Yuan YP, Shen S, Liu F, Mao CY, Liu H, Zhang S, Hu ZW, Fan LY, Li MJ, Fan SH, Liu XJ, Xu YM. NOTCH2NLC Intermediate-Length Repeat Expansions Are Associated with Parkinson Disease. Ann Neurol. 2021 Jan;89(1):182-187.

- Shoubridge C, Gecz J. Polyalanine tract disorders and neurocognitive phenotypes. Adv Exp Med Biol. 2012;769:185-203.

- Sofola OA, Jin P, Qin Y, Duan R, Liu H, de Haro M, Nelson DL, Botas J. RNA-binding proteins hnRNP A2/B1 and CUGBP1 suppress fragile X CGG premutation repeat-induced neurodegeneration in a Drosophila model of FXTAS. Neuron 55: 565-571.

- Sone J, Tanaka F, Koike H, Inukai A, Katsuno M, Yoshida M, Watanabe H, Sobue G. Skin biopsy is useful for the antemortem diagnosis of neuronal intranuclear inclusion disease.

Neurology. 2011 Apr 19;76(16):1372-6. doi

- Sone J, Mori K, Inagaki T, Katsumata R, Takagi S, Yokoi S, Araki K, Kato T, Nakamura T, Koike H, Takashima H, Hashiguchi A, Kohno Y, Kurashige T, Kuriyama M, Takiyama Y, Tsuchiya M, Kitagawa N, Kawamoto M, Yoshimura H, Suto Y, Nakayasu H, Uehara N, Sugiyama H, Takahashi M, Kokubun N, Konno T, Katsuno M, Tanaka F, Iwasaki Y, Yoshida M, Sobue G. Clinicopathological features of adult-onset neuronal intranuclear inclusion disease. Brain. 2016 Dec;139(Pt 12):3170-3186.

- Sone J, Mitsuhashi S, Fujita A, Mizuguchi T, Hamanaka K, Mori K, Koike H, Hashiguchi A, Takashima H, Sugiyama H, Kohno Y, Takiyama Y, Maeda K, Doi H, Koyano S, Takeuchi H, Kawamoto M, Kohara N, Ando T, Ieda T, Kita Y, Kokubun N, Tsuboi Y, Katoh K, Kino Y, Katsuno M, Iwasaki Y, Yoshida M, Tanaka F, Suzuki IK, Frith MC, Matsumoto N, Sobue G. Long-read sequencing identifies GGC repeat expansions in NOTCH2NLC associated with neuronal intranuclear inclusion disease. Nat Genet. 2019 Aug;51(8):1215-1221.

- Sonobe Y, Ghadge G, Masaki K, Sendoel A, Fuchs E, Roos RP. Translation of dipeptide repeat proteins from the C9ORF72 expanded repeat is associated with cellular stress. Neurobiol Dis. 2018 Aug;116:155-165.

- Sonobe Y, Aburas J, Krishnan G, Fleming AC, Ghadge G, Islam P, Warren EC, Gu Y, Kankel MW, Brown AEX, Kiskinis E, Gendron TF, Gao FB, Roos RP, Kratsios P. A C. elegans model of C9orf72-associated ALS/FTD uncovers a conserved role for eIF2D in RAN translation. Nat Commun. 2021 Oct 15;12(1):6025.

- Strømme P, Mangelsdorf ME, Shaw MA, Lower KM, Lewis SM, Bruyere H, Lütcherath V, Gedeon AK, Wallace RH, Scheffer IE, Turner G, Partington M, Frints SG, Fryns JP, Sutherland GR, Mulley JC, Gécz J. Mutations in the human ortholog of Aristaless cause X-linked mental retardation and epilepsy. Nat Genet. 2002 Apr;30(4):441-5.

- Sugiyama A, Sato N, Kimura Y, Maekawa T, Enokizono M, Saito Y, Takahashi Y, Matsuda H, Kuwabara S. MR Imaging Features of the Cerebellum in Adult-Onset Neuronal Intranuclear Inclusion Disease: 8 Cases. AJNR Am J Neuroradiol. 2017 Nov;38(11):2100-2104.

- Suzuki IK, Gacquer D, Van Heurck R, Kumar D, Wojno M, Bilheu A, Herpoel A, Lambert N, Cheron J, Polleux F, Detours V, Vanderhaeghen P. Human-Specific NOTCH2NL Genes Expand Cortical Neurogenesis through Delta/Notch Regulation. Cell. 2018 May 31;173(6):1370-1384.e16.

- Sznajder ŁJ, Swanson MS. Short Tandem Repeat Expansions and RNA-Mediated Pathogenesis in Myotonic Dystrophy. Int J Mol Sci. 2019 Jul 9;20(13):3365.

- Tabet R, Schaeffer L, Freyermuth F, Jambeau M, Workman M, Lee CZ, Lin CC, Jiang J, Jansen-West K, Abou-Hamdan H, Désaubry L, Gendron T, Petrucelli L, Martin F, Lagier-Tourenne C. CUG initiation and frameshifting enable production of dipeptide repeat proteins from ALS/FTD C9ORF72 transcripts. Nat Commun. 2018 Jan 11;9(1):152.

- Takahashi-Fujigasaki J. Neuronal intranuclear hyaline inclusion disease. Neuropathology. 2003 Dec;23(4):351-9.

- Takahashi-Fujigasaki J, Nakano Y, Uchino A, Murayama S. Adult-onset neuronal intranuclear hyaline inclusion disease is not rare in older adults. Geriatr Gerontol Int. 2016 Mar;16 Suppl 1:51-6.

- Tassone F, Hagerman RJ, Loesch DZ, Lachiewicz A, Taylor AK, Hagerman PJ. Fragile X males with unmethylated, full mutation trinucleotide repeat expansions have elevated levels of FMR1 messenger RNA. Am J Med Genet 94: 232–236.

- Tassone F, Beilina A, Carosi C, Albertosi S, Bagni C, Li L, Glover K, Bentley D, Hagerman PJ. Elevated FMR1 mRNA in premutation carriers is due to increased transcription. RNA. 2007;13(4):555-62

- The Huntington's Disease Collaborative Research Group. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell. 1993 Mar 26;72(6):971-83.

- Tian Y, Wang JL, Huang W, Zeng S, Jiao B, Liu Z, Chen Z, Li Y, Wang Y, Min HX, Wang XJ, You Y, Zhang RX, Chen XY, Yi F, Zhou YF, Long HY, Zhou CJ, Hou X, Wang JP, Xie B, Liang F, Yang ZY, Sun QY, Allen EG, Shafik AM, Kong HE, Guo JF, Yan XX, Hu ZM, Xia K, Jiang H, Xu HW, Duan RH, Jin P, Tang BS, Shen L. Expansion of Human-Specific GGC Repeat in Neuronal Intranuclear Inclusion Disease-Related Disorders. Am J Hum Genet. 2019 Jul 3;105(1):166-176.

- Todd, P. K., Oh, S. Y., Krans, A., He, F., Sellier, C., Frazer, M., Renoux, A. J., Chen, K. C., Scaglione, K. M., Basrur, V., et al. (2013). CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. Neuron 78(3):440-55.

- Toko M, Ohshita T, Kurashige T, Morino H, Kume K, Yamashita H, Sobue G, Iwasaki Y, Sone J, Kawakami H, Maruyama H. FXTAS is difficult to differentiate from neuronal intranuclear inclusion disease through skin biopsy: a case report. BMC Neurol. 2021 Oct 12;21(1):396.

- Tran T, Childs-Disney JL, Liu B, Guan L, Rzuczek S, Disney MD. Targeting the r(CGG) repeats that cause FXTAS with modularly assembled small molecules and oligonucleotides. ACS Chem Biol. 2014 Apr 18;9(4):904-12.

- Uyama E, Uchino M, Chateau D, Tomé FM. Autosomal recessive oculopharyngodistal myopathy in light of distal myopathy with rimmed vacuoles and oculopharyngeal muscular dystrophy. Neuromuscul Disord. 1998 Apr;8(2):119-25.

- van Kuilenburg ABP, Tarailo-Graovac M, Richmond PA, Drögemöller BI, Pouladi MA, Leen R, Brand-Arzamendi K, Dobritzsch D, Dolzhenko E, Eberle MA, Hayward B, Jones MJ, Karbassi F, Kobor MS, Koster J, Kumari D, Li M, MacIsaac J, McDonald C, Meijer J, Nguyen C, Rajan-Babu IS, Scherer SW, Sim B, Trost B, Tseng LA, Turkenburg M, van Vugt JJFA, Veldink JH, Walia JS, Wang Y, van Weeghel M, Wright GEB, Xu X, Yuen RKC, Zhang J, Ross CJ, Wasserman WW, Geraghty MT, Santra S, Wanders RJA, Wen XY, Waterham HR, Usdin K, van Karnebeek CDM. Glutaminase Deficiency Caused by Short Tandem Repeat Expansion in GLS. N Engl J Med. 2019 Apr 11;380(15):1433-1441.

- van der Sluijs BM, ter Laak HJ, Scheffer H, van der Maarel SM, van Engelen BG. Autosomal recessive oculopharyngodistal myopathy: a distinct phenotypical, histological, and genetic entity. J Neurol Neurosurg Psychiatry.

2004 Oct;75(10):1499-501.

- Verkerk AJ, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DP, Pizzuti A, Reiner O, Richards S, Victoria MF, Zhang FP, et al. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell. 1991 May 31;65(5):905-14.

- Verma AK, Khan E, Mishra SK, Mishra A, Charlet-Berguerand N, Kumar A. Curcumin Regulates the r(CGG)(exp) RNA Hairpin Structure and Ameliorate Defects in Fragile X-Associated Tremor Ataxia Syndrome. Front Neurosci. 2020 Apr 7;14:295.

- Wenzel HJ, Murray KD, Haify SN, Hunsaker MR, Schwartzer JJ, Kim K, La Spada AR, Sopher BL, Hagerman PJ, Raske C, Severijnen LWFM, Willemsen R, Hukema RK, Berman RF. Astroglial-targeted expression of the fragile X CGG repeat premutation in mice yields RAN translation, motor deficits and possible evidence for cell-to-cell propagation of FXTAS pathology. Acta Neuropathol Commun. 2019 Feb 26;7(1):27.

- Wheeler TM, Thornton CA. Myotonic dystrophy: RNA-mediated muscle disease. Curr Opin Neurol. 2007 Oct;20(5):572-6.

- Wieben ED, Aleff RA, Tosakulwong N, Butz ML, Highsmith WE, Edwards AO, Baratz KH. A common trinucleotide repeat expansion within the transcription factor 4 (TCF4, E2-2) gene predicts Fuchs corneal dystrophy. PLoS One. 2012;7(11):e49083.

- Wieben ED, Aleff RA, Tang X, Butz ML, Kalari KR, Highsmith EW, Jen J, Vasmatzis G, Patel SV, Maguire LJ, Baratz KH, Fautsch MP. Trinucleotide Repeat Expansion in the Transcription Factor 4 (TCF4) Gene Leads to Widespread mRNA Splicing Changes in Fuchs' Endothelial Corneal Dystrophy. Invest Ophthalmol Vis Sci. 2017 Jan 1;58(1):343-352.

- Wilburn B, Rudnicki DD, Zhao J, Weitz TM, Cheng Y, Gu X, Greiner E, Park CS, Wang N, Sopher BL, La Spada AR, Osmand A, Margolis RL, Sun YE, Yang XW. An antisense CAG repeat transcript at JPH3 locus mediates expanded polyglutamine protein toxicity in Huntington's disease-like 2 mice. Neuron. 2011 May 12;70(3):427-40.

- Willemsen R, Hoogeveen-Westerveld M, Reis S, Holstege J, Severijnen LA, Nieuwenhuizen IM, Schrier M, van Unen L, Tassone F, Hoogeveen AT, et al. The FMR1 CGG repeat mouse displays ubiquitin-positive intranuclear neuronal inclusions; implications for the cerebellar tremor/ataxia syndrome. Hum Mol Genet 12: 949-959.

- Woulfe JM. Abnormalities of the nucleus and nuclear inclusions in neurodegenerative disease: a work in progress. Neuropathol Appl Neurobiol. 2007 Feb;33(1):2-42.

- Xi J, Wang X, Yue D, Dou T, Wu Q, Lu J, Liu Y, Yu W, Qiao K, Lin J, Luo S, Li J, Du A, Dong J, Chen Y, Luo L, Yang J, Niu Z, Liang Z, Zhao C, Lu J, Zhu W, Zhou Y. 5' UTR CGG repeat expansion in GIPC1 is associated with oculopharyngodistal myopathy. Brain. 2021 Mar 3;144(2):601-614.

- Yamaguchi N, Mano T, Ohtomo R, Ishiura H, Almansour MA, Mori H, Kanda J, Shirota Y, Taira K, Morikawa T, Ikemura M, Yanagi Y, Murayama S, Shimizu J, Sakurai Y, Tsuji S, Iwata A. An Autopsy Case of Familial Neuronal Intranuclear Inclusion Disease with Dementia and Neuropathy. Intern Med. 2018 Dec 1;57(23):3459-3462.

- Yang WY, He F, Strack RL, Oh SY, Frazer M, Jaffrey SR, Todd PK, Disney MD.Small Molecule Recognition and Tools to Study Modulation of r(CGG)(exp) in Fragile X-Associated Tremor Ataxia Syndrome. ACS Chem Biol. 2016 Sep 16;11(9):2456-65.

- Yang D, Cen Z, Wang L, Chen X, Liu P, Wang H, Ouyang Z, Chen Y, Zhang F, Xie F, Wang B, Wu S, Yin H, Jiang B, Wang Z, Ji J, Luo W. Neuronal intranuclear inclusion disease tremor-dominant subtype: A mimicker of essential tremor. Eur J Neurol. 2021 Nov 9.

- Xinzhuang Yang, Dingding Zhang, Pidong Li, Jingwen Niu, Dan Xu, Xueyu Guo, Zhen Wang, Yanhuan Zhao, Haitao Ren, Chao Ling, Yang Wang, Jianxiong Shen, Yicheng Zhu, Depeng Wang, Liying Cui, Lin Chen, Yi Dai. Expansion of 5' UTR CGG repeat in RILPL1 is associated with oculopharyngodistal myopathy. medRxiv, posted September 29, 2021. https://doi.org/10.1101/2021.09.18.21263669

- Yeetong P, Pongpanich M, Srichomthong C, Assawapitaksakul A, Shotelersuk V, Tantirukdham N, Chunharas C, Suphapeetiporn K, Shotelersuk V. TTTCA repeat insertions in an intron of YEATS2 in benign adult familial myoclonic epilepsy type 4. Brain. 2019 Nov 1;142(11):3360-3366.

- Yu J, Deng J, Guo X, Shan J, Luan X, Cao L, Zhao J, Yu M, Zhang W, Lv H, Xie Z, Meng L, Zheng Y, Zhao Y, Gang Q, Wang Q, Liu J, Zhu M, Zhou B, Li P, Liu Y, Wang Y, Yan C, Hong D, Yuan Y, Wang Z. The GGC repeat expansion in NOTCH2NLC is associated with oculopharyngodistal myopathy type 3. Brain. 2021 Jul 28;144(6):1819-1832.

- Yuan Y, Liu Z, Hou X, Li W, Ni J, Huang L, Hu Y, Liu P, Hou X, Xue J, Sun Q, Tian Y, Jiao B, Duan R, Jiang H, Shen L, Tang B, Wang J. Identification of GGC repeat expansion in the NOTCH2NLC gene in amyotrophic lateral sclerosis. Neurology. 2020 Dec 15;95(24):e3394-e3405.

- Yuen P, Baxter DW. The morphology of Marinesco bodies (paranucleolar corpuscles) in the melanin-pigmented nuclei of the brain-stem. J Neurol Neurosurg Psychiatry. 1963 Apr;26(2):178-83.

- Zhang GJ, Wu D, Zhu YX, Ni HF, Zhang ZJ. Clinicopathological features of neuronal intranuclear inclusion disease diagnosed by skin biopsy. Neurol Sci. 2021 Aug 13.

- Zhao J, Liu J, Xiao J, Du J, Que C, Shi X, Liang W, Sun W, Zhang W, Lv H, Yuan Y, Wang Z. Clinical and muscle imaging findings in 14 mainland chinese patients with oculopharyngodistal myopathy. PLoS One. 2015 Jun 3;10(6):e0128629.

- Zhong S, Lian Y, Luo W, Luo R, Wu X, Ji J, Ji Y, Ding J, Wang X. Upstream open reading frame with NOTCH2NLC GGC expansion generates polyglycine aggregates and disrupts nucleocytoplasmic transport: implications for polyglycine diseases. sActa Neuropathol. 2021 Dec;142(6):1003-1023.

- Zhuchenko O, Bailey J, Bonnen P, Ashizawa T, Stockton DW, Amos C, Dobyns WB, Subramony SH, Zoghbi HY, Lee CC. Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the alpha 1A-voltage-dependent calcium channel. Nat Genet. 1997 Jan;15(1):62-9.

- Zu T, Gibbens B, Doty NS, Gomes-Pereira M, Huguet A, Stone MD, Margolis J, Peterson M, Markowski TW, Ingram MA, Nan Z, Forster C, Low WC, Schoser B, Somia NV, Clark HB, Schmechel S, Bitterman PB, Gourdon G, Swanson MS, Moseley M, Ranum LP. Non-ATG-initiated translation directed by microsatellite expansions. Proc Natl Acad Sci U S A. 2011 Jan 4;108(1):260-5.

2.4. Hypothèses et objectifs

Ces dernières années, grâce au progrès en termes de séquençage, de nouvelles mutations dont de « courtes » (< 200) expansions de répétitions de tri-nucléotides CGG situées dans des régions dites non-codantes, notamment dans les régions 5'UTR, de différents gènes ont été identifiées dans différentes pathologies : NIID, OPDM et OPML. Cependant, le mécanisme de toxicité de ces expansions reste à élucider. Mon travail de thèse a consisté à éclaircir par quel mécanisme ces répétitions CGG pouvaient être toxiques. Trois mécanismes sont actuellement proposés pour expliquer la toxicité des expansions de répétitions. Nous avons écarté l'hypothèse d'une toxicité au niveau ADN puisque ces maladies ont des symptômes communs ainsi que des caractéristiques histopathologiques similaires, bien que les répétitions de nucléotides CGG soient situées dans des gènes différents. De plus, ces « courtes » expansions, toutes situées dans la région 5'UTR des gènes, ne sont pas suffisantes pour induire des modifications épigénétiques conduisant à la perte d'expression du gène. En effet, aucune diminution de l'expression au niveau ARN et protéique de ces différents gènes n'a pas été observée chez les patients.

Au vu de la taille de ces expansions, nous avons supposé que ce nombre de répétitions (< 200) n'était pas suffisant pour titrer de manière suffisante des protéines de liaison à l'ARN, éliminant donc également un mécanisme de toxicité au niveau ARN. Cette supposition est soutenue par l'absence de foci d'ARN dans les noyaux cellulaires chez ces patients. Dans FXTAS, les répétitions CGG dans la région 5'UTR du gène *FMR1* sont traduites en une protéine composée de polyGlycine, qui forme des inclusions intranucléaires et dont l'expression est toxique dans différents modèles. Nous avons alors fait l'hypothèse que les expansions CGG dans les gènes *NOTCH2NLC* et *GIPC1* étaient, elles aussi, potentiellement traduites en protéines formant des inclusions nucléaires typiques de ces maladies. En effet, les similitudes de séquence (CGG), de taille (50 à 200), de localisation (5'UTR) des expansions de répétitions ainsi que caractéristiques histopathologiques communes (agrégats intranucléaires positifs à l'ubiquitine et p62) entre FXTAS, NIID et l'OPDM suggèrent un mécanisme pathogénique commun entre ces différentes maladies.

Cette question a fait l'objet de mon travail de thèse et les résultats sont présentés sous forme de publication (Publication 1) et de résultats. Enfin, après m'être intéressée à des expansions de répétitions CGG situées dans des régions non-codantes, je vais maintenant présenter une autre pathologie causée, elle, par des expansions GGGGCC localisées dans une région intronique : la sclérose latérale amyotrophique (SLA) associée à la démence fronto-temporale (DFT).

INTRODUCTION

- CHAPITRE 3 -

La sclérose latérale amyotrophique (SLA) et la démence fronto-temporale (DFT)

3. LA SCLEROSE LATERALE AMYOTROPHIQUE (SLA) ET LA DEMENCE FRONTO-TEMPORALE (DFT)

3.1. Epidémiologie et clinique

3.1.1. La sclérose latérale amyotrophique (SLA)

La sclérose latérale amyotrophique (SLA), également appelée maladie de Charcot ou encore maladie de Lou Gherig, est la troisième maladie neurodégénérative la plus commune après les maladies d'Alzheimer et de Parkinson. Son incidence est de 2,16 cas pour 100 000 personnes par an pour une prévalence d'environ 1 / 20 000 personnes (Logroscino et al.,2010). La SLA touche environ 1,5 homme pour 1 femme et apparaît généralement entre 55 et 65 ans, et très rarement au-delà de 80 ans. Bien que la majorité des cas de SLA soient sporadiques (SLAS), environ 10% des patients présentent une forme familiale (SLAF), généralement avec une transmission autosomale dominante (Hanby et al., 2011).

La SLA est caractérisée par une dégénérescence progressive des motoneurones supérieurs dans le cortex moteur, et inférieurs au niveau du tronc cérébral et de la colonne vertébrale. Cette neurodégénérescence engendre une atrophie et une faiblesse musculaire qui mène la majorité des patients à la paralysie, puis à la mort par déficience respiratoire en moyenne 3 à 5 ans après le diagnostic. Cliniquement, on distingue 2 formes distinctes de SLA, qui peuvent se succéder ou survenir simultanément (Figure 26 ; Swash et al., 2000 ; Swinnen et Robberecht, 2014) :

- La forme spinale concerne 2/3 des cas et commence par l'atteinte des membres, inférieurs ou supérieurs. Elle est caractérisée par une dégénérescence des motoneurones inférieurs conduisant à des sensations de faiblesse musculaire, des crampes et des sensations de raideur ;
- La forme bulbaire concerne 1/3 des cas et débute par des problèmes de déglutition, de phonation et d'élocution. Elle est due à la dégénérescence des motoneurones supérieurs.

A l'exception du Riluzole qui peut prolonger la vie des patients de quelques mois (Shaw et Ince, 1997), il n'existe actuellement aucun traitement pour cette maladie.

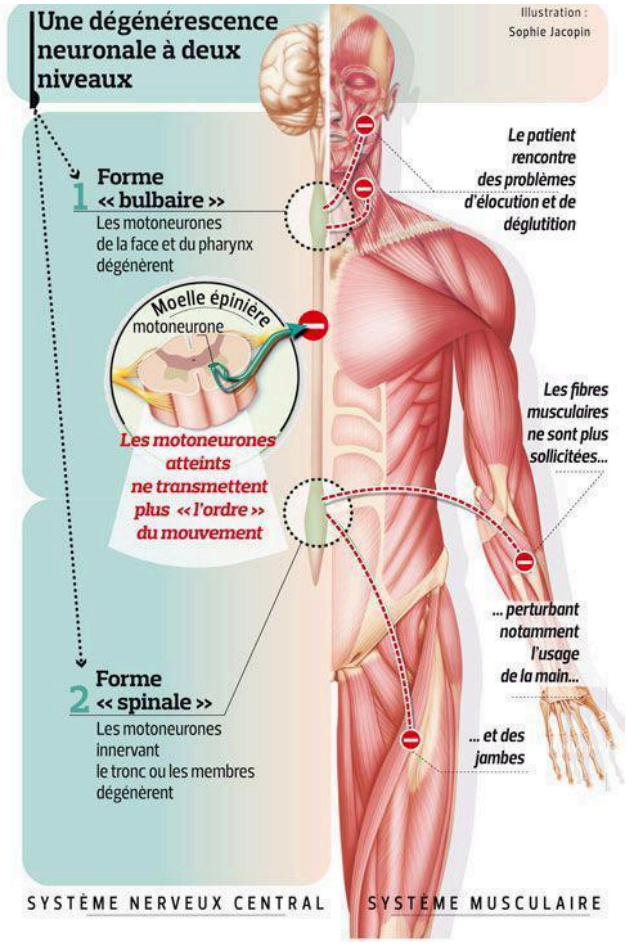


Figure 26. Les 2 formes de SLA

La SLA est une maladie affectant les neurones moteurs du cortex moteur et du tronc cérébral conduisant à la forme « bulbaire » caractérisée par des problèmes de déglutition, etc. ou ; de la moelle épinière conduisant à la forme « spinale » caractérisée par de la faiblesse musculaire des membres. Image du journal *Le Figaro*.

3.1.2. La démence fronto-temporale (DFT)

Après la maladie d'Alzheimer et les démences à corps de Lewy, la démence fronto-temporale (DFT) est la troisième forme de démence la plus fréquente (Vieira et al., 2013). Son incidence est d'environ 1,6 cas pour 100 000 personnes par an, pour une prévalence de 15 à 22 cas pour 100 000 personnes (Coyle-Gilchrist et al., 2016). Cette pathologie touche autant les hommes que les femmes et apparaît généralement entre 55 et 70 ans, avec environ 13% des patients développant la maladie avant l'âge de 50 ans (Mercy et al., 2008). Entre 30 à 50% des cas de DFT auraient une origine génétique.

La DFT est causée par une atrophie progressive des lobes frontaux et temporaux du cerveau, engendrant des troubles cliniquement variés (Figure 27). Ceux-ci peuvent être classifiés en 2 catégories majeures basées soit sur une altération comportement, soit sur une altération du langage (pour revue, Olney et al., 2017) :

- Le variant comportemental se caractérise par des changements dans le comportement, la personnalité et le contrôle émotionnel comme des altérations de la conduite en société, une désinhibition, une apathie affective, etc. ;
- L'aphasie progressive primaire se caractérise par des difficultés de langage comme des difficultés articulatoires, une élocution lente, la perte de la connaissance de certains mots, etc.

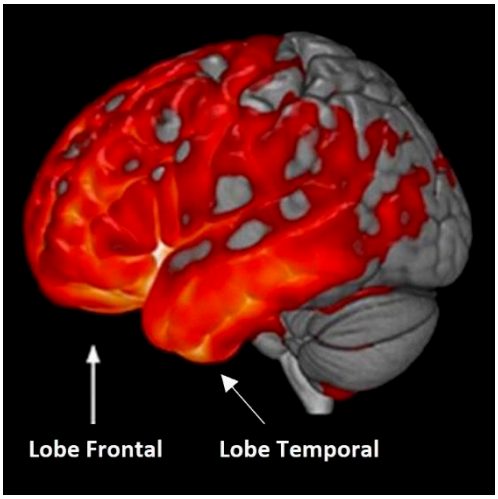


Figure 27. Zones du cerveau atteintes chez les patients DFT
Les neurones des lobes frontaux et temporaux du cortex dégénèrent chez les patients atteints de DFT. Image adaptée de wilkesadultdaycare.org.

3.1.3. SLA et DFT : un continuum

La SLA et la DFT peuvent exister sous des formes dites « pures », mais aussi appartenir à un même continuum. En effet, au niveau clinique, environ 15 % des patients atteints de SLA présentent aussi une DFT et, de même, parmi les patients atteints de DFT, 40% des sujets présentent aussi des déficits moteurs dont 15% sont compatibles avec une SLA (Ringholz et al., 2005).

De plus, de nombreuses similitudes au niveau histopathologique ont été identifiées, ainsi que des causes génétiques communes, qui sont décrites ci-dessous (cf Introduction 3.2 et Introduction 3.3).

3.2. Histopathologie

La SLA et la DFT présentent des caractéristiques histopathologiques communes, que ce soit pour les formes familiales ou sporadiques, la plus notable étant la présence d'agrégats de protéine TDP-43 (codée par le gène *TARDBP*) tronquée et phosphorylée qui s'accumule dans le cytoplasme des neurones de ces patients (Figure 28 ; Neumann et al., 2006). Ainsi, approximativement 95% des patients touchés par la SLA et 50% des patients atteints de DFT présentent des agrégats de protéines TDP-43 qui sont localisés dans les zones atteintes du cerveau et de la moelle épinière des patients (Arai et al., 2006 ; Neumann et al., 2006).

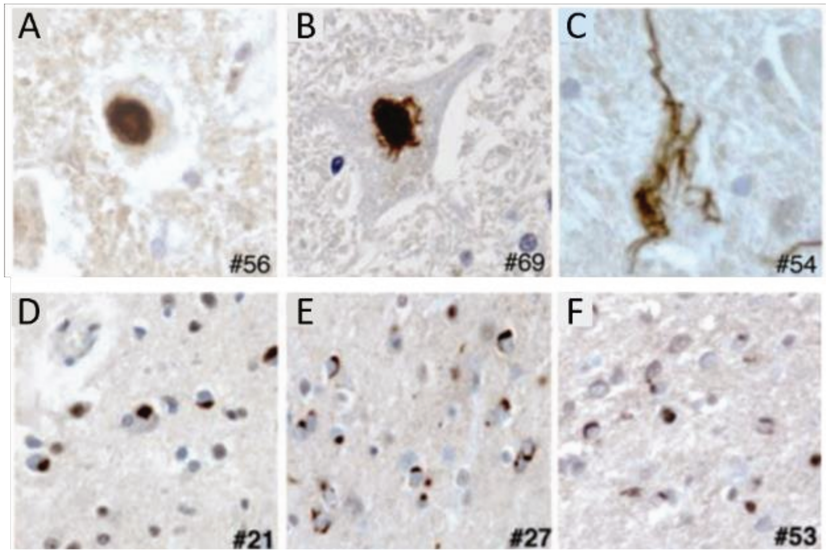


Figure 28. Présence d'agrégats de la protéine TDP-43 chez des patients SLA et DFT
IHC réalisée contre la protéine TDP-43 (**A**, **B**, **C**) sur des motoneurones de patients SLA et (**D**, **E**, **F**) dans le cortex frontal des patients DFT. Images de Neumann et al., 2006.

Les formes de SLA et de DFT ne présentant pas d'agrégats de TDP-43 affichent des agrégats d'autres protéines dont SOD1, FUS, Tau et UBQLN2 en accord avec le gène muté dans la pathologie (Kato et al., 1999 ; Deng et al., 2010 ; pour revue, Blokhuis et al., 2013).

Enfin, les cas de SLA, de DFT ou de SLA/DFT causés par une expansion de répétitions GGGGCC dans le gène *C9ORF72* présentent des agrégats de protéines composées de dipeptides répétés (protéines DPR) codées par les répétitions (Gendron et al., 2013 ; Mackenzie et al., 2013 ; Mori et al., 2013a ; Mori et al., 2013b ; Zu et al., 2013). De façon étonnante, ces agrégats de protéines DPR et les agrégats de TDP-43 sont situées dans des zones distinctes. En effet, les agrégats de protéines DPR sont majoritairement observés dans les zones du cerveau

peu ou pas affectées par la pathologie, contrairement aux inclusions de TDP-43 (Mackenzie et al., 2015).

Outre la présence d'agrégats protéiques, l'examen post-mortem des tissus des patients atteints de SLA révèle une perte des motoneurones dans la corne ventrale de la moëlle épinière et dans le cortex moteur (Figure 29), cette mort neuronale étant associée à une réaction inflammatoire dont une hyper-prolifération des astrocytes (astrogliose) et de la microglie ainsi qu'une dénervation des muscles (pour revue, Philips et Robberecht, 2011). Enfin, l'analyse par IRM des patients atteints de DFT révèle, à un stade tardif de l'évolution de la pathologie, une atrophie corticale sévère des circonvolutions des lobes temporaux et frontaux.

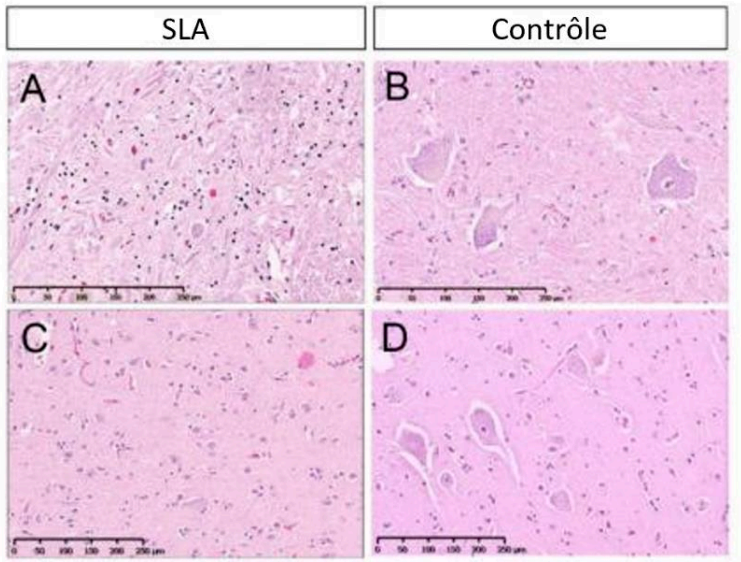


Figure 29. Perte des neurones moteurs chez un patient atteint de SLA
Perte des motoneurones de (**A**) la corne antérieure de la moelle épinière et (**B**) du cortex moteur chez des patients de SLA en comparaison à (**B**, **D**) un individu contrôle. Images de Saberi et al., 2015.

3.3. Mutations

En 1993, des mutations faux-sens dans le gène *SOD1* ont été la première cause génétique identifiées dans la SLA (Rosen et al., 1993) et ; en 1998, ce sont des mutations faux-sens dans le gène *MAPT* qui ont été décrites dans la DFT (Hutton et al., 1998). Puis, la découverte de mutations dans les gènes *TARDP*, *FUS*, *C9ORF72*, *TBK1*, etc. pouvant conduire soit à une SLA, soit à une DFT soit à une SLA-DFT a définitivement confirmé l'existence d'un continuum entre ces deux maladies (Benajiba et al., 2009 ; Kwiatkowski et al., 2009 ; Vance et al., 2009 ; DeJesus-Hernandez et al., 2011 ; Renton et al., 2011 ; Freischmidt et al., 2015). Toutefois, il est à noter que certaines mutations sont à l'origine de formes de SLA pure comme des mutations dans le gène *SOD1* ou de formes de DFT pures comme des mutations dans les gènes *MAPT* ou *PGRN* (Rosen et al., 1993 ; Beker et al., 2006).

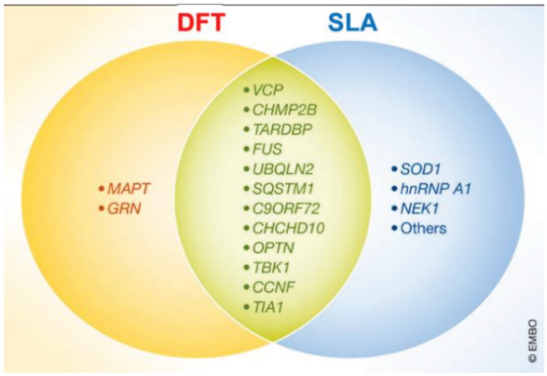


Figure 30. Gènes mutés chez les patients atteins de SLA, de DFT ou de SLA/DFT
Des mutations dans certains gènes peuvent conduire à des formes de DFT pures, des formes de SLA pures ou des cas de SLA/DFT chez une même famille ou un même patient. Image de Gao et al., 2017.

A ce jour, une quarantaine de gènes sont décrits comme étant associés à la SLA, à la DFT ou à la SLA/DFT (Tableau 3). Néanmoins, parmi ces nombreux gènes pouvant être impliqués, des mutations dans les gènes *C9ORF72*, *SOD1*, *TARDBP* et *FUS* sont responsables de plus de 50% des cas de SLAF et plus de 70% des cas de DFT sont causés par des mutation dans les gènes *C9ORF72*, *MAPT* ou *GRN* (Figure 31 ; Renton et al., 2011 ; pour revue, Olszewska et al., 2016).

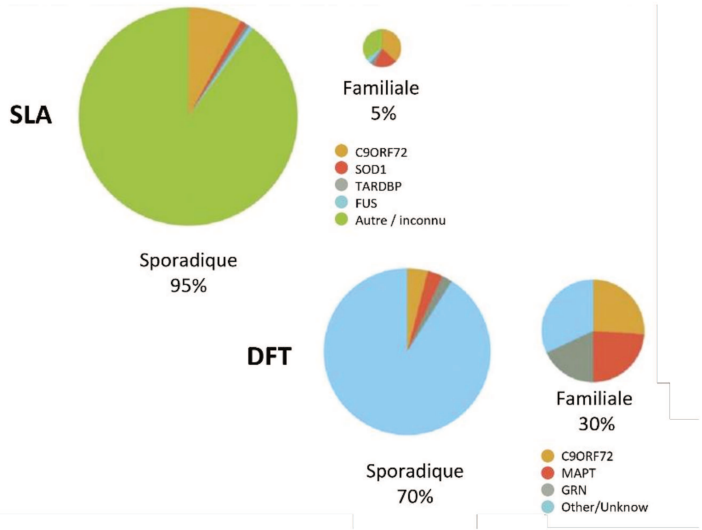


Figure 31. Principaux gènes impliqués dans la SLA et la DFT
Une mutation dans le gène *C9ORF72* est la principale cause de SLA et de DFT. Turner et al., 2017.

Année de découverte	Gène	Protéine	SLA, DFT ou SLA/DFT	Référence
1993	*SOD1*	Cu/Zn superoxide dismutase 1	SLA	Rosen et al.
1994	*NEFH*	Neurofilament, heavy polypeptide	SLA	Rooke et al.
1995	*CHMP2B*	Charged multivesicular body protein 2B	SLA/DFT	Brow et al.
1998	*MAPT*	Microtubule-associated protein	DFT	Hutton et al.
1998	*SETX*	Senataxin	SLA	Chance et al.
2000	*PSEN1*	Presenilin 1	DFT	Raux et al.
2001	*ALS2*	Alsin 2	SLA	Hadano et al. ; Yang et al.
2002	*ALS3*	Alsin 3	SLA	Hand et al.
2003	*ALS7*	Alsin 7	SLA	Sapp et al.
2003	*DCTN1*	Dynactin subunit 1	SLA	Puis et al.
2004	*PRPH*	Peripherin	SLA	Gros-Louis et al. ; Leung et al.
2004	*VAPB*	VAMP-associated protein B and C	SLA	Nishimura et al.
2006	*ANG*	Angiogenin	SLA	Greenway et al.
2006	*GRN*	Progranulin	DFT	Baker et al. ; Cruts et al.
2006	*PON1-3*	Paraoxonase 1-3	SLA	Saeed et al.
2008	*PNPLA6*	Neuropathy target esterase	SLA	Rainier et al.
2008	*TARDBP*	TAR DNA binding protein 43	SLA/DFT	Kabashi et al. ; Sreedharan et al. ; Van Deerlin et al.
2009	*CHRNA3,4,B4*	Neuronal acetylcholine receptor subunit α-3, α-4, β-4	SLA	Sabatelli et al.
2009	*ELP3*	Elongator acetyltransferase complex subunit 3	SLA	Simpson et al.
2009	*FIG4*	Lipid phosphatase FIG4	SLA	Chow et al.
2009	*FUS*	Fused in Sarcoma	SLA/DFT	Kwiatkowski et al. ; Vance et al.
2010	*ATXN2*	Ataxin 2	SLA	Elden et al.
2010	*DAO*	D-amino-acid oxidase	SLA	Mitchell et al.
2010	*OPTN*	Optineurine	SLA/DFT	Maruyama et al.
2010	*VCP*	Valosin-containing protein	SLA/DFT	Johnson et al.
2011	*C9ORF72*	Chromosome 9 open reading frame 72	SLA/DFT	DeJesus-Hernandez et al. ; Renton et al.
2011	*SIGMAR1*	Sigma non-opioid intracellular receptor 1	SLA	Al-Saif et al.
2011	*SQSTM1*	Sequestosome 1 / p62	SLA/DFT	Fecto et al.
2011	*TAF15*	TATA-box binding protein associated factor 15	SLA	Ticozzi et al.
2011	*UBQLN2*	Ubiquilin 2	SLA/DFT	Deng et al.
2012	*C19ORF12*	Chromosome 19 open reading frame 12	SLA	Deschauer et al.
2012	*EWSR1*	EWS RNA binding protein 1	SLA	Couthouis et al.
2012	*PFN1*	Profilin-1	SLA	Wu et al.
2012	*SPG11*	Spastic Paraplegia 11	SLA	Daoud et al.
2012	*UBQLN1*	Ubiquilin 1	SLA/DFT	Gonzalez-Perez et al.
2013	*ARHGEF28*	Rho GEF 28	SLA	Droppelmann et al.
2013	*ERBB4*	Erb-b2 receptor tyrosine kinase 4	SLA	Takahashi et al.
2013	*HNRNPA1*	Heterogenous nuclear ribonucleoprotein A1	SLA	Kim et al.
2013	*HNRNPA2B1*	Heterogenous nuclear ribonucleoprotein A2/B1	SLA	Kim et al.
2014	*CHCHD10*	Coiled-coil-helix-coiled-coil-helix domain containing protein 10	SLA/DFT	Bannwarth et al.
2014	*MATR3*	Matrin 3	SLA	Johnson et al.
2014	*SS18L1*	Calcium-responsive transactivator	SLA	Teyssou et al.
2014	*TUBA4A*	A-tubulin 4A chain	SLA	Smith et al.
2015	*GLE1*	GLE1 RNA export mediator	SLA	Keneb et al.
2015	*TBK1*	TANK-binding kinase 1	SLA/DFT	Freischmidt et al. ; Cirulli et al.
2016	*C21ORF2*	Chromosome 21 open reading frame 2	SLA	Van Rheenen et al.
2016	*CCNF*	Cyclin F	SLA/DFT	Williams et al.
2016	*NEK1*	Serine/threonine kinase NIMA protein 1	SLA	Brenner et al.
2016	*SCFD1*	Sec1 family domain containing protein 1	SLA	Van Rheenen et al.
2017	*TIA1*	Intracellular antigen-1	SLA/DFT	Mackenzie et al.
2018	*KIF5A*	Kinesin heavy chain isoform 5A	SLA	Brenner et al. ; Nicolas et al.
2019	*GLT8D1*	Glycosyltransferase 8 domain conteining 1	SLA	Cooper-Knock et al.
2019	*LGALSL*	Galectin like prtein	SLA	Gelfman et al.
2019	*DNAJC7*	DnaJ Heat Shock Protein Family (Hsp40) Member C7	SLA	Farhan et al.

Tableau 3. Gènes mutés responsables de la SLA, la DFT ou la SLA/DFT
Protéines codées par des gènes pouvant être mutés dans des cas de SLA, des cas de DFT ou des cas de SLA/DFT classées par ordre de découverte. Adapté de Mejzini et al., 2019 et Olszewska et al., 2016.

3.4. Expansion d'hexa-nucléotides GGGGCC dans le gène *C9ORF72*

En 2011, la cause génétique principale de SLA et de DFT a été identifiée ; il s'agit d'une expansion de répétitions d'hexa-nucléotides GGGGCC dans le premier intron du gène *C9ORF72* (DeJesus-Hernandez et al., 2011 ; Renton et al., 2011). En effet, en Europe et en Amérique du Nord, cette expansion de répétitions est retrouvée dans 80 à 90% des cas de SLA/DFT familiaux, 20 à 50% des cas de SLA familiaux, 10 à 30% des cas de DFT familiaux et en moindre proportion, de 5 à 10%, chez les cas sporadiques (Majounie et al., 2012). Dans la population générale, le nombre de répétitions de microsatellites GGGGCC est compris entre 2 et 70 alors que chez les patients atteints de SLA/DFT, l'expansion peut s'étendre des centaines voire des milliers de fois (DeJesus-Hernandez et al., 2011 ; Benussi et al., 2014 ; Xi et al., 2015).

La sévérité des symptômes et le phénomène d'anticipation observé dans d'autres pathologies dues à des expansions de répétitions ne semblent pas se produire dans la SLA/DFT. En effet, les différentes études tentant de mettre en corrélation la taille de l'expansion et les symptômes de la maladie sont en désaccord (Benussi et al., 2014 ; Suh et al., 2015 ; Gijsenlinck et al., 2016). Les variations entre ces différentes études peuvent peut-être s'expliquer par la grande instabilité de l'expansion de répétitions GGGGCC entre les différents tissus et même entre les différentes zones du cerveau au sein d'un même individu (Beck et al., 2013). Néanmoins, il a été démontré que la toxicité des expansions de répétitions GGGGCC augmente avec l'âge des patients et la pénétrance chez les patients porteurs de la mutation est de 0% à l'âge de 35 ans, de 50% à 58 ans et une pénétrance quasi complète à plus de 90% à 80 ans (Majounie et al., 2012 ; Benussi et al., 2013).

A ce jour, le mécanisme par lequel l'expansion de répétitions GGGGCC dans le gène *C9ORF72* est toxique reste controversé et une contribution des 3 mécanismes, à savoir une toxicité au niveau ADN, ARN et protéique, ont été proposés et sont présentés ci-dessous (Figure 32).

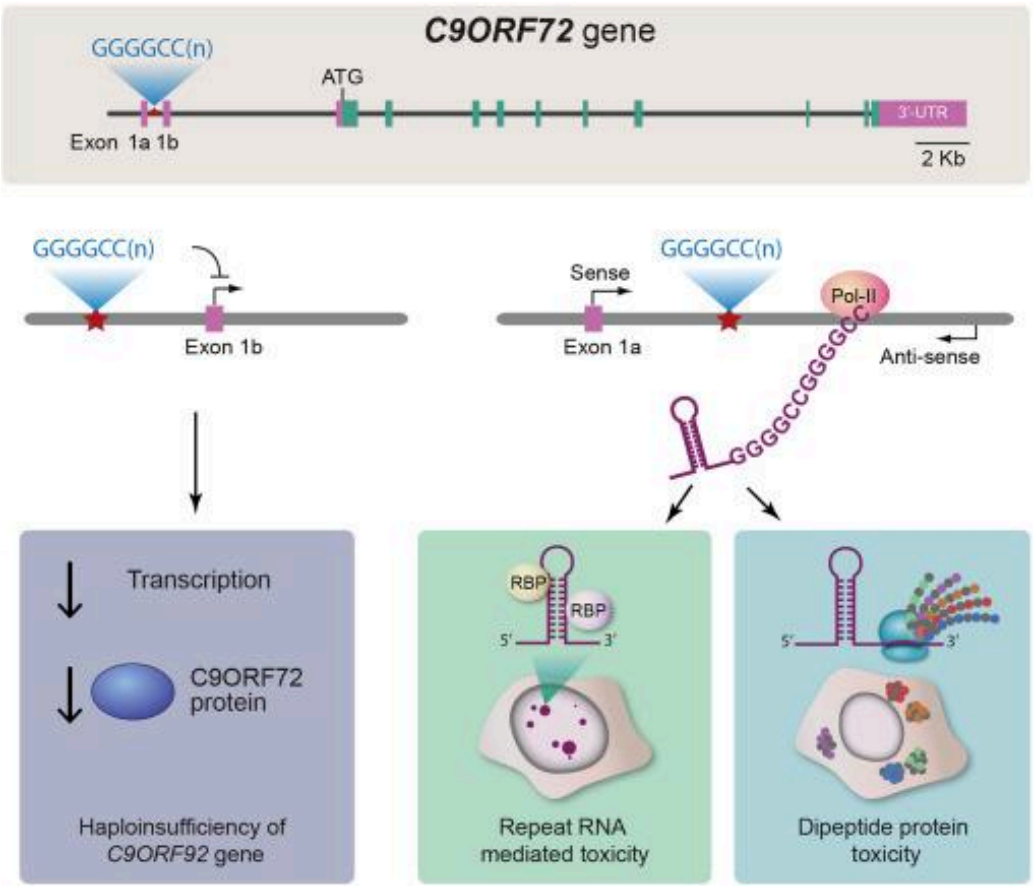


Figure 32. Mécanismes de toxicité de l'expansion GGGGCC dans le gène *C9ORF72*
L'expansion de répétitions GGGGCC dans le gène *C9ORF72* inhiberait l'expression de la protéine C9ORF72 conduisant à sa perte de fonction partielle. Ces répétitions seraient également transcrites en ARN séquestrant des protéines spécifiques et ; ces ARN contenant les répétitions seraient traduits en protéines DPR. Gitler et Tsuji., 2016.

3.4.1. Hypothèse de la perte de fonction de la protéine C9ORF72

Le gène *C9ORF72* pour « Chromosome 9 Open Reading Frame 72 » est composé de 11 exons avec 2 promoteurs alternatifs situés dans les exons 1a et 1b. Suivant l'épissage alternatif de l'intron 5, ce gène peut coder pour 2 isoformes protéiques : une forme longue où la protéine C9ORF72, composée de 481 acides aminées et de poids moléculaire de 54 kDa, est encodée par les exons 2 à 11 et ; une forme courte où la protéine c-C9ORF72, composée de 222 acides aminés et de poids moléculaire d'environ 24 kDa, est codée par les exons 2 à 5 (Figure 33 ; DeJesus-Hernandez et al., 2011 ; Renton et al., 2011). Cette insoforme courte est instable en transfection cellulaire et, le développement d'anticorps spécifiquement dirigés contre ces 2 isoformes montrent que l'isoforme longue est largement prédominante dans le cerveau humain et murin (Frick et al., 2018). Ce gène n'est pas présent chez les insectes, les plantes et la plupart des champignons, mais sa séquence est hautement conservée chez les différentes espèces le possédant (Levine et al., 2013).

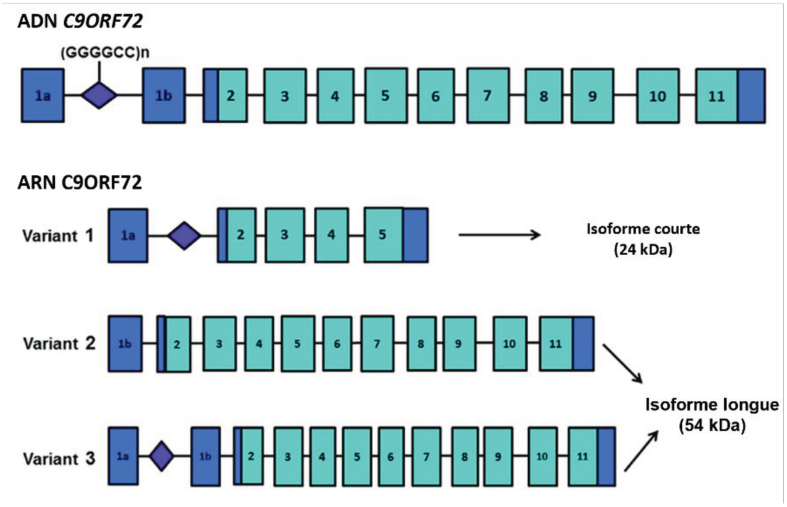


Figure 33. Epissage alternatif de C9ORF72 produisant 2 isoformes protéiques
Le gène *C9ORF72* est composé de 11 exons et 2 promoteurs alternatifs situés dans les exons 1a et 1b. Selon l'épissage de l'intron 5, ce gène peut coder pour 2 isoformes protéiques : une courte de 24 kDa ou une longue de 54 kDa. Barker et al., 2017.

Dès la découverte de l'expansion de répétitions GGGGCC dans le gène *C9ORF72*, des niveaux plus faibles d'ARNm C9ORF72 ont été décrits (DeJesus-Hernandez et al., 2011 ; Renton et al., 2011), suggérant une réduction de l'expression de la protéine C9ORF72. Cette diminution de la transcription s'explique par le fait que l'expansion de répétitions GGGGCC conduit à une hyper-méthylation des îlots CpG en 5' de l'expansion (Xi et al., 2013 ; Xi et al., 2014) ainsi que par une tri-méthylation excessive des histones H3 et H4 liées aux répétitions GGGGCC (Belzil

et al., 2013). La diminution d'expression de C9ORF72 a été confirmée au niveau protéique, puisque les patients présentent une diminution d'environ 50% de l'expression de la protéine C9ORF72 spécifiquement dans les zones atteintes par la DFT, à savoir le cortex frontal (Waite et al., 2014 ; Xiao et al., 2015 ; Saberi et al., 2018), mais également dans des zones non-atteintes comme le cortex occipital (Saberi et al., 2018) ou le cervelet (Frick et al., 2018). Néanmoins, la seule perte de fonction de la protéine C9ORF72 ne parait pas suffisante pour expliquer le développement de la SLA/DFT car aucune mutation perte de fonction dans le gène *C9ORF72* n'a été reportée chez des patients, exceptée une mutation non-sens, c'est-à-dire l'insertion d'un codon stop prématurée chez un seul patient (Liu et al., 2016a).

Enfin, bien que la diminution d'expression de la protéine C9orf72 conduit à des altérations locomotrices chez le poisson zèbre (Ciura et al., 2013 ; Butti et al., 2021) et que la surexpression de C9ORF72 dans des motoneurones différenciés à partir de cellules iPS de patients abolit la mort neuronale observée de ces cellules (Shi et al., 2018), la déplétion de C9orf72 chez les souris ne provoque pas de SLA/DFT mais conduit à un phénotype inflammatoire (Atanasio et al., 2016 ; Burberry et al., 2016 ; Jiang et al., 2016 ; O'Rourke et al., 2016 ; Sudria-Lopez et al., 2016).

3.4.1.1. *Fonction de la protéine C9ORF72*

Des analyses bio-informatiques basées sur l'alignement de séquences permettant des prédictions structurales ont permis de mettre en évidence que la protéine C9ORF72 contient des domaines DENN pour « Differentially Expressed in Normal and Neoplastic cells » caractéristiques des facteurs d'échange GFP/GTP (GEF) pour les protéines Rab GTPases (Zhang et al., 2012 ; Levine et al., 2013). Ces domaines sont composés de 400 à 500 acides aminés et forment 2 lobes pouvant se lier l'un à l'autre ainsi qu'aux Rab GTPases. De plus, différentes études basées sur les interactions protéine-protéine ont permis d'établir que la protéine C9ORF72 forme un complexe avec deux autres protéines : SMCR8 et WDR41, qui sera appelé par la suite complexe CSW (Amick et al., 2016 ; Sellier et al., 2016 ; Sullivan et al., 2016 ; Ugolino et al., 2016 ; Yang et al., 2016). WDR41 (« WD repeat containing protein 41 ») est une protéine de 52 kDa composée de 6 domaines WD40. SMCR8 (« Smith Magenis syndrome Chromosome Region candidate 8 ») est une protéine de 105 kDa qui contient, tout

comme C9ORF72, deux domaines DENN (Zhang et al., 2012). La fonction de WDR41 et de SMCR8 sont inconnues.

Les protéines C9ORF72 et SMCR8 pourraient intervenir dans la régulation des Rab (« Ras-related proteins in brain »), qui font partie de la superfamille des petites GTPases Ras, comprenant également les protéines des familles Ras, Rho, Arf, Arl et Ran (Figure 34). Les Arf GTPases au nombre de 6, les Arl GTPases au nombre de 18 et les Rab GTPases forment une famille d'environ 80 protéines chez l'Homme. Ces protéines Arf, Arl et Rab sont localisées sur la face cytosolique des membranes et sont impliquées dans de nombreuses étapes du trafic vésiculaire comme la formation de vésicules, leur migration le long des réseaux d'actine et de tubuline ainsi que leur fusion à une vésicule ou une membrane cible. Comme toutes les petites GTPases Ras, les protéines Rab, Arl et Arf existent sous 2 conformations distinctes, une forme où elles sont liées au GTP et dites actives et ; une forme où elles sont liées au GDP et dites inactives. Le passage de l'état inactif à l'état actif est catalysé par les facteurs d'échange GDP/GTP (GEF) alors que l'inactivation des Rab, Arl et Arf est accélérée par des protéines GAP (« GTPase-activating protein ») (Figure 35 ; pour revue, Stenmark et Olkkonen, 2001).

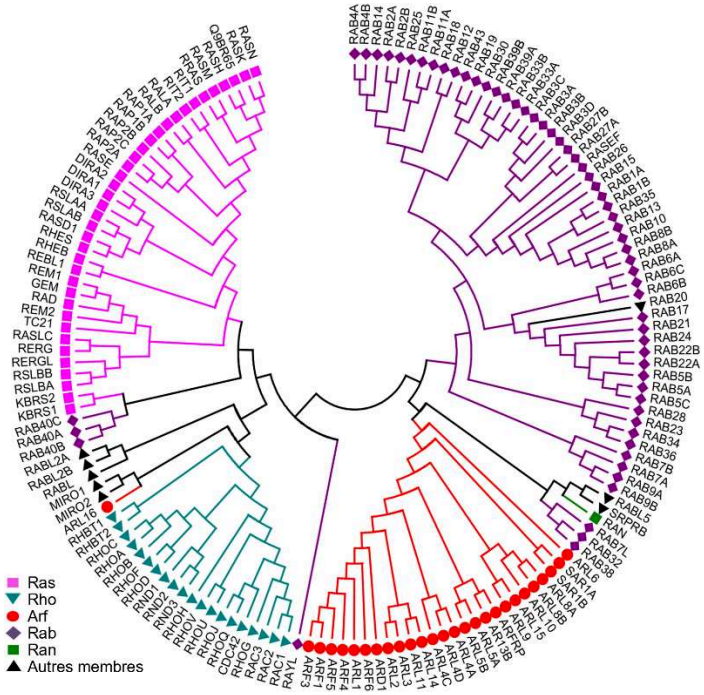


Figure 34. La superfamille des GTPases Ras
La superfamille des Ras GTPases est composée des Ras, des Rho, des Arf, des Rab et des Ran. Image de Qu et al., 2019.

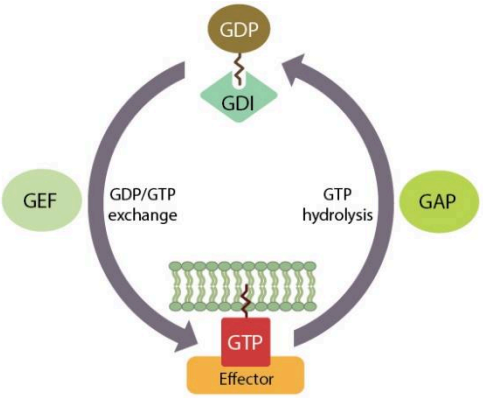


Figure 35. Conformation des petites GTPases Ras
Les GTPases Ras sont sous forme inactives lorsqu'elles lient le GDP et sous forme actives lorsqu'elles lient le GTP, où elles interagissent avec des effecteurs spécifiques. Le passage d'un état à l'autre fait intervenir des facteurs GEF ou GAP. Image de mechanobio.info.

A la suite de ces découvertes, le complexe C9ORF72-SMCR8-WDR41 a été décrit comme régulant de nombreuses Rab GTPases comme Rab8a et Rab39b (Sellier et al., 2016 ; Yang et al., 2016), Rab1a (Webster et al., 2016) et / ou Rab5 (Farg et al., 2014 ; Shi et al., 2018) afin de réguler l'autophagie ; Rab7 et Rab11 pour réguler l'endocytose (Farg et al., 2014) ; ou encore Rab3 pour réguler l'exocytose (Frick et al., 2018). Néanmoins, la résolution récente de la structure par Cryo-EM du complexe C9ORF72-SMCR8-WDR41 révèle que ce complexe n'agit non pas comme un facteur d'échange GDP/GTP (GEF) permettant d'activer les Rab GTPases, mais comme une protéine GAP accélérant leur activité GTPase, inactivant alors les Rab et les Arf GTPases (Figure 36A ; Su et al., 2020 ; Tang et al., 2020). En effet, ce complexe ressemble à un autre complexe, le complexe FLNC-FNIP2, qui agit comme une GAP pour des petites GTPases particulières régulant l'activité de mTOR au lysosome et qui existent uniquement sous forme de dimère : RagC/D ou RagB/D. La comparaison des structures de ces 2 complexes révèle que, tout comme FLNC-FNIP2, SMCR8 possède un résidu arginine (Arg147) critique pour son activité GTPasique envers Rab8a et Rab11 (Figure 36A, 36B) Tang et al., 2020) et / ou envers Arf1, Arf5 et Arf6 (Su et al., 2020 ; Su et al., 2021).

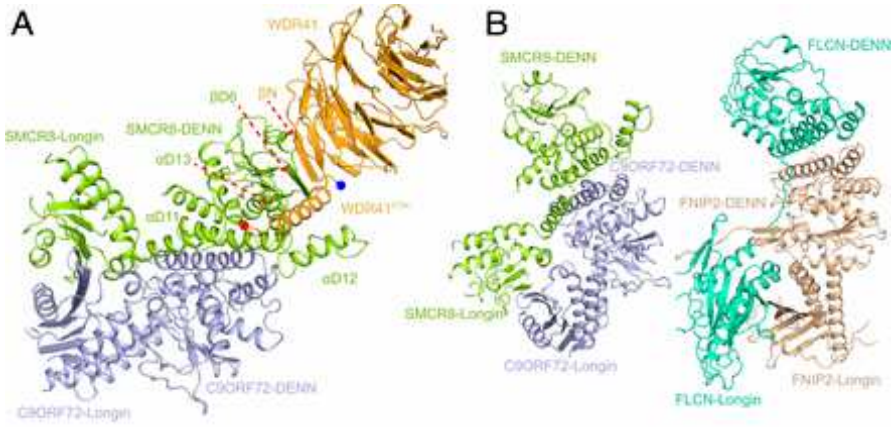


Figure 36. Structure des complexes C9ORF72-SMCR7-WDR41 et FLNC-FNIP2
(A) Structure résolue par cryo-EM du complexe C9ORF72-SMCR7-WDR41. (B) Comparaison des structures des complexes C9ORF72-SMCR7-WDR41 et FLNC-FNIP2. Images de Tang et al., 2020.

3.4.1.2. *Implication de la protéine C9ORF72 dans l'autophagie*

La protéine C9ORF72 a été identifiée comme étant impliquée dans la régulation de l'autophagie, je vais décrire ce mécanisme de dégradation ci-dessous.

3.4.1.2.1. L'autophagie

L'autophagie est un mécanisme catabolique intracellulaire permettant la dégradation des agrégats protéiques, des bactéries, des organelles dysfonctionnelles comme des mitochondries, des peroxysomes, etc., et qui sont des éléments trop massifs pour être dégradés par le protéasome (pour revue, Glick et al., 2010). Ce mécanisme est conservé de la levure à l'Homme et est essentiel pour éliminer des structures, des micro-organismes et des organelles dysfonctionnelles potentiellement toxiques, mais aussi pour fournir de l'énergie à la cellule en recyclant des composants cellulaires. Il existe 3 formes majeures d'autophagie (Figure 37) :

- L'autophagie médiée par les protéines chaperonnes (CMA) où la protéine chaperonne Hsc70 reconnaît et interagit avec une séquence KFERQ présente sur les protéines cibles à dégrader, conduisant au transport et au transfert de ces protéines dans la lumière du lysosome directement à travers sa membrane (pour revue, Tekirdag et Cuervo, 2018) ;
- La micro-autophagie où le lysosome, via une invagination ou une protrusion, engloutit directement des composants présents dans le cytoplasme comme des protéines ou des organelles (pour revue, Li et al., 2012) ;
- La macro-autophagie ou autophagie où une vésicule composée d'une double membrane enveloppe des protéines ou des organelles, formant alors l'autophagosome qui est transporté et fusionne avec le lysosome où le matériel est dégradé. Ce mécanisme est décrit plus en détail ci-dessous.

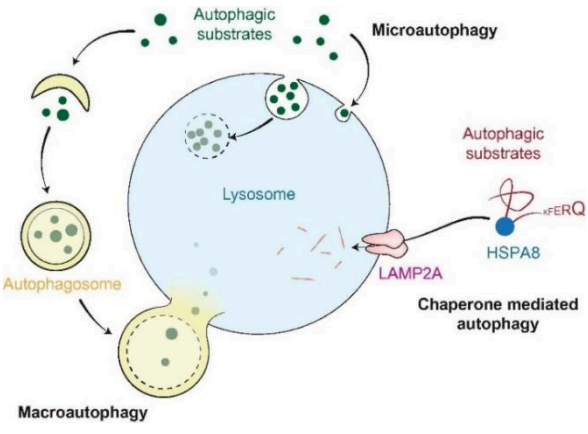


Figure 37. Les 3 formes d'autophagie
L'autophagie est classée en 3 mécanismes : l'autophagie médiée par les protéines chaperonnes qui dégrade spécifiquement les protéines contenant une séquence KFERQ ; la micro-autophagie où le lysosome dégrade directement des composants cytoplasmiques et ; la macro-autophagie faisant intervenir une double membrane appelée autophagosome qui englobe le matériel à dégrader puis, fusionne au lysosome. Image de Hazari et al., 2020.

Le déroulement de l'autophagie fait appel à plusieurs étapes successives dont l'initiation et la nucléation conduisant à la formation d'une membrane isolée appelée le phagophore. Lors de la maturation, ce phagophore s'allonge autour des éléments à dégrader pour former l'autophagosome, une vésicule close qui transporte ces éléments jusqu'aux lysosomes avec lesquels il fusionne. Les lysosomes sont des organelles contenant des hydrolases acides permettant la dégradation du matériel se trouvant dans l'autophagosome. Cette voie dépend d'une cascade de signalisation faisant intervenir de nombreuses protéines, présentées ci-dessous.

Un régulateur majeur de l'autophagie est le complexe mTORC1 composé de 5 protéines : mTOR, RAPTOR, PRAS40, DEPTOR et mLST8. En condition basale, le complexe mTORC1 est actif et phosphoryle le complexe ULK1 qui est alors inactif. Lors d'une privation en acides aminés, le complexe mTORC1 est inactivé, ce qui aboutit à l'activation du complexe ULK, initiateur de l'autophagie (Figure 38 ; pour revue, Parzych et Klionsky, 2014).

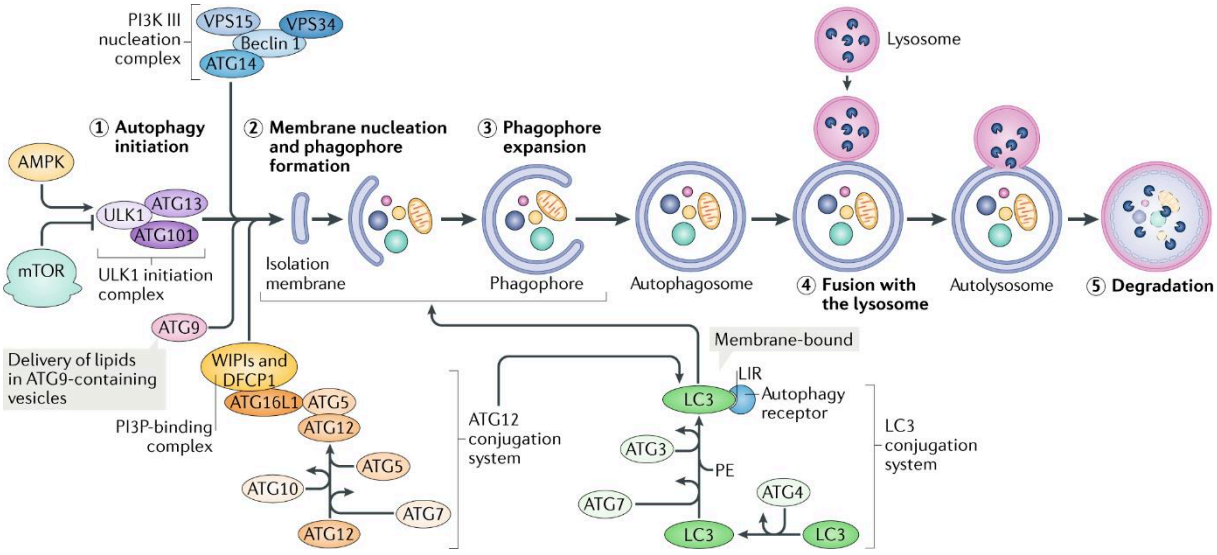


Figure 38. Protéines impliquées dans la cascade de signalisation dans l'autophagie
L'initiation fait intervenir le complexe ULK1 qui phosphoryle et active le complexe PI3KC3 qui permet la génération de PI3P. Ces PI3P recrutent les protéines DFCP1 et WIPI qui vont permettre, via l'intervention des systèmes de conjugaison, l'ancrage des protéines LC3 à l'autophagosome. Cet autophagosome mature fusionne alors aux lysosomes où son contenu est dégradé. Image de Hansen et al., 2018.

Le complexe ULK est composé de 4 protéines : la protéine kinase ULK1 (ou ULK2), ATG13, FIP200 et ATG101. Le complexe ULK actif s'autophosphoryle et recrute le complexe PI3KC3 (« class III phosphatidylinositol 3-kinase complex 3 ») composé de VPS15, VPS34, Beclin-1 (BECN1) et ATG14L et l'active en phosphorylant BECN1 et VPS15 (Mercer et al., 2021). La

protéine BECN1 phosphorylée active VPS34 qui phosphoryle les phosphoinositides (PI) en phosphatidylinositol-3-phosphate (PI3P) de la membrane du réticulum endoplasmique, générant ainsi une membrane isolée, donnant naissance au phagophore (Figure 38 ; pour revue, Parzych et Klionsky, 2014).

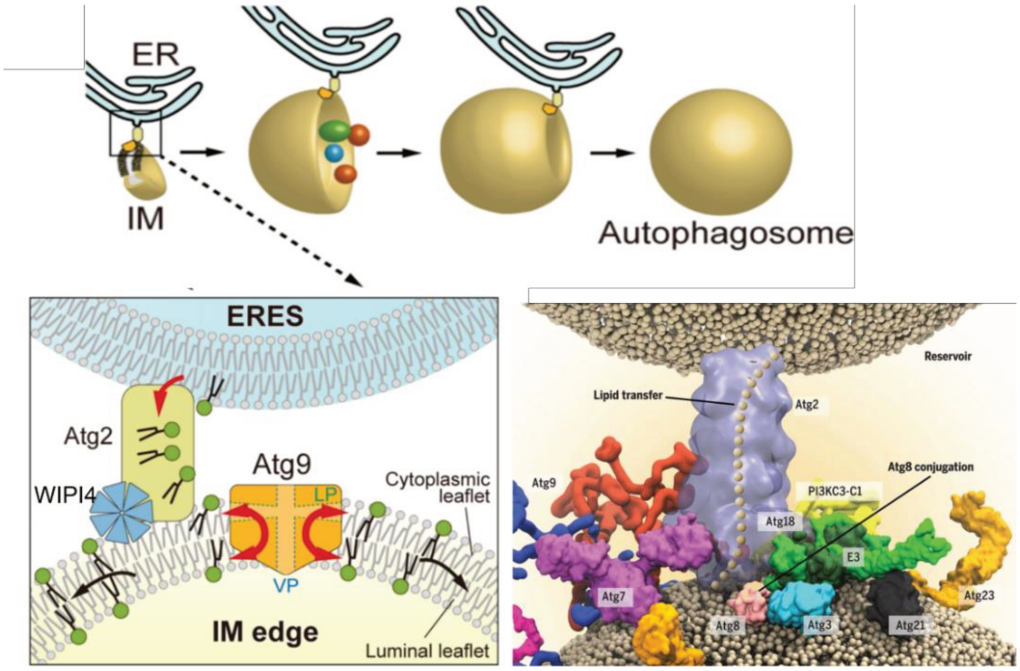


Figure 39. Incorporation des lipides au niveau de l'autophagosome
WIPI4 recrute le canal lipidique Atg2 qui transfère les lipides depuis le réticulum endoplasmique jusqu'à l'autophagosome en formation. Atg9 est une scramblase qui répartit les lipides entre les 2 membranes de l'autophagosome. Images de Matoba et Noda, 2020 et Sawa-Makarska et al., 2020.

La liaison des protéines DFCP1 et WIPI2 / WIPI4 aux PI3P permet la formation de la double membrane et l'élongation du phagophore. En effet, la protéine WIPI4 permet le recrutement du canal lipidique ATG2 qui transfert les lipides du réticulum endoplasmique jusqu'à l'autophagosome en formation (Figure 39 ; Chowdhury et al., 2018 ; Kotani et al., 2018 ; Osawa et al., 2019 ; Valverde et al., 2019). La distribution des lipides entre les 2 feuillets lipidiques de la membrane de l'autophagosome est assurée par la scramblase ATG9 (Figure 39 ; Maeda et al., 2020 ; Matoba et al., 2020 ; Sawa-Makarska et al., 2020). La protéine WIPI2, quant à elle, interagit avec ATG16L1 conduisant au recrutement du premier système de conjugaison composé des protéines ATG12, ATG5 et ATG16L1 où ATG12 est préalablement conjugué à ATG5 par l'action de ATG7, une protéine activatrice de type E1 et, ATG10, une enzyme de conjugaison de type E2 (Figure 40). Le complexe ATG16L1 agit comme une ubiquitine ligase de type E3 et attire les protéines de la famille LC3 (LC3A, LC3B, LC3C, GABARAP,

GABARAPL1 et GABARAPL2), appelées Atg8 chez la levure, à l'autophagosome en formation (pour revue, Parzych et Klionsky, 2014). Néanmoins, avant d'être incorporées à la double membrane de l'autophagosome, les protéines LC3 sont clivées à leur extrémité C-terminale par la protéase ATG4 puis, sont transférées sur ATG7 puis ATG3, permettant alors la liaison au complexe ATG16L1. Enfin, les protéines LC3 sont lipidées (LC3-II) par leur liaison à une phosphatidyléthanolamine (PE) et ancrées à la double membrane de l'autophagosome (Figure 40).

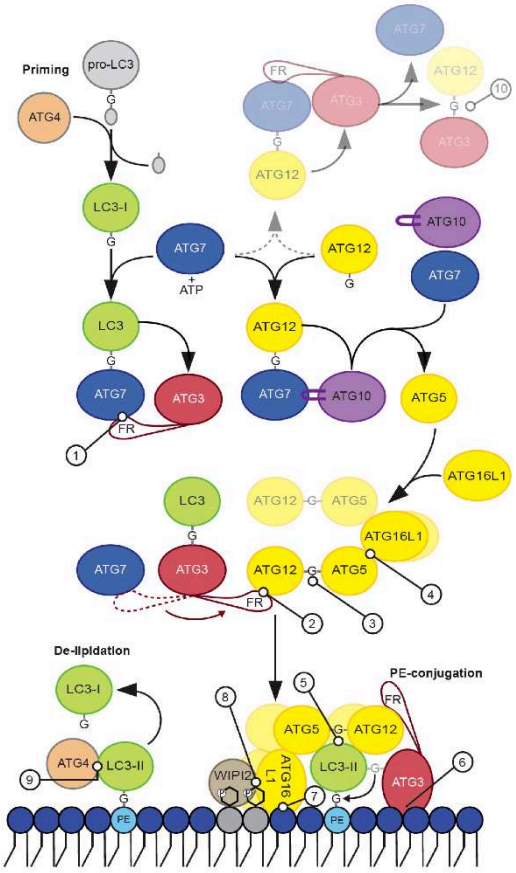


Figure 40. Systèmes de conjugaison impliqués dans le recrutement des protéines LC3 à l'autophagosome
LC3 est clivé par ATG4, transféré à ATG7 puis à ATG3. ATG7 et ATG10 permettent la liaison entre ATG12 et ATG5, qui vont alors se lier à ATG16L1. ATG3 liée à LC3 va alors intergir avec ATG12, conduisant à la lipidation de LC3 et à son ancrage à la double membrane de l'autophagosome. Schéma de Lystad et Simonsen, 2019.

Une fois refermé, l'autophagosome est transporté par les microtubules grâce à différentes Rab GTPases dont Rab7, vers les lysosomes où ces deux types de vésicules fusionnent par l'action de plusieurs protéines SNARE, VAMP7/8, SNAP29 et STX17, formant alors l'auto-lysosome. Le contenu de l'autophagosome est alors dégradé par les protéases acides lysosomales et les acides aminés et autres produits obtenus après dégradation peuvent être réutilisés par la cellule (pour revue, Parzych et Klionsky, 2014).

Il est à noter que lorsque l'autophagie est nécessaire pour dégrader spécifiquement une organelle, des agrégats protéiques ou des bactéries, on parle alors d'autophagie sélective. Ce mécanisme est induit par la poly-ubiquitinylation en lysine 63 du substrat à dégrader (pour

revue, Shaid et al., 2013). Les protéines ubiquitine sont alors reconnues par des protéines adaptatrices, appelées récepteurs autophagiques, qui font le lien entre les protéines ubiquitinylées et les protéines LC3-II ancrées à la double membrane de l'autophagosome. De multiples récepteurs autophagiques ont été décrits, dont p62, OPTN, NBR1, NDP52, TAX1BP1, etc. et sont relativement spécifiques au substrat à dégrader mais peuvent aussi être redondants entre eux. Par exemple, p62 et TAX1BP1 sont principalement impliquées dans la dégradation des agrégats protéiques. Enfin, l'autophagie sélective est activée par l'autophosphorylation de la protéine kinase TBK1 qui est recrutée au site de formation de l'autophagosome par l'OPTN où elle phosphoryle alors de nombreuses protéines comme Rab7 mais aussi OPTN et p62, ce qui renforce leur interaction avec les substrats ubiquitinylés à dégrader et les protéines LC3-II (pour revue, Shaid et al., 2013). De manière intéressante, des mutations dans les gènes *OPTN*, *UBQNL2*, *SQSTM1* (codant pour la protéine p62), et *TBK1* sont associées à la SLA et à la DFT (Maruyama et al., 2010 ; Deng et al., 2011 ; Fecto et al., 2011 ; Cirulli et al., 2015 ; Freischmidt et al., 2015).

3.4.1.2.2. Rôle potentiel de C9ORF72 dans l'autophagie

La protéine C9ORF72 a été identifiée comme pouvant jouer un rôle dans l'autophagie et ceci, à de nombreuses étapes, de l'initiation de l'autophagie à la fusion des autophagosomes aux lysosomes.

Premièrement, plusieurs études montrent une fonction du complexe C9ORF72-SMCR8-WDR41 (CSW) dans l'initiation dans l'autophagie. En effet, il a été démontré que la déplétion de C9ORF72 conduit à une accumulation de puncta du récepteur autophagique p62 en motoneurones différenciés à partir de cellules iPS de patients C9-SLA/DFT (Almeida et al., 2013) et dans des lignées de cellules immortalisées (Sellier et al., 2016 ; Webster et al., 2016). De plus, l'absence de C9ORF72 altère le flux autophagique, induisant notamment avec une diminution de la lipidation de la protéine LC3b après induction de l'autophagie, suggérant que le complexe CSW participe à l'induction de l'autophagie (Sellier et al., 2016 ; Sullivan et al., 2016 ; Ugolino et al., 2016 ; Webster et al., 2016 ; Jung et al., 2017).

Deuxièmement, l'activité du complexe CSW semble régulée après l'initiation de l'autophagie, notamment par phosphorylation puisque la protéine SMCR8 est phosphorylée sur de nombreux acides aminés par diverses kinases impliquées dans l'autophagie dont ULK1 (Behrends et al., 2010 ; Sellier et al., 2016), TBK1 (Sellier et al., 2016), mTOR (Hsu et al., 2011) et AMPK (Hoffman et al., 2015). De plus, il a été montré que ce complexe interagit avec le complexe ULK1 (Behrends et al., 2010 ; Sellier et al., 2016 ; Sullivan et al., 2016 ; Webster et al., 2016 ; Yang et al., 2016 ; Jung et al., 2017). Toutefois, les études ne sont pas en accord sur la sous-unité du complexe ULK1 permettant l'interaction avec le complexe CSW ; il pourrait s'agir de FIP200 (Sullivan et al., 2016 ; Webster et al., 2016) ou de ATG101 (Yang et al., 2016). Des travaux supplémentaires suggèrent que la régulation par le complexe CSW de la protéine ULK1, ou inversement, n'est pas directe. En effet, le complexe CSW permettrait de faciliter l'interaction entre ULK1 et Rab1 pour induire l'initiation de la formation de l'autophagosome (Webster et al., 2016). De plus, la déplétion de FIP200 n'altère pas la relocalisation au lysosome du complexe CSW après induction de l'autophagie, suggérant que le complexe CSW agit en amont du complexe ULK1 (Amick et al., 2018). Ces données suggèrent une fonction du complexe CSW dans la régulation du complexe ULK1, potentiellement à la fois à l'initiation de l'autophagie, mais aussi au niveau du lysosome.

En effet, de manière intéressante, le complexe CSW est localisé au lysosome, notamment après induction de l'autophagie (Amick et al., 2016 ; Frick et al., 2018). Cette relocalisation du cytoplasme et de l'appareil de Golgi jusqu'au lysosome est dépendante de la liaison entre WDR41 et PQLC2, un transporteur lysosomal d'acides aminés cationiques (Amick et al., 2018 ; Amick et al., 2020 ; Talaia et al., 2021). De plus, de multiples altérations des lysosomes ont été décrites après une déplétion de C9ORF72, dont des altération de leur biogénèse dans des motoneurones différenciés à partir de cellules iPS de patients C9-SLA/DFT (Shi et al., 2018), des altérations de leur taille et de leur localisation en cellules HeLa (Amick et al., 2016), de leur nombre en MEF développé à partir de souris KO pour *C9orf72* (Ugolino et al., 2016) et de leur acidification en macrophages purifiés à partir de souris KO *C9orf72* ou *Smcr8* (Shao et al., 2020).

La protéine mTOR est un régulateur majeur de l'autophagie et est localisée au lysosome, plusieurs études se sont donc penchées sur une possible régulation de mTOR par le complexe CSW (Amick et al., 2016). Ainsi, plusieurs études ont rapporté que la déplétion de C9ORF72 ou de SMCR8 ne modifie pas la localisation de mTOR mais modifie son activité. En effet, le complexe CSW semble nécessaire pour l'activation de mTOR, suggérant que C9ORF72 est un régulateur négatif de l'autophagie (Ugolino et al., 2016 ; Jung et al., 2017 ; Amick et al., 2018 ; Wang et al., 2020), bien qu'une hyperactivation de mTOR ait été observée dans une étude portant sur des macrophages de souris KO pour *C9orf72* ou *Smcr8* (Shao et al., 2020). Une des cibles de mTOR est le facteur de transcription TFEB qui, lors de l'activation de l'autophagie, est transloqué dans le noyau où il active la transcription de nombreux gènes *Atg* (« Autophagy-related gene »), impliqués dans l'autophagie ainsi que dans la biogénèse des lysosomes. Il est à noter que la déplétion de C9ORF72 conduit à une sous-activation de mTOR, ce qui aboutit à une moindre phosphorylation du facteur TFEB qui devient alors actif et nucléaire (Ugolino et al., 2016 ; Wang et al., 2020).

En conclusion, il est désormais établi que la protéine C9ORF72 est impliqué dans l'autophagie. Toutefois, comment et à quelle étape le complexe CSW intervient reste encore discuté.

3.4.2. Hypothèse d'un gain de fonction de l'ARN

Le deuxième mécanisme pouvant expliquer la toxicité des expansions de répétitions dans le gène *C9ORF72* est leur transcription sens en ARN contenant des répétitions GGGGCC, ainsi que leur transcription antisens en ARN contenant des répétitions CCCCGG. Ces ARN s'accumulent sous forme de foci dans les noyaux et / ou le cytoplasme des cellules des patients SLA/DFT (Figure 41 ; DeJesus-Hernandez et al., 2011 ; Gendron et al., 2013 ; Lagier-Tourenne et al., 2013 ; Lee et al., 2013 ; Mori et al., 2013b ; Mizielinska et al., 2013 ; Zu et al., 2013) ainsi que des neurones différenciés partir de cellules iPS de patients (Almeida et al., 2013 ; Donnelly et al., 2013 ; Sareen et al., 2013). Des études *in vitro* sur la structure de ces ARN ont mis en évidence qu'ils étaient capables de former des quadruplexes-G (Reddy et al., 2013 ; Conlon et al., 2016) et lier de nombreuses protéines de liaison à l'ARN comme hnRNPA1, Purα, ADARB2, SRSF1, hnRNP-H, la nucléoline, SRSF2, ALYREF, hnRNPA3, SRSF1 et SFPQ (Donnelly et al., 2013 ; Lee et al., 2013 ; Reddy et al., 2013 ; Sareen et al., 2013 ; Xu et al., 2013 ; Cooper-Knock et al., 2014 ; Haeusler et al., 2014 ; Conlon et al., 2016 ; Davidson et al., 2017 ; Hautbergue et al., 2017 ; Bajc Česnik et al., 2019).

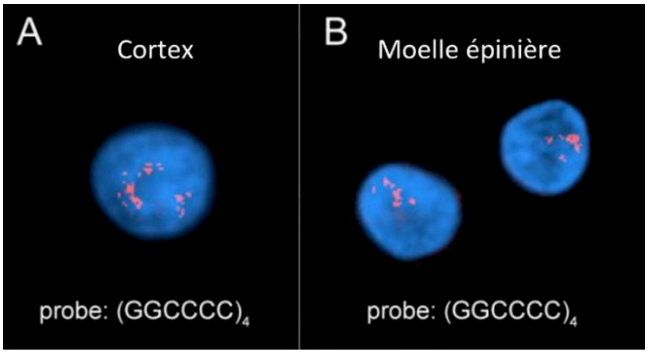


Figure 41. Foci d'ARN compose de répétitions GGGGCC chez les patients SLA/DFT
(**A**, **B**) Présence de foci d'ARN sens composés de répétitions GGGGCC chez un patient SLA/DFT dans des cellules de (**A**) cortex et (**B**) de moelle épinière. Images de DeJesus-Hernandez et al., 2011.

Toutefois, contrairement aux dystrophies myotoniques, la titration de ces protéines de liaison à l'ARN semble faible et ne conduit qu'à des altérations subtiles de l'épissage chez les patients atteints de SLA/DFT ; même si une contribution particulière du facteur d'épissage hnRNP-H soit suggérée chez les patients porteurs de la mutation dans le gène *C9ORF72* ainsi que chez les patients sporadiques (Prudencio et al., 2015 ; Conlon et al., 2016 ; Conlon et al., 2018).

3.4.3. Hypothèse d'un gain de fonction de protéines composées de dipeptides répétés

Le troisième mécanisme potentiel pour expliquer la toxicité de l'expansion de répétitions GGGGCC est la traduction des ARN sens et antisens en protéines composées de dipeptides répétés, également appelées protéines DPR. Cette traduction se ferait par le mécanisme « RAN » (cf Introduction 1.4.1.4) et permettrait de produire 5 protéines DPR différentes selon le cadre de lecture ouvert (Ash et al., 2013 ; Gendron et al., 2013 ; Mori et al., 2013a ; Zu et al., 2013). Ainsi, les répétitions GGGGCC seraient potentiellement traduites en protéines composées de polyGlycine-Alanine (polyGA), polyGlycine-Proline (polyGP) et polyGlycine-Arginine (polyGR). Les répétitions CCCCGG sont, elles, traduites en protéines composées de polyProline-Alanine (polyPA) et polyProline-Arginine (polyPR), le dernier cadre de lecture possible codant pour des protéines composées de polyProline-Glycine ayant la même composition en acides aminés que la protéines polyGlycine-Proline provenant de la traduction de l'ARN sens (Figure 42).

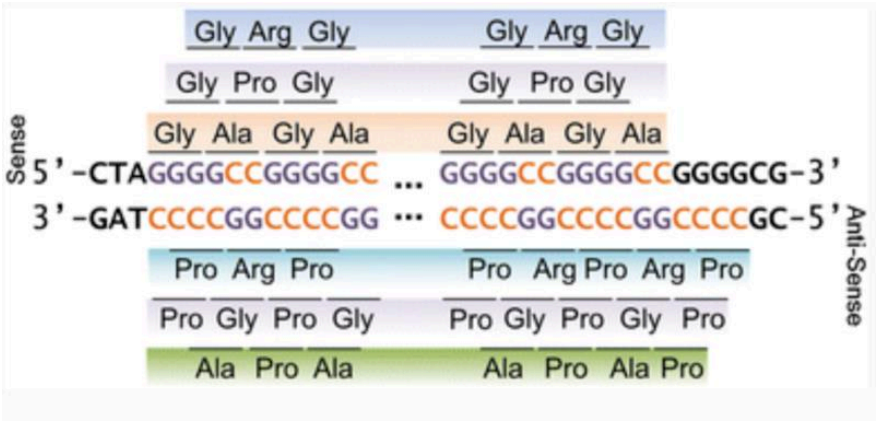


Figure 42. Traduction des ARN sens et antisens en protéines DPR
En fonction du cadre de lecture ouvert considéré, l'ARN sens composé de répétitions GGGGCC peut être traduit en protéines composées polyGA, polyGP et polyGR et ; l'ARN antisens peut être traduit en protéines composées de polyPR, polyGP et polyPA. Schéma de Gendron et al., 2013.

Une confirmation de la présence de ces protéines DPR fut le développement d'anticorps dirigés contre les parties répétées de ces protéines, qui ont alors permis d'observer la présence de ces différentes protéines DPR sous forme d'inclusions cytoplasmiques dans les neurones des patients C9-SLA/DFT (Figure 43), qui colocalisent avec des protéines impliquées dans l'autophagie, l'ubiquitine et p62 (Gendron et al., 2013 ; Mackenzie et al., 2013 ; Mori et al., 2013b ; Zu et al., 2013).

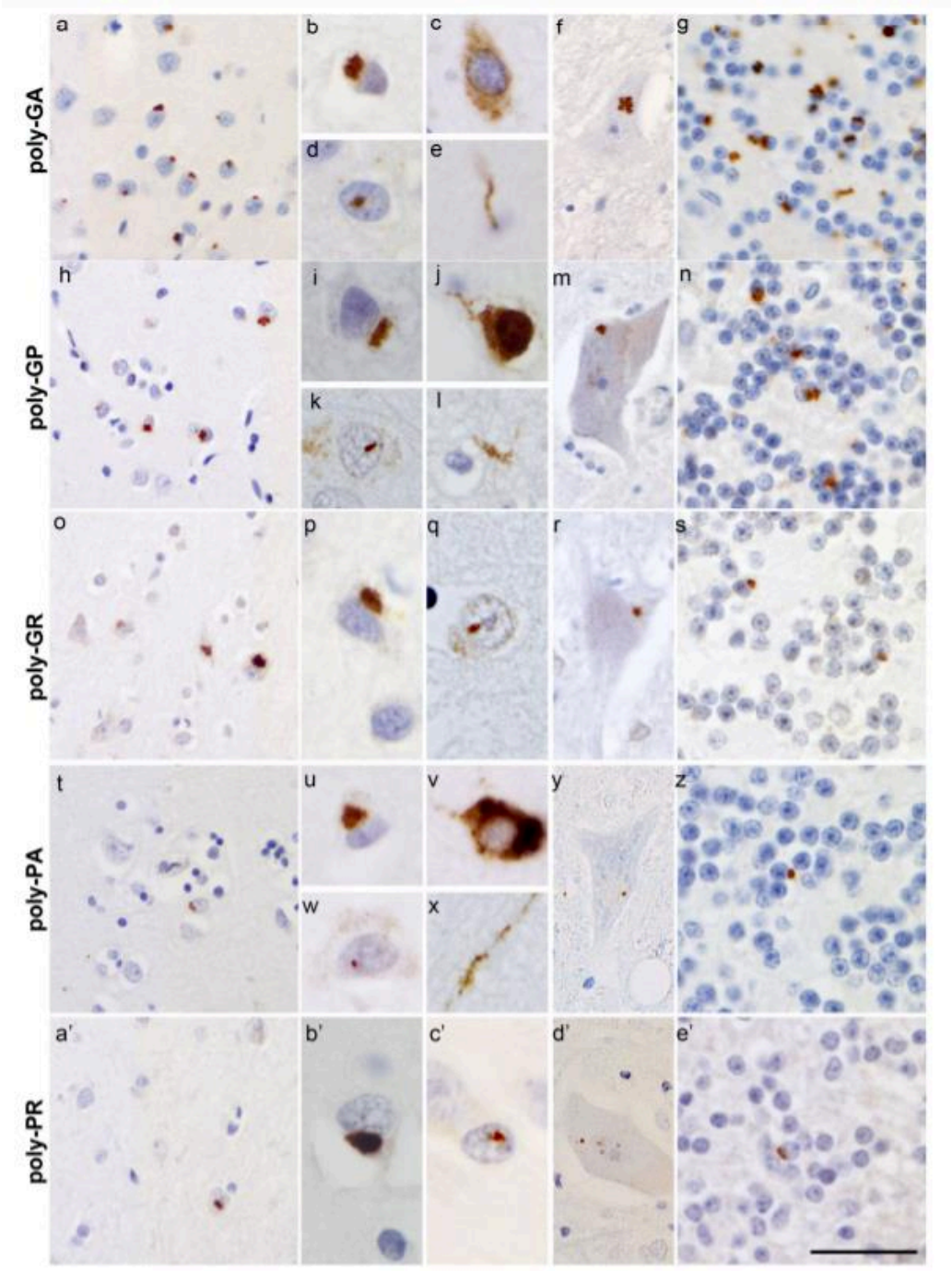


Figure 43. Présence des différentes protéines DPR chez un patient C9-SLA/DFT
IHC dirigées contre les protéines DPR (**a**-**g**) polyGA, (**h**-**n**) polyGP, (**o**-**s**) polyGR, (**t**-**z**) polyPA et (**a'**-**e'**) polyPR dans (**a**, **b**, **c**, **d**, **e**, **h**, **i**, **j**, **k**, **o**, **p**, **q**, **t**, **u**, **v**, **w**, **x**, **a'**, **b'**, **c'**) le cortex frontal, (**f**, **m**, **r**, **y**, **d'**) la moelle épinière où les inclusions sont rares et (**g**, **n**, **s**, **z**, **e'**) les cellules granulaires du cervelet où les inclusions sont nombreuses. Images de Mackenzie et al., 2015.

Toutefois, des analyses histopathologiques sur des lames de patients C9-SLA/DFT montrent que l'abondance des 5 protéines DPR dans le système nerveux central n'est pas homogène. En effet, la protéine la plus abondante est celle composée de polyGA, suivie de celle composée de polyGP, polyGR, polyPA et enfin polyPR (Figure 44 ; Mackenzie et al., 2015).

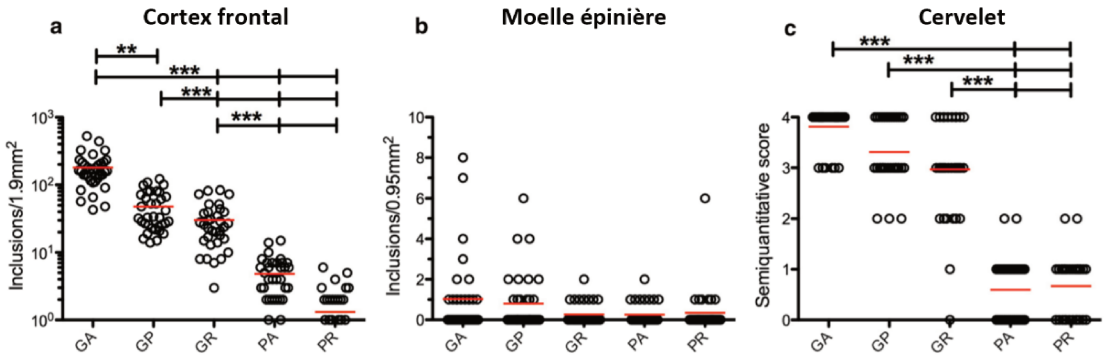


Figure 44. Quantification des agrégats de protéines DPR dans le cerveau de patients SLA
Nombre d'inclusions de protéines DPR polyGA, polyGP, polyGR, polyPA et polyPR dans (a) le cortex frontal, (b) la moelle épinière et (c) le cervelet de patients C9-SLA/DFT. Graphiques de Mackenzie et al., 2015.

De plus, de nombreuses études basées sur des techniques d'histopathologie ont tenté de corréler l'expression des protéines DPR avec les zones atteintes chez les patients, la sévérité, l'âge d'apparition ou la durée de la maladie. Plusieurs de ces travaux ne décèlent aucune corrélation entre les agrégats de protéines DPR et la neurodégénérescence. En effet, les inclusions formées par les protéines DPR sont principalement situées dans les cortex frontaux et temporaux, l'hippocampe, la couche granulaire du cervelet et beaucoup plus rarement dans la moelle épinière (Figure 44a, 44c) ; contrairement aux agrégats de TDP-43 où leur présence est détectée uniquement dans les zones atteintes (Barborie et al., 2015 ; Gomez-Deza et al., 2015 ; Mackenzie et al., 2015 ; Schludi et al., 2015 ; Davidson et al., 2016). Toutefois, il a été montré dans des études plus récentes une éventuelle corrélation entre les agrégats de polyGR et la zone de neurodégénérescence (Saberi et al., 2017 ; Sakae et al., 2018). Enfin, une dernière étude sur la part soluble des protéines DPR démontre une corrélation entre la protéine polyGR et l'âge d'apparition et la durée de la maladie ; ainsi qu'entre la protéine polyGP et l'âge d'apparition de la maladie (Quaegebeur et al., 2020). Néanmoins, l'absence des protéines DPR dans la moelle épinière soulève la question de leur contribution dans la mort des motoneurones (Figure 44b). Il a ainsi été proposé que les protéines DPR pourraient être extrêmement toxiques pour les motoneurones sous une forme soluble, résultant ainsi en une absence d'inclusion dans ces cellules.

Enfin, la surexpression de certaines de ces protéines DPR se révèle toxique dans de nombreux modèles comme les motoneurones différenciés à partir de cellules iPS, les drosophiles (les poissons zèbres et les souris en altérant de nombreuses voies cellulaires (Figure 45 ; Mizielinska et al., 2014 ; Wen et al., 2014 ; Tran et al., 2015 ; Yang et al., 2015 ; Lee et al., 2016 ; Lopez-Gonzalez et al., 2016 ; Westergard et al., 2016 ; Ohki et al., 2017 ; Schludi et al., 2017 ; Mordes et al., 2018 ; Swaminathan et al., 2018 ; Zhang et al., 2018 ; Hao et al., 2019 ; LaClair et al., 2020 ; Li et al., 2020 ; Maor-Nof et al., 2021). Ces protéines DPR, notamment celles riches en acides aminés arginine (polyGR et polyPR), semblent particulièrement toxiques sont décrites plus en détail ci-dessous.

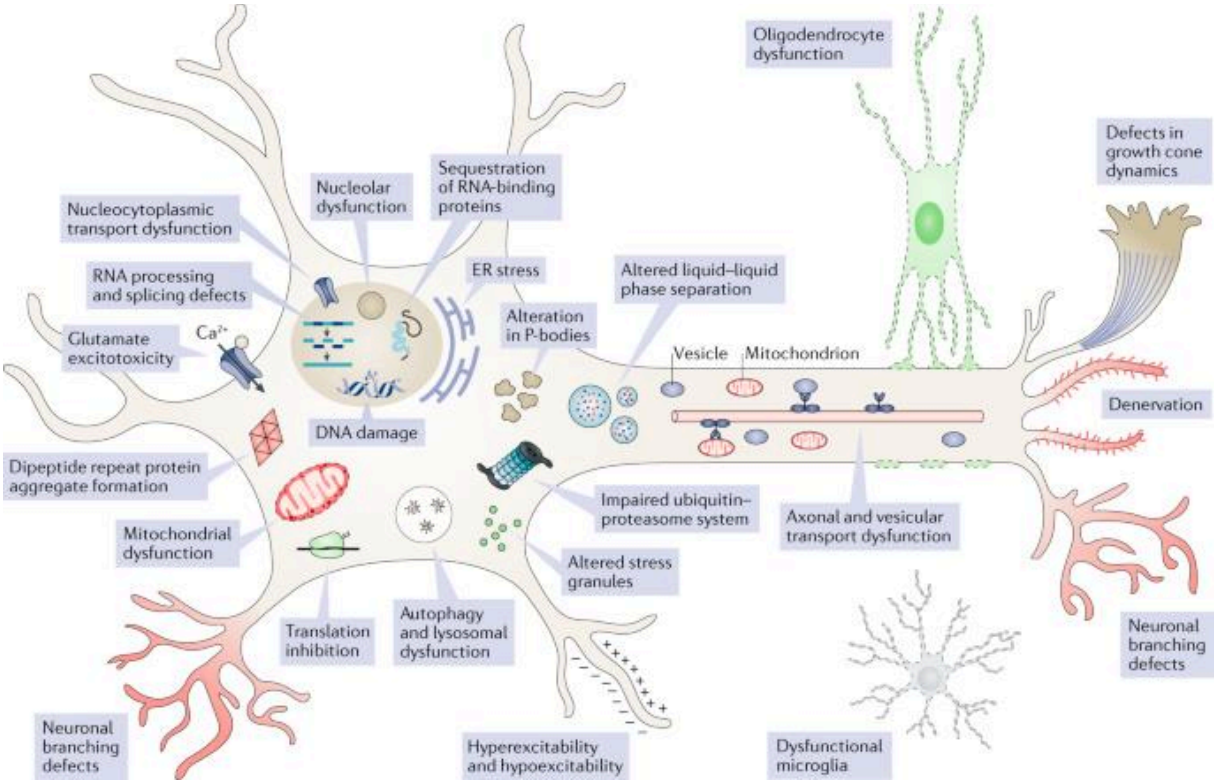


Figure 45. Voies cellulaires altérées par les protéines DPR.
L'expression des protéines DPR sont décrites comme altérant de nombreuses voies cellulaires et organelles comme les mitochondries, le réticulum, le protéasome, etc. Image de Balendra et Isaacs, 2018.

Il a été montré que la traduction de ces différentes protéines DPR faisait appel au mécanisme « RAN » (Ash et al., 2013 ; Gendron et al., 2013 ; Mori et al., 2013a ; Zu et al., 2013). La nature de ce mécanisme de traduction étant encore incertain, l'étude du mécanisme de traduction de ces différentes protéines fait partie de mon travail de thèse.

3.4.3.1. *Toxicité de la protéine DPR polyGlycine-Alanine (polyGA)*

La protéine polyGA est la protéine DPR retrouvée la plus abondamment chez les patients (Mackenzie et al., 2015). Cette protéine exprimée sous un ATG artificiel est moins toxique que les protéines DPR polyGR et polyPR. Elle est décrite pour former des fibres amyloïdes composées de feuillets β, éventuellement capables de se transmettre d'une cellule à une autre (Chang et al., 2016 ; Westergard et al., 2016 ; Brasseur et al., 2020 ; Khosvari et al., 2020). La protéine polyGA altérerait plusieurs mécanismes cellulaires. Tout d'abord, les agrégats de polyGA agissent sur l'import nucléaire de protéines en altérant la voie dépendante de l'importine-α/β, conduisant à la mauvaise localisation des protéines TDP-43 et PSCM4, une ATPase de la sous-unité du protéasome 26S (Khosvari et al., 2017 ; Khosvari et al., 2020). De plus, la séquestration de nombreuses sous-unités du protéasome a été rapportée dont Unc119, HR23 ou encore UBQLN2 (May et al., 2014 ; Schludi et al., 2015 ; Zhang et al., 2016 ; Schludi et al., 2017 ; Riemslagh et al., 2019 ; Zhang et al., 2021). Enfin, la résolution par Cryo-EM des agrégats cellulaires de polyGA confirme la présence de nombreuses sous-unités du protéasome 26S dans ces inclusions (Figure 46 ; Guo et al., 2018). Ces résultats semblent indiquer que les agrégats de polyGA altèrent l'activité du protéasome, ce qui pourrait conduire à un stress du réticulum endoplasmique et à la mort neuronale (Zhang et al., 2014). De plus, il a été montré que les agrégats de polyGA affectent également le second mécanisme de dégradation, l'autophagie, via la séquestration de la protéine VCP (Božič et al., 2021). Enfin, il a récemment été suggéré que dans des cultures primaires de neurones corticaux et moteurs, la protéine polyGA forme des agrégats dans les axones et les dendrites. La présence de ces agrégats modifie l'influx calcique et donc l'exocytose des vésicules synaptiques, notamment via la diminution de l'expression de la protéine SV2 (« Synaptic vesicle-associated protein 2 ») (Jensen et al., 2020).

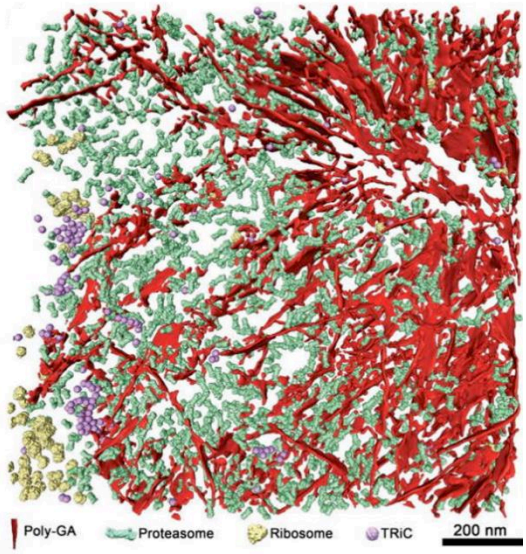


Figure 46. PolyGA séquestre le protéasome
Cryo-EM montrant la protéine polyGA séquestre des sous-unités du protéasome. Image de Guo et al., 2018.

Afin de confirmer la toxicité de cette protéine polyGA en modèle animal, plusieurs équipes ont développé des modèles murins exprimant spécifiquement cette protéine polyGA. En 2016, un premier modèle murin a été développé par l'injection d'un virus AAV1 codant pour une protéine GFP-$(GA)_{50x}$ traduite grâce à un codon ATG. L'expression de la protéine polyGA conduit à une atrophie du cerveau due à une mort neuronale dans le cortex, dont le cortex moteur, l'hippocampe mais aussi à une perte des cellules de Purkinje (Zhang et al., 2016). De plus, cette neurodégénérescence est associée à une astrogliose, sans activation de la microglie. Enfin, ces souris présentent des anomalies comportementales comme une hyperactivité, une anxiété et des altérations locomotrices (Zhang et al., 2016). L'année suivante, une lignée transgénique exprimant une protéine ATG-$(GA)_{150x}$-CFP sous le contrôle du promoteur *Thy1* spécifique des neurones a été développée. Ces souris présentent de nombreux agrégats de polyGA dans le tronc cérébral, le cervelet et la moelle épinière et ainsi qu'une hyper-phosphorylation de la protéine TDP-43, un marqueur histopathologique de la SLA/DFT. L'expression de la protéine polyGA ne conduit pas à une mort neuronale, mais provoque une activation de la microglie sans astrogliose et, ces souris développent des anomalies locomotrices révélées par le test de la marche sur poutre (Schludi et al., 2017). La génération d'un autre modèle murin exprimant une protéine ATG-GFP-$(GA)_{175x}$ sous le contrôle d'un promoteur *Nestin*, permettant, là encore, une expression neuronale de la protéine, révèle que la protéine polyGA induit une mort des neurones de l'hippocampe associée à une infiltration de la microglie ; et des motoneurones de la moelle épinière, conduisant à une dénervation des muscles ainsi qu'à une mauvaise localisation de la protéine TDP-43 (LaClair et al., 2020).

Il est à noter que toutes ces études ont été réalisées avec des répétitions GGGGCC traduites en protéine polyGA grâce à un codon ATG artificiel situé en amont de ces répétitions. En effet, la seule expression des répétitions GGGGCC ne conduit pas à l'expression d'une protéine polyGA détectable chez le poisson zèbre, ce qui n'est alors pas toxique. Néanmoins, l'expression de la protéine polyGA grâce à un codon ATG conduit à de la neurodégénérescence chez ces poissons zèbres (Ohki et al., 2017). Il reste donc à déterminer si la protéine polyGA exprimée sous sa séquence naturelle conduit à des altérations, identiques ou différentes, que celles observées lorsque cette protéine est exprimée sous un ATG artificiel sans séquence bordante, ainsi qu'à déterminer comment cette protéine peut-elle être traduite malgré l'absence de codon d'initiation canonique ATG. Ces questions ont fait l'objet de mon travail de thèse.

3.4.3.2. *Les protéines DPR non-toxiques : polyGlycine-Proline et polyProline-Alanine*

La protéine polyGP est la seule protéine DPR pouvant être traduite à partir de l'ARN sens GGGGCC et de l'ARN antisens CCCCGG. Les protéines DPR polyGP et polyPA exprimées à partir d'un codon ATG artificiel ne sont pas toxiques en modèle de drosophile (Mizielinska et al., 2014, Wen et al., 2014 ; Freibaum et al., 2015 ; Lee et al., 2016). De plus, des expériences de transfection en culture de cellules immortalisées ont montré que la protéine polyGP s'agrège moins que les autres protéines DPR (Yamakawa et al., 2015).

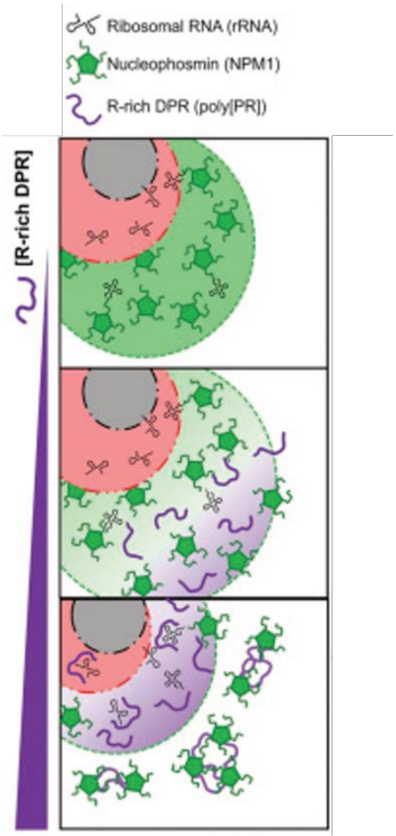
La protéine polyGP existant sous forme soluble, elle peut être dosée dans le liquide cérébro-spinal ainsi que dans les cellules mononucléées périphériques du sang des patients C9-SLA/DFT. Le niveau d'expression de la protéine polyGP permet de discriminer les patients atteints de SLA/DFT des individus sains, ainsi que de détecter des patients présymptomatiques. Toutefois, la quantité de polyGP n'évolue pas avec l'âge des patients et ne permet donc pas de déterminer l'avancement de la pathologie. Ainsi, le dosage de la protéine polyGP pourrait servir de biomarqueur précoce afin d'identifier de potentiels patients (Gendron et al., 2017 ; Lehmer et al., 2017 ; Meeter et al., 2018).

3.4.3.3. *Les protéines DPR toxiques : polyGlycine-Arginine et polyProline-Arginine*

Les protéines DPR riches en arginine, la protéine polyGR traduite à partir de l'ARN sens et la protéine polyPR traduite à partir de l'ARN antisens, partagent de nombreuses similarités. Tout d'abord, ces 2 protéines DPR sont chargées positivement et sont hautement polaires, dues à leur richesse en arginine, et sont moins sujettes à l'agrégation que la protéine polyGA (Kwon et al., 2014). De plus, l'interactome de ces protéines révèle une grande similarité, avec plus de 40% d'interactants communs (Lee et al., 2016 ; Božič et al., 2021). Enfin, de toutes les protéines DPR, ce sont les plus toxiques lorsqu'elles sont exprimées sous un ATG artificiel en modèles cellulaires, de drosophile, de poisson zèbre ou murins (Mizielinska et al., 2014 ; Wen et al., 2014 ; Freibaum et al., 2015 ; Tao et al., 2015 ; Boeynaems et al., 2016 ; Lee et al., 2016 ; Zhang et al., 2018 ; Corman et al., 2019 ; Hao et al., 2019 ; Cook et al., 2020 ; Li et al., 2020 ; Maor-Nof et al., 2021 ; Riemslagh et al., 2021). En effet, les différents modèles de souris exprimant les protéines DPR riches en arginine présentent un phénotype sévère dont une forte neurodégénérescence conduisant à une atrophie du cerveau, de l'astrogliose, une activation de

la microglie, une perte de poids ainsi que des altérations motrices (Zhang et al., 2018 ; Hao et al., 2019 ; Cook et al., 2020 ; Maor-Nof et al., 2021).

Bien que ces protéines DPR riches en arginine ont été décrites comme compromettant plusieurs mécanismes cellulaires comme le trafic intracellulaire par l'altération des réseaux de microtubules (Fumagalli et al., 2021) ou encore l'intégrité mitochondriale (Lopez-Gonzalez et al., 2016 ; Choi et al., 2019), la découverte que ces protéines DPR lient et séquestrent la protéine Nucléophosmine 1 (NPM1) apporte une explication à de nombreuses altérations qui avaient été précédemment observées (Figure 47 ; Farg et al., 2017). En effet, la protéine NPM1 est impliquée dans de nombreux processus cellulaires comme la biogénèse et le transport du ribosome, la réparation de l'ADN et la séparation de phase liquide-liquide permettant de la formation nombreuses organelles sans membrane comme les granules de stress et les nucléoles ; tous affectés par l'expression des protéines DPR polyGR et polyPR (Kwon et al., 2014 ; Freibaum et al., 2015 ; Tao et al., 2015 ; Lee et al., 2016 ; Lin et al., 2016 ; Boeynaems et al., 2017 ; Farg et al., 2017 ; Suzuki et al., 2018b ; Zhang et al., 2018 ; White et al., 2019 ; Andrade et al., 2020; Hayes et al., 2020).



Les protéines polyGR et polyPR sont très similaires néanmoins, ces protéines ne sont pas localisées au même endroit lors d'expériences de transfection dans des cellules. La protéine polyGR est dans le cytoplasme alors que la protéine polyPR se trouve dans les nucléoles (Lee et al., 2016). Il est à noter que ces 2 protéines sont majoritairement cytoplasmiques chez les patients (Mori et al., 2013a ; Zu et al., 2013).

Enfin, il est à noter que ces études portant sur la toxicité des protéines polyGR et polyPR ont utilisé des constructions contenant un ATG artificiel pour initier leur traduction. Il reste donc à déterminer la toxicité de ces protéines entourées de leur séquence naturelle du gène *C9ORF72*.

Figure 47. La protéine NPM1 est délocalisée par les protéines polyGR et polyPR
Les protéines polyGR et polyPR séquestrent et altèrent la localisation de la protéine NPM1, conduisant à l'altération de nombreux processus cellulaires. White et al., 2019.

3.4.4. Exemple de modèles exprimant les répétitions d'hexa-nucléotides GGGGCC

De nombreux modèles, cellulaires et animaux, ont été développés dans le but d'attester de la toxicité des expansions de répétitions GGGGCC dans le gène *C9ORF72*. Ces modèles sont plus proches de la situation des patients toutefois, ils ne permettent pas ou difficilement de discriminer le mécanisme (ADN, ARN ou protéique) responsable des altérations observées.

Ainsi, des lignées de cellules iPS générées à partir de fibroblastes de patients C9-SLA/DFT puis, différenciées en neurones ou en motoneurones (iMN) présentent quasiment systématiquement des foci d'ARN nucléaires. Toutefois, selon les études et les lignées iPS utilisées, une forte variabilité du niveau d'expression de l'ARNm C9ORF72 est observée, suggérant que les modifications épigénétiques de l'ADN et l'haploinsuffisance ne sont pas systématiquement reproduites (Almeida et al., 2013 ; Donnelly et al., 2013 ; Sareen et al., 2014 ; Su et al., 2014 ; Freibaum et al., 2015 ; Dafinca et al., 2016 ; Shi et al., 2018 ; Simone et al., 2018). De plus, les protéines DPR semblent difficilement détectables dans ces modèles cellulaires et leur présence, principalement la protéine polyGP, est révélée majoritairement par test Elisa et, éventuellement par dot blot (Almeida et al., 2013 ; Donnelly et al., 2013 ; Su et al., 2014 ; Dafinca et al., 2016 ; Simone et al., 2018). Les différences entre ces différentes lignées iPS compliquent les analyses permettant d'incriminer les mécanismes perte de fonction et / ou gain de fonction. Ainsi, selon les études, les expansions GGGGCC dans le gène *C9ORF72* ont été montré comme pouvant induire des altérations des mitochondries et du réticulum endoplasmique conduisant à des concentrations calciques anormales (Dafinca et al., 2016 ; Dafinca et al., 2020), de l'autophagie (Almeida et al., 2013 ; Shi et al., 2018), du transport nucléocytoplasmique (Zhang et al., 2015), de la formation des granules de stress (Dafinca et al., 2016) ou encore présentent une sensibilité accrue au glutamate (Donnelly et al., 2013) pouvant conduire à de la mort cellulaire (Shi et al., 2018).

Concernant les modèles murins, deux approches distinctes ont été utilisées : l'approche virale et l'insertion d'un chromosome artificiel bactérien (BAC) codant pour les répétitions GGGGCC. A ce jour, 2 groupes ont réalisé des expériences basées sur l'injection de virus chez des souris naissantes. Toutes 2 ont utilisé des AAV2/9 porteurs de 66 répétitions (Chew et al., 2015) ou 102 répétitions (Herranz-Martin et al., 2017) d'hexa-nucléotides GGGGCC. Ces expansions de répétitions sont transcrites en ARN formant des foci et sont aussi traduits en protéines DPR polyGA, polyGP et polyGR (Chew et al., 2015 ; Herranz-Martin et al., 2017). Seul le modèle de Chew et al. présente des agrégats de TDP-43, une astrogliose ainsi que des

altérations comportementales comme de l'hyperactivité, de l'anxiété et des déficits locomoteurs. Concernant les modèles transgéniques, 4 groupes ont généré des modèles murins par l'insertion du BAC contenant des répétitions GGGGCC entourés d'une séquence plus ou moins longue du gène *C9ORF72* humain (O'Rourke et al., 2015 ; Peters et al., 2015 ; Jiang et al., 2016 ; Liu et al., 2016b). Tous ces modèles développent des foci d'ARN. Les 3 groupes ayant utilisé le fond génétique C57BL6 détectent la protéine polyGP ; seul un groupe détecte également les 2 autres protéines DPR provenant de l'ARN sens polyGA et polyGR (Jiang et al., 2016), mais ces souris ne développent ni altération locomotrice ni neurodégénérescence, suggérant que les foci d'ARN et les protéines DPR ne sont pas suffisants pour induire la SLA/DFT (O'Rourke et al., 2015 ; Peters et al., 2015 ; Jiang et al., 2016). L'équipe de Liu et al. a, elle, utilisé des souris de fond génétique FVB et, bien qu'aucune protéine DPR n'ait été détectée lors de cette première étude, des travaux supplémentaires sur ce modèle ont révélé la présence des protéines DPR polyGA, polyGP et polyGR (Nguyen et al., 2020). Ces souris affichent une longévité plus faible, une paralysie associée à une dénervation des muscles et une perte des motoneurones (Liu et al., 2016b). Ces résultats ont été récemment contesté par Mordes et al. qui n'a pas observé de telles altérations phénotypiques chez ces souris soulignant que les souris FVB présentent naturellement de l'épilepsie, une plus faible longévité ainsi que des dysfonctions locomotrices (Mordes et al., 2020).

En conclusion, ces modèles iPS et murins soulignent la complexité des mécanismes sous-jacents à l'expansion de répétitions GGGGCC dans le gène *C9ORF72* impliqués dans cette pathologie.

3.5. Hypothèses et objectifs

Une expansion de répétitions d'hexa-nucléotides GGGGCC située dans le premier intron du gène *C9ORF72* est la première cause génétique de SLA, de DFT et de SLA/DFT. Différents mécanismes non-exclusifs sont proposés pour expliquer la pathogénicité de cette expansion, à savoir (1) la perte de fonction de la protéine C9ORF72 impliquée dans l'autophagie, (2) les foci d'ARN sens et antisens séquestrant des protéines spécifiques et (3) la traduction des ARN contenant ces répétitions par un mécanisme « RAN » en protéines DPR toxiques.

Avant mon arrivée au laboratoire, au vu de la taille de l'expansion de répétitions GGGGCC (plusieurs centaines à plusieurs milliers), mon équipe a tout d'abord testé l'hypothèse d'un gain de fonction toxique de l'ARN. Pour ce faire, mon équipe a développé un ARN composé de répétitions GGGGCC synthétique et qui a été lié à une colonne. Puis, le passage de protéines extraites à partir de cerveaux de souris sur cette colonne a été réalisé, suivi de l'élution des protéines accrochées à l'ARN. L'analyse par spectrométrie de masse de ces protéines a permis d'identifier notamment la protéine hnRNP-H. Cette protéine colocalise avec les foci d'ARN composés de répétitions GGGGCC mais peu d'altérations d'épissage des ARN cibles de la protéine hnRNP-H ont été détectées (Almeida et al., 2013). C'est pourquoi par la suite, mon équipe s'est intéressée à la fonction de la protéine C9ORF72 et a mis en évidence que cette protéine interagit avec les protéines SMCR8 et WDR41, ce complexe jouant un rôle dans l'autophagie (Seillier et al., 2016). Mon travail de thèse a alors été de tester si la déplétion de la protéine C9ORF72, qui conduit à une autophagie sous-optimale, ne pouvait pas conduire à l'accumulation des protéines DPR. En effet, ces protéines forment des agrégats cytoplasmiques trop massifs pour être dégradées par le protéasome et sont donc éliminées par autophagie.

En parallèle, nous nous sommes également intéressés au mécanisme de traduction de ces protéines DPR. En effet, la majeure partie des études portant sur la toxicité des protéines DPR utilise un codon artificiel ATG situé en amont des répétitions permettant de traduire les répétitions GGGGCC dans un cadre de lecture choisi et à un niveau élevé. Or, les 5 protéines DPR ne sont pas en observées en abondance égale chez les patients SLA/DFT, suggérant des mécanismes de traduction différents de ces protéines. Nous avons voulu comprendre pourquoi il existe cette variabilité d'expression entre les protéines DPR et, pour cela, nous nous sommes donc intéressés à leur mécanisme de traduction. Ces objectifs ont fait l'objet de mon travail de thèse et les résultats obtenus sont présentés sous forme de publication (Publication 2).

INTRODUCTION

- CHAPITRE 4 -

Stratégies thérapeutiques

4. STRATEGIES THERAPEUTIQUES

A l'heure actuelle, la SLA reste une maladie incurable. En effet, seul le riluzole, un inhibiteur de la libération du glutamate semble ralentir la mort des motoneurones, toutefois avec une efficacité clinique limitée ne permettant aux patients atteints de SLA de gagner, en moyenne, que 2 mois de vie. De même, les pathologies dues à des expansions de répétitions qui sont traduites (protéines à polyA, polyG et polyQ) sont à ce jour, comme la SLA, incurables. La recherche de molécules et le développement de nouvelles stratégies thérapeutiques pour ces maladies sont donc des domaines de recherche en pleine expansion.

4.1. Molécules liant l'ADN et / ou l'ARN

Le point commun à toutes ces pathologies précédemment présentées est la présence d'expansion de répétitions de nucléotides CAG, CGN, CGG, GGGGCC, etc. qui présentent des similarités en termes de séquences mais pas forcément au niveau des structures formées par l'ADN et l'ARN. En effet, les ARN composés de CAG forment des hélices A-A (Kiliszek et al., 2010), alors que les ARN composés de CGG et de GGGGCC forment des quadruplexes-G (Darnell et al., 2001 ; Darnell et al., 2005 ; Didiot et al., 2008 ; Reddy et al., 2013 ; Conlon et al., 2016). Trouver des molécules liant spécifiquement ces structures et conduisant à la contraction des expansions de répétitions au niveau ADN, ou à la déstabilisation des ARN empêchant leur agrégation sous forme de foci d'ARN ou leur traduction s'avèrerait donc une approche de choix afin de soigner ces différentes pathologies, surtout qu'une même molécule pourrait éventuellement agir sur des répétitions similaires et donc cibler plusieurs maladies.

Concernant les maladies à polyQ causées par des expansions de tri-nucléotides CAG, une molécule appelée naphthyridine-azaquinolone (NA) a été montrée comme se liant spécifiquement au niveau ADN aux expansions de répétitions CAG de taille intermédiaire ou longue, ne se fixant donc pas sur l'allèle non-muté. En modèle murin de la maladie de Huntington, l'injection intracérébrale de NA conduit à une diminution de la taille de l'expansion CAG, conduisant à une moindre agrégation de la protéine HTT-polyQ et à une meilleure survie des souris. La contraction de l'expansion de répétitions CAG par le NA ferait intervenir le système de reconnaissance et de réparation des mésappariements de l'ADN (Nakamori et al., 2020).

Dans le cadre de l'étude des maladies FXS, FXTAS, C9-SLA/DFT, etc., des molécules ciblant spécifiquement les ARN riches en GC et pouvant moduler leur traduction sont activement recherchées. Par exemple, il a été montré que le composé 1a (9-hydroxy-5,11-dimethyl-2-(2-(piperidin-1-yl)ethyl)-6H-pyrido[4,3-b]carbazol-2-ium) diminuait le nombre de foci d'ARN composé de répétitions CGG ainsi que l'agrégation de la protéine FMRpolyG en culture primaire de neurones (Disney et al., 2012 ; Haify et al., 2021). Ce même composé semble également fonctionné dans la C9-SLA/DFT puisqu'il inhibe la formation de foci d'ARN et la génération de protéines DPR dans les iNeurones différenciés à partir de fibroblastes de patients (Su et al., 2014).

De même, la porphyrine TMPyP4 lie et déstabilise les quadruplexes-G formés par les ARN composés de répétitions CGG, conduisant à une augmentation de la traduction de la protéine FMRP, l'absence de cette protéine conduisant à FXS (Ofer et al., 2009). Il a été montré par la suite que cette même molécule TMPyP4 est capable de lier également les répétitions GGGGCC de l'ARN de C9ORF72 impliquées dans la SLA/DFT et de déstructurer ces ARN, conduisant à une moindre séquestration des protéines liant cet ARN ainsi qu'à une plus faible traduction en protéines DPR (Zamiri et al., 2014 ; Mori et al., 2021).

Au contraire de ces molécules qui déstabilisent les quadruplexes-G des ARN comportant des répétitions GGGGCC, un composé, le DB1273, stabilise ces structures secondaires diminuant ainsi l'export de ces ARN et leur traduction en protéines DPR dans des modèles de neurones différenciés à partir de cellules iPS de patients C9-SLA/DFT. Ce composé permet aussi une meilleure survie des drosophiles exprimant 36 motifs GGGGCC (Simone et al., 2018).

Enfin, une stratégie récente basée sur l'attraction d'une ribonucléase, capable de dégrader les ARN, par une molécule reconnaissant et liant les expansions ARN GGGGCC conduit à une diminution d'expression des protéines DPR en motoneurones différenciés à partir de cellules iPS de patients C9-SLA/DFT ainsi que dans un modèle murin porteurs des expansions de répétitions GGGGCC (Bush et al., 2021).

4.2. Molécules activant l'autophagie

La plupart des maladies neurodégénératives comme les maladies d'Alzheimer, de Parkinson, à polyQ, à polyG et la SLA/DFT, etc. sont caractérisées par l'expression de protéines toxiques formant des agrégats protéiques, qui peuvent potentiellement être ciblés par le mécanisme d'autophagie. Des molécules activant l'autophagie sont donc une piste thérapeutique intéressante, puisqu'une molécule pourrait permettre le traitement de plusieurs pathologies à agrégats protéiques (Figure 48). Parmi ces molécules, trois classes se distinguent : celles activant l'autophagie dépendamment de mTOR comme la rapamycine et ses dérivés ; celles agissant sur l'AMPK comme la metformine ; et celles activant l'autophagie indépendamment de mTOR comme le tréhalose, le vérapamile, le lopéramide, la clonidine ou encore la calpastatine (pour revue, Park et al., 2020).

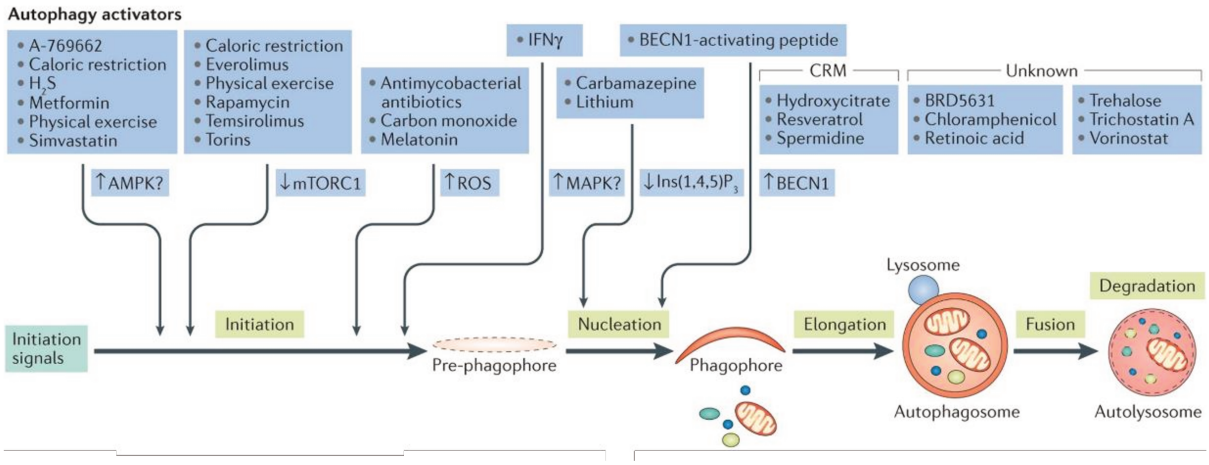


Figure 48. Molécules activant l'autophagie
L'autophagie peut être stimulée à des étapes différentes, notamment durant l'initiation et la nucléation, par différentes molécules qui agissent sur des protéines de la cascade autophagique. Galluzzi et al., 2017.

Le traitement de modèles murins de SLA par des inhibiteurs de mTOR, notamment la rapamycine, donne un résultat mitigé. En effet, l'administration de rapamycine chez des souris porteuses d'une mutation dans le gène *TARDP* améliore leur état général avec une diminution de la mort neuronale (Wang et al., 2012). De plus, la rapamycine ainsi que le torkinib, un second inhibiteur de mTOR, améliore le phénotype des cultures primaires de neurones et des drosophiles exprimant une version mutée de la protéine FUS impliquée dans la SLA, suggérant que l'activation de l'autophagie pourrait tout de même constituer une piste thérapeutique possible (Marrone et al., 2018 ; Marrone et al., 2019). Néanmoins, les souris ayant une mutation dans le gène *SOD1* et traitées avec la rapamycine meurent plus rapidement (Zhang et al., 2011).

Il serait intéressant de tester l'effet de la rapamycine chez un modèle murin C9-SLA/DFT même si les inhibiteurs de mTOR sont rarement utilisés en clinique car mTOR est une protéine kinase centrale du métabolisme cellulaire ne jouant pas uniquement un rôle dans l'autophagie, mais aussi dans la croissance, la prolifération, la mobilité et la survie cellulaires ainsi que dans la transcription et la traduction de nombreux gènes (pour revue, Saxton et Sabatini, 2017). De ce fait, ces composés provoquent de nombreux effets secondaires indésirables chez l'Homme comme des pneumopathies, de la toxicité pour les reins, une hypercholestérolémie, etc. (Buhaescu et al., 2006). Malgré ces effets secondaires, la rapamycine reste utilisée comme immunosuppresseur lors d'une transplantation rénale (Mahalati et Kahan, 2001) et l'évérolimus (RAD001) contre le cancer et la sclérose tubéreuse (Hasskarl, 2018).

Concernant l'AMPK, cette kinase répond au niveau énergétique (AMP) de la cellule et permet d'induire l'autophagie lorsque le niveau en énergie est faible. La metformine, un stimulateur de l'AMPK, est utilisé comme antidiabétique et est décrite comme améliorant le phénotype des souris FXS et HD respectivement causées par des expansions de répétitions CGG et CAG (Ma et al., 2006 ; Gantois et al., 2017 ; Arnoux et al., 2018). L'ajout de metformine dans l'eau des souris exprimant des expansions de répétitions GGGGCC du gène *C9ORF72* permet une diminution de l'expression des protéine DPR et une amélioration des capacités locomotrices de ces souris (Zu et al., 2020). Ces résultats ont conduit à tester la metformine chez les patients C9-SLA/DFT et, cette étude est actuellement en phase clinique II (NCT04220021). Un second stimulateur de l'AMPK, le bosutinib, approuvé par l'administration de la nourriture et des médicaments (FDA) et l'agence européenne de médecine pour le traitement de leucémie myéloïde chronique, augmente la survie des motoneurones différenciés à partir de cellules iPS de patients SLA sporadiques ou porteurs de mutations dans les gènes *SOD1*, *TARDP* et *C9ORF72*, suggérant que cette molécule pourrait être utilisée chez une majorité des patients SLA (Imamura et al., 2017 ; Osaki et al., 2018). Ce composé est actuellement à l'étude en phase clinique I au Japon.

Enfin, de nombreux composés, activant l'autophagie indépendamment de mTOR ou d'AMPK ont également été testés. Dans la SLA/DFT, le lithium améliore le phénotype des souris porteurs d'une mutation dans le gène *SOD1* (Fornai et al. 2008), ainsi que la durée de vie des patients SLA/DFT porteurs d'une mutation dans le gène *UNC13A* mais pas des patients porteurs de l'expansion de répétitions GGGGCC dans le gène *C9ORF72* (Van Eijk et al., 2017).

De plus, une étude réalisée en 2010 sur des cultures primaires de neurones d'hippocampe, de striatum et de cortex a permis de mettre en évidence que des composés de la classe des tricycliques permettaient d'activer l'autophagie dans ces cellules ; améliorant ainsi leur survie quand elles surexprimaient la protéine à polyQ impliquée dans HD (Figure 49 ; Tsvetkov et al., 2010). Les tricycliques et ses dérivés sont approuvés par la FDA, notamment comme antidépresseurs. Ces composés, comme la trifluopérazine, la chlorpromazine, la mésoridazine, la thioridazine, le pimozide (5 antipsychotiques utilisés pour le traitement de schizophrénie), la promazine, la triflupromazine, (2 antipsychotiques avec de forts effets sédatifs, et donc peu utilisés en clinique), la prométhazine (un anti-histaminique utilisé dans le traitement de l'allergie et de l'insomnie) et dans une moindre mesure le 10-(4′-(N-diethylamino)butyl)-2-chlorophenoxazine (NCP) pourraient être des candidats potentiels pour activer l'autophagie pour le traitement de nombreuses pathologies caractérisées par la présence d'agrégats protéiques (Figure 49).

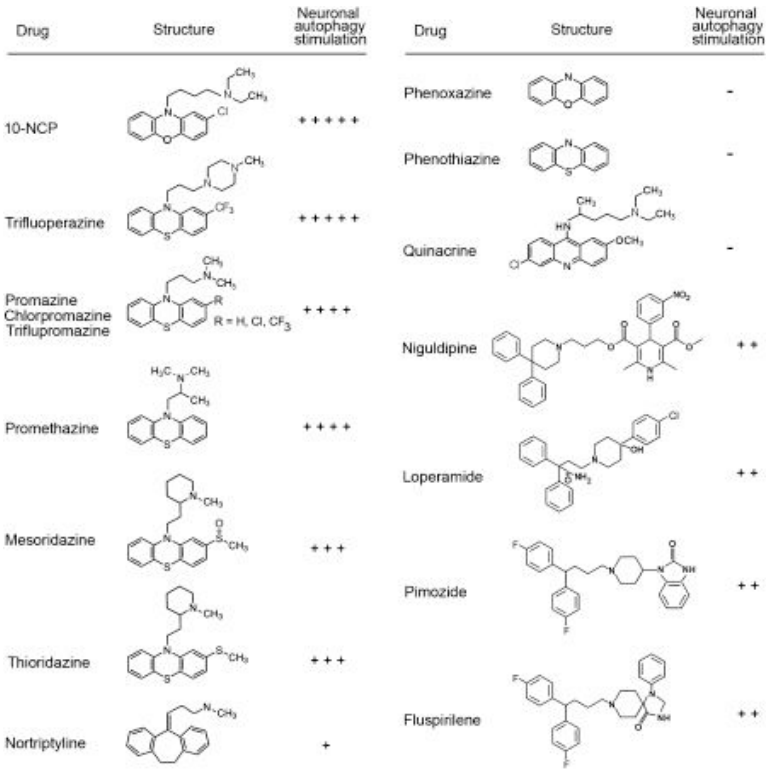


Figure 49. Composés de la classe des tricycles activant ou non l'autophagie
Les composés de la classe des tricycles comme le NCP, la trifluopérazine, la promazine, la chlorpromazine, la triflupromazine et la prométhazine activent fortement l'autophagie en cultures primaires de neurones. Tsvetkov et al., 2010.

Par exemple, il a été montré que le pimozide permet un meilleur maintien des jonctions neuromusculaires chez le poisson zèbre et la souris exprimant des formes mutées des protéines TDP-43 et SOD1 retrouvées dans la SLA (Patten et al., 2017). De plus, le traitement de cultures primaires de neurones avec de la prométhazine, mais également avec de la paroxétine, un

antidépresseur, diminue le nombre de granules de stress induites par des mutations dans la protéine FUS, impliquée dans la SLA (Marrone et al., 2018). Enfin, des cultures primaires de neurones et d'astrocytes exprimant, là encore, une protéine TDP-43 mutée traitée avec de la fluphénazine, utilisée comme antipsychotique aux propriétés anti-dopaminergiques, améliore la survie de ces cellules (Baramada et al., 2014).

Il serait donc intéressant de tester ces molécules aux effets bénéfiques sur des modèles de HD et de SLA et / ou DFT causées par des mutations dans les gènes *TARDBP*, *SOD1*, *FUS*, etc., dans des modèles de SLA/DFT avec une expansion de répétitions GGGGCC dans le gène *C9ORF72* mais aussi dans des modèles de FXTAS exprimant la protéine FMRpolyG.

4.3. Anticorps dirigés contre les protéines toxiques

La maladie d'Alzheimer (AD) est une démence caractérisée notamment par une perte de la mémoire causée par la dégénérescence des neurones de l'hippocampe qui s'étend par la suite au reste du cerveau. L'analyse histopathologique des cerveaux des patients AD révèle l'accumulation du peptide β-amyloïde sous forme de plaques séniles ainsi que l'agrégation et la phosphorylation anormale de la protéine Tau, la mutation de cette protéine Tau pouvant conduire à une DFT. Cette année, la FDA a autorisé l'utilisation de l'aducanumab, un anticorps dirigé contre le peptide β-amyloïde chez les patients AD atteints de forme précoce (Sevigny et al., 2016 ; Mullard, 2021). Toutefois, bien que cet anticorps aducanumab permet une réduction de la taille des plaques β-amyloïdes, le bénéfice de la destruction de ces plaques est encore discuté au sein de la communauté scientifique (Holmes et al., 2008). C'est pourquoi des essais cliniques utilisant des anticorps dirigés, cette fois, contre la protéine Tau mutée sont en cours (pour revue, Congdon et Sigurdsson, 2018).

Les résultats plutôt encourageants obtenus sur les patients atteints par AD ainsi que l'observation de la transmission entre cellules des agrégats de protéines DPR, notamment de la protéine polyGA, codées à partir des expansions de répétitions GGGGCC dans le gène *C9ORF72* ont conduit les chercheurs à s'intéresser à cette approche thérapeutique dans le cadre de la SLA/DFT (Chang et al., 2016 ; Westergard et al., 2016 ; Brasseur et al., 2020 ; Khosvari et al., 2020). De manière intéressante, un anticorps dirigé contre la protéine polyGA inhibe son agrégation et sa transmission d'une cellule à l'autre en cultures primaires de neurones (Zhou et al., 2017). Ces résultats ont été confirmé dans un souris modèle murin exprimant les répétitions GGGGCC où l'injection intra-péritonéale de l'anticorps α-GA_1 glycosylé dirigé contre la protéine polyGA conduit à une diminution de l'expression des protéines DPR polyGA, mais aussi polyGP et polyGR, à une diminution de la neuro-inflammation et de la neurodégénérescence ainsi qu'à une augmentation de la survie de ces souris (Nguyen et al., 2021).

4.4. Oligonucléotides antisens (ASO)

Les oligonucléotides antisens (ASO) sont des oligonucléotides mesurant de 12 à 30 nucléotides liant spécifiquement l'ARN et qui sont modifiés chimiquement. Les modifications les plus communes étant l'ajout d'un groupement 2'-O-méthyl sur le ribose et, d'un soufre sur le phosphate (phosphorothioate), qui augmentent sa stabilité et son affinité pour l'ARN cible ainsi que sa pénétrance dans les tissus et cellules. Il existe un mécanisme dit bloquant où l'ASO se fixe à l'ARN cible et agit par encombrement allostérique. Il peut alors modifier l'épissage alternatif ou inhiber la traduction en bloquant le passage du ribosomesur l'ARN cible (Figure 50). Il existe également un mécanisme dit dégradant où des ASO, appelés gapmer, sont composés d'ARN modifiés chimiquement entourant une séquence d'ADN et, l'appariement entre l'ADN du gapmer et l'ARN cible induit le recrutement de la RNase H1 et donc la dégradation de l'ARN (Figure 50 ; pour revue, Bennet et al., 2019).

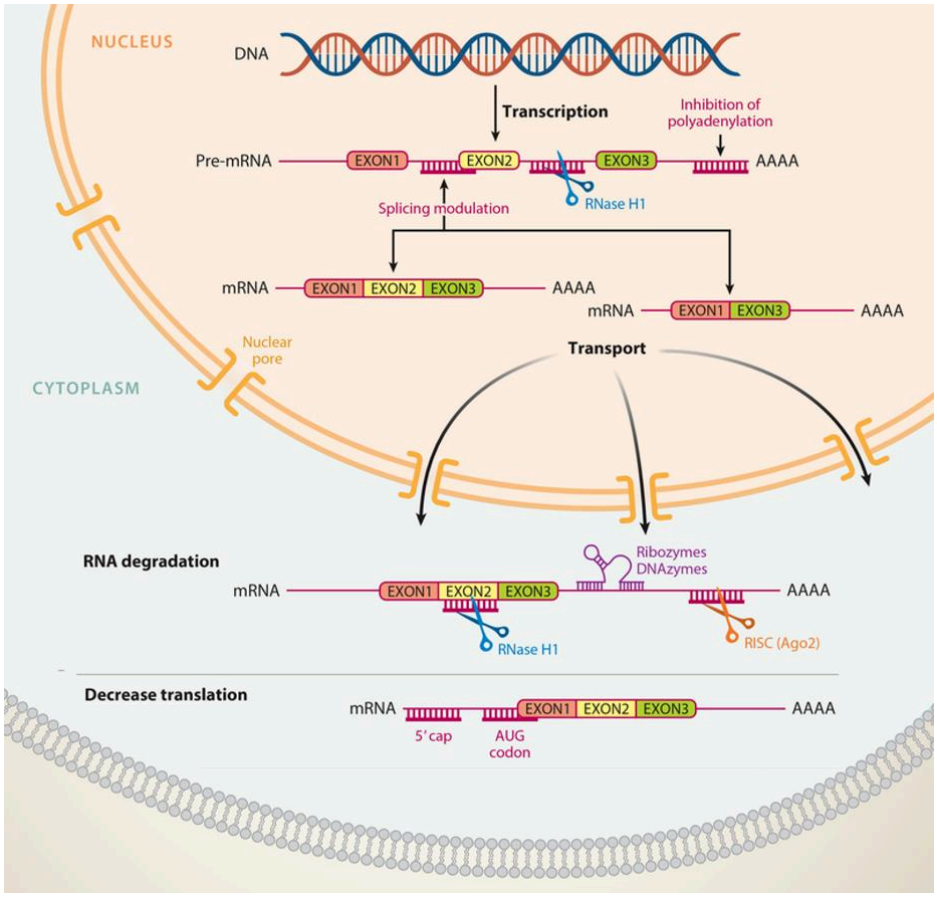


Figure 50. Mécanismes d'action des ASO
Les ASO de type gapmer se fixent sur l'ARN et le duplexe ADN/ARN est reconnu par la RNase H1 qui clive cet ARN. Les ASO de type bloquant se fixent à l'ARN et, par encombrement stérique, altère l'épissage ou diminue la traduction de cet ARN. Image de Bennet et al., 2019.

Ce type de stratégie par ASO a notamment permis d'améliorer considérablement les conditions de vie des patients atteints d'amyotrophie spinale proximale (SMA) liée à *SMN1*. Cette maladie neurodégénérative affecte les motoneurones et conduit à des altérations des capacités motrices et respiratoires chez l'enfant. Cette pathologie est due à une mutation perte de fonction du gène *SMN1*, conduisant donc à l'absence d'expression de la protéine SMN1 (Matthijs et al., 1996). Dans 95% de la population générale et chez tous les patients atteints de SMA, il existe un gène quasiment identique au gène *SMN1*, le gène *SMN2*. Toutefois, ce gène *SMN2* présente une variation d'un nucléotide dans son exon 7 qui ne modifie pas sa séquence codant mais qui modifie son épissage. L'absence de cet exon conduit à la production d'une protéine instable ne pouvant pas alors compenser la perte de la protéine SMN1 (Campbell et al., 1997). Un ASO, appelé Nusinersen, permet de bloquer l'accès de la protéine hnRNP à un site intronique régulant l'épissage de l'ARN SMN2. La protéine hnRNP stimule normalement l'excision de l'exon 7, son absence permet ainsi le maintien de l'exon 7 dans la protéine SMN2 et la production d'une protéine fonctionnelle (Figure 51 ; Hua et al., 2008). En 2017, le Nusinersen a été approuvé par la FDA aux Etats-Unis et permet une amélioration notable des capacités motrices des enfants (Aartsma-Rus, 2017 ; Darras et al., 2019).

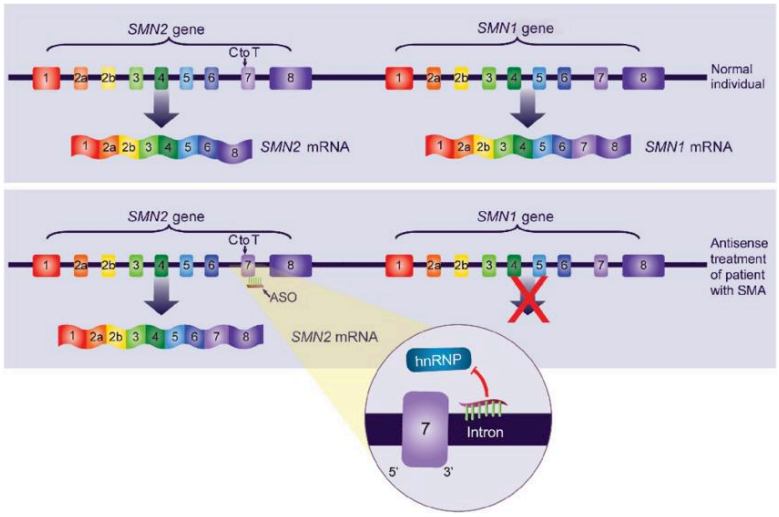


Figure 51. Action du Nusinersen sur l'épissage de SMN2
Le Nusinersen se fixe sur l'ARN SMN2 et empêche la fixation de la protéine hnRNP qui promeut l'éjection de l'exon 7, résultant en une protéine SMN2 similaire à la protéine SMN1. Image de Chiriboga et al., 2016.

Toutefois, cette stratégie par ASO n'est pas toujours couronnée de succès. En effet, deux essais cliniques en phase III recourant à différents ASO (Tominersen, WVE-120101 et WVE-120102) sur des patients touchés par la maladie de Huntington ont été stoppés car sans bénéfice et avec plusieurs effets secondaires après 1 à 2 ans de traitement (Kingwell, 2021 ; Kwon, 2021). Le

Tominersen, un ASO de type gapmer, conduit à la dégradation de la HTT mutée mais également de la HTT non-mutée, conduisant donc à une abolition de l'expression de la protéine HTT. La perte d'expression de cette protéine pouvant peut-être conduire à des effets secondaires, de nouveaux ASO ciblant spécifiquement la HTT-polyQ sont en cours de développement (Kwon, 2021). Néanmoins, les ASO WVE-120101 et WVE-120102 ciblent spécifiquement la forme mutée de la HTT et n'ont pas permis d'améliorer le phénotype des patients, peut-être à cause d'une mauvaise pénétrance des ASO dans les tissus. C'est pourquoi le développement de nouvelles modifications chimiques des ASO pour augmenter leur pénétrance dans les tissus est en cours.

Dans le cas de la SLA/DFT due à une expansion d'hexa-nucléotides GGGGCC dans le gène *C9ORF72*, plusieurs ASO induisant la dégradation ou bloquant la traduction de l'ARN de C9ORF72 contenant les répétitions ont été décrits comme réduisant le nombre de foci d'ARN et / ou diminuant la traduction de ces ARN en protéines DPR en modèles cellulaires et murins (Donnelly et al. 2013, Gendron et al. 2017, Jiang et al. 2016, Lagier-Tourenne et al. 2013, Sareen et al. 2013 ; Tabet et al., 2018). Toutefois, la localisation de l'ASO induisant la dégradation des ARN par la RNase H1 peut influer sur l'expression de la protéine C9ORF72. En effet, lorsque les ASO sont dirigés contre la région en 3' des répétitions GGGGCC, l'expression de la protéine C9ORF72 est diminuée (Donnelly et al. 2013 ; Sareen et al. 2013), ce qui n'est pas le cas lorsque l'ASO est situé en 5' des répétitions GGGGCC (Lagier-Tourenne et al. 2013 ; Jiang et al., 2016). La perte de fonction de la protéine C9ORF72 pouvant jouer un rôle dans le développement de la SLA, un essai clinique utilisant un ASO dirigé contre la région en 5' des répétitions GGGGCC est en cours (BIIB078 ; NCT03626012). Il est à noter que d'autres essais cliniques ciblant des mutations dans les gènes *SOD1* (BIIB067 / Tofersen ; NCT02623699) et *FUS* (ION363 / Jacifusen ; NCT04768972) sont également en cours.

FXTAS est causée par une expansion de répétitions CGG dans la région 5'UTR du gène *FMR1* qui est transcrite en ARN pouvant être ciblé par des ASO, puis traduite en une protéine toxique, FMRpolyG, En cultures primaires de neurones de souris FXTAS ainsi qu'en neurones différenciés à partir de cellules iPS de patients FXTAS, la transfection d'un ASO dirigé contre le codon d'initiation de la protéine FMRpolyG bloque sa traduction et permet ainsi une meilleure survie des cellules (Rodriguez et al., 2020). Cette stratégie d'ASO a également fait ses preuves en modèle murin puisque l'injection intraventriculaire d'un ASO dirigé directement contre les répétitions de nucléotides CGG diminue le nombre d'agrégats formés par la protéine FMRpolyG et améliore le phénotype locomoteur des souris (Derbis et al., 2021). Dans ces 2

études, le type d'ASO utilisé est de type bloquant, c'est-à-dire qu'il empêche la traduction de l'ARNm par le ribosome au niveau des structures doubles brins formées par l'ARNm et l'ASO.

En conclusion, l'étude des mécanismes sous-jacents à ces différentes pathologies pourraient permettre le développement de thérapies ciblées, notamment par l'utilisation d'ASO et / ou d'anticorps dirigés contre les ARN ou les protéines toxiques, respectivement.

Nous allons maintenant aborder la partie résultat, composée de 2 publications portant sur les mécanismes de toxicité des expansions de répétitions, soit CGG localisées dans la région 5'UTR du gène *NOTCH2NLC* responsables de NIID (Publication 1), soit GGGGCC situées dans le premier intron du gène *C9ORF72* causant la SLA/DFT (Publication 2) ainsi que d'une partie comprenant des résultats préliminaires sur les expansions de répétitions CGG localisées dans la région 5'UTR du gène *GIPC1* responsables de l'OPDM2.

RESULTATS

RESULTATS

- ARTICLE 1 -

Translation of GGC repeat expansions into a toxic polyglycine protein in NIID defines a novel class of human genetic disorders : the polyG diseases

Article 1

Translation of GGC repeat expansions into a toxic polyglycine protein in NIID defines a novel class of human genetic disorders: The polyG diseases

Boivin M, Deng J, Pfister V, Grandgirard E, Oulad-Abdelghani M, Morlet B, Ruffenach F, Negroni L, Koebel P, Jacob H, Riet F, Dijkstra AA, McFadden K, Clayton WA, Hong D, Miyahara H, Iwasaki Y, Sone J, Wang Z, Charlet-Berguerand N.

Publié dans *Neuron 2021 Jun 2;109(11):1825-1835.e5. doi: 10.1016/j.neuron.2021.03.038*

La maladie à inclusion intranucléaire neuronale (NIID) est une maladie neurodégénérative rare causée par une expansion de répétitions de tri-nucléotides CGG dans la région 5'UTR du gène *NOTCH2NLC*. Cette pathologie est caractérisée par la présence d'inclusions intranucléaires d'origine inconnue Mon travail de thèse a porté sur le mécanisme de toxicité de cette mutation. Des expériences de transfection cellulaire ont permis de mettre en évidence que, malgré la localisation dans une région 5'UTR dite « non-codante », ces répétitions CGG étaient traduites dans un cadre de lecture unique, produisant alors une nouvelle protéine composée d'une suite de résidus glycines, que nous avons appelé **uN2CpolyG**. Des analyses de spectrométrie de masse ont alors montré que l'initiation de la traduction de cette protéine se fait à un codon ATG en amont des répétitions CGG.

Afin d'attester de l'existence de cette nouvelle protéine uN2CpolyG chez les patients NIID, nous avons développé des anticorps dirigés contre sa partie C-terminale. Des immunofluorescences réalisées sur des lames de cerveaux ainsi que sur biopsies cutanées de patients NIID a révélé la présence de cette protéine polyGlycine dans les agrégats intranucléaires caractéristiques de cette pathologie.

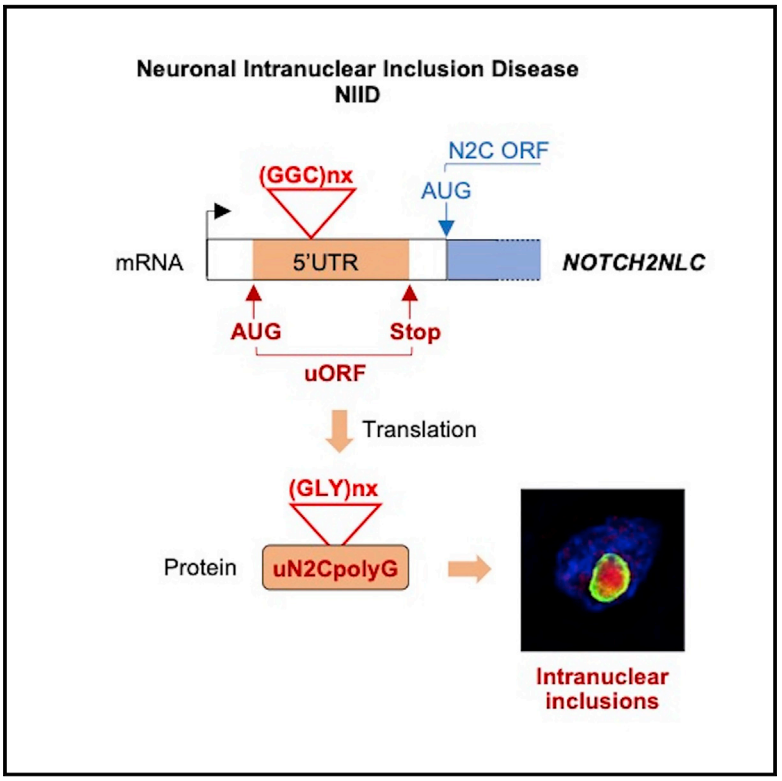
Enfin, l'expression de la protéine uN2CpolyG en culture primaire de neurones d'embryons de souris ou dans le système nerveux central de souris adulte conduit à la formation d'agrégats intranucléaires, à une mort cellulaire et provoque une hyperactivité, une ataxie ainsi qu'une mort prématurée des animaux. Ces résultats suggèrent que la seule expression de la protéine uN2CpolyG est suffisante pour récapituler les caractéristiques principales de NIID.

Neuron

Report

Translation of GGC repeat expansions into a toxic polyglycine protein in NIID defines a novel class of human genetic disorders: The polyG diseases

Graphical abstract



Highlights

- NIID is a neurodegenerative disease caused by expansion of GGC repeats in NOTCH2NLC
- These GGC repeats are translated into a polyglycine (polyG) protein
- The polyG protein is toxic and forms intranuclear inclusions in cells and animals
- Similarities between FXTAS and NIID define a new set of disorders: polyG diseases

Authors

Manon Boivin, Jianwen Deng, Véronique Pfister, ..., Jun Sone, Zhaoxia Wang, Nicolas Charlet-Berguerand

Correspondence

ncharlet@igbmc.fr

In brief

The neurodegenerative disease NIID is caused by an expansion of GGC repeats in NOTCH2NLC. Boivin et al. found that these repeats are translated into a toxic polyglycine (polyG) protein that forms intranuclear inclusions. An identical mechanism exists in FXTAS, unveiling a novel group of genetic pathologies, the polyG diseases.

Boivin et al., 2021, Neuron *109*, 1825–1835
June 2, 2021 © 2021 The Author(s). Published by Elsevier Inc.
https://doi.org/10.1016/j.neuron.2021.03.038

CellPress

Neuron

CellPress
OPEN ACCESS

Report

Translation of GGC repeat expansions into a toxic polyglycine protein in NIID defines a novel class of human genetic disorders: The polyG diseases

Manon Boivin,[1] Jianwen Deng,[2] Véronique Pfister,[1] Erwan Grandgirard,[1] Mustapha Oulad-Abdelghani,[1] Bastien Morlet,[1] Frank Ruffenach,[1] Luc Negroni,[1] Pascale Koebel,[1] Hugues Jacob,[1] Fabrice Riet,[1] Anke A. Dijkstra,[3] Kathryn McFadden,[4] Wiley A. Clayton,[5] Daojun Hong,[6] Hiroaki Miyahara,[7] Yasushi Iwasaki,[7] Jun Sone,[7,8] Zhaoxia Wang,[2] and Nicolas Charlet-Berguerand[1,9,*]

[1]Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), INSERM U 1258, CNRS UMR 7104, University of Strasbourg, 67404 Illkirch, France
[2]Department of Neurology, Peking University First Hospital, Beijing 100034, China
[3]Department of Pathology, Amsterdam University Medical Centre, Amsterdam Neuroscience, VUmc, Amsterdam, the Netherlands
[4]Department of Pathology, IWK Health Centre, Halifax, NS B3K 6R8, Canada
[5]Department of Pathology, University of Pittsburgh Medical Center, Pittsburgh, PA 15213, USA
[6]Department of Neurology, First Affiliated Hospital of Nanchang University, Nanchang, China
[7]Department of Neuropathology, Institute for Medical Science of Aging, Aichi Medical University, Nagakute, Japan
[8]Department of Neurology, Suzuka National Hospital, Suzuka 513-8501, Japan
[9]Lead contact
*Correspondence: ncharlet@igbmc.fr
https://doi.org/10.1016/j.neuron.2021.03.038

SUMMARY

Neuronal intranuclear inclusion disease (NIID) is a neurodegenerative disease characterized by the presence of intranuclear inclusions of unknown origin. NIID is caused by an expansion of GGC repeats in the 5′ UTR of the NOTCH2NLC (N2C) gene. We found that these repeats are embedded in a small upstream open reading frame (uORF) (uN2C), resulting in their translation into a polyglycine-containing protein, uN2CpolyG. This protein accumulates in intranuclear inclusions in cell and mouse models and in tissue samples of individuals with NIID. Furthermore, expression of uN2CpolyG in mice leads to locomotor alterations, neuronal cell loss, and premature death of the animals. These results suggest that translation of expanded GGC repeats into a novel and pathogenic polyglycine-containing protein underlies the presence of intranuclear inclusions and neurodegeneration in NIID.

INTRODUCTION

More than 40 inherited human diseases are caused by expansion of microsatellites, short DNA tandem repeats variable in size and sequences. These microsatellite expansions are pathogenic by three main non-exclusive mechanisms (Nelson et al., 2013; Gao and Richter, 2017; Rodriguez and Todd, 2019). First, they can promote DNA epigenetic changes that inhibit transcription, resulting in loss of function of the allele carrying the repeats. Second, RNA transcripts containing the expanded repeats can bind to specific RNA binding proteins, potentially altering their localization and function. Third, expanded repeats can be translated by canonical initiation to AUG or near-cognate start codons or by translation initiation directly within the repeat through a novel mechanism called repeat-associated non-AUG (RAN) translation (Zu et al., 2011), into proteins containing a pathogenic stretch of repeated amino acids. Progress in long-read and whole-genome sequencing has recently unveiled a dozen novel microsatellite expansions located in the "non-coding" part of the human genome as pathogenic causes, notably an intronic AG-rich repeat expansion in the *RFC1* gene in cerebellar ataxia with neuropathy and bilateral vestibular areflexia syndrome (CANVAS) (Cortese et al., 2019; Rafehi et al., 2019); 6 similar intronic AT-rich repeat expansions in benign adult familial myoclonic epilepsy (BAFME) 1–6 and 5 similar 5′ UTR-embedded GC-rich repeat expansions in fragile X-associated tremor/ataxia syndrome (FXTAS), neuronal intranuclear inclusion disease (NIID), oculopharyngodistal myopathy (OPDM), and oculopharyngeal myopathy with leukoencephalopathy (OPML), neuromuscular and neurodegenerative syndromes with some overlapping symptoms and similar histopathological features (Hagerman et al., 2001; Ishiura et al., 2019; Sone et al., 2019; Deng et al., 2019, 2020; Tian et al., 2019; Xi et al., 2021; review in Ishiura and Tsuji, 2020).

Among these later disorders, NIID, also known as neuronal intranuclear hyaline inclusion disease (NIHID) and intranuclear

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

inclusion body disease (INIBD), is a rare genetic disease characterized by the presence of intranuclear inclusions in the central and peripheral nervous systems and in multiple other organs (Lindenberg et al., 1968; Munoz-Garcia and Ludwin, 1986; Sone et al., 2016). NIID age of onset is variable, and three subgroups (infant, juvenile, and adult) have been defined (Takahashi-Fujigasaki, 2003). Clinically, NIID is also tentatively divided in three subgroups: dementia dominant, parkinsonism dominant, and muscle weakness dominant. However, NIID symptoms are highly heterogenous, and overlap between subgroups is observed frequently, with variable muscle weakness associated with various dysfunctions of the central and peripheral nervous systems, which can include progressive dementia and cognitive impairment, parkinsonism, cerebellar ataxia, sensory disturbance, autonomic dysfunction, and/or peripheral neuropathy (Takahashi-Fujigasaki, 2003; Sone et al., 2016; Takahashi-Fujigasaki et al., 2016). Moreover, individuals with NIID with atypical presentation, such as essential tremors, multiple-systems atrophy, and amyotrophic lateral sclerosis, and various acute symptoms, including stroke-like episodes, epileptic seizures, and/or encephalitic episodes, have also been reported (Sone et al., 2016; Fang et al., 2020; Li et al., 2020; Sun et al., 2020; Yuan et al., 2020). As a consequence of these diverse ages of onset and clinical presentations, NIID diagnosis is most often confirmed by the widespread presence of characteristic eosinophilic intranuclear inclusions in neurons and glial cells in the central and peripheral nervous systems and in various other tissues (Chen et al., 2020a; Liu et al., 2008; Sone et al., 2005, 2011, 2014). These intranuclear inclusions are immunoreactive for various markers of the proteasomal and autophagic degradation pathways, including ubiquitin, sumo, and p62 (Pountney et al., 2003; Mori et al., 2012; Nakamura et al., 2014; Sone et al., 2016). Importantly, an expansion of GGC repeats located in the 5′ UTR of the NOTCH2NLC (Notch 2 N-terminal like C, N2C) gene has been found recently to be associated with individuals with familial and sporadic NIID, mostly in people of Asian origin (Sone et al., 2019; Ishiura et al., 2019; Deng et al., 2019; Tian et al., 2019). The NOTCH2NLA, NOTCH2NLB, and NOTCH2NLC genes are human-specific paralogs of *NOTCH2* exons 1–5, which encode Notch 2 N-terminal like (N2L) proteins that regulate Notch signaling to expand human neuronal progenitors during brain development (Fiddes et al., 2018; Suzuki et al., 2018). Alteration of NOTCH2NLC (N2C) protein function in NIID is unlikely because GGC repeats are located more than 100 nt upstream of the ATG start codon initiating the N2C open reading frame (ORF). Furthermore, NOTCH2NLC mRNA levels are unaltered in individuals with NIID (Sone et al., 2019; Ishiura et al., 2019; Tian et al., 2019). Thus, it remains to be determined how expansion of GGC repeats embedded in a predicted non-coding genomic region can lead to formation of intranuclear inclusions and cause neuronal cell death.

Here we find that the NOTCH2NLC GGC repeats are embedded in a small upstream open reading frame (uORF) located ahead of the main N2C ORF and encodes a small protein, uN2C. Of clinical importance is that GGC repeats embedded into the uN2C ORF are translated into a polyglycine stretch, resulting in expression of a uN2C polyglycine-containing protein (uN2CpolyG) in NIID. Antibodies developed against uN2CpolyG revealed its presence in the typical intranuclear inclusions in skin and brain sections of individuals with NIID. Furthermore, expression of uN2CpolyG in cell and animal models drives formation of p62-positive inclusions. Finally, expression of uN2CpolyG in mice is toxic, resulting in locomotor alterations, neuronal cell loss, and a reduced lifespan. These results are reminiscent of FXTAS, where an expansion of GGC repeats is embedded in a small upstream ORF, resulting in expression of a polyglycine-containing protein that, like uN2CpolyG, forms intranuclear inclusions and is toxic in cell and animal models (Hukema et al., 2015; Sellier et al., 2017; Todd et al., 2013). These data suggest the existence of a novel class of human genetic diseases where expanded GCC repeats are translated into pathogenic polyglycine (polyG) proteins.

RESULTS

NOTCH2NLC GGC repeats are translated into a polyG-containing protein

Because translation of expanded repeats into pathogenic stretches of repeated amino acids is an established mechanism of pathogenicity in microsatellite diseases, we investigated the potential translation of GGC repeats in NIID. First we cloned GGC repeats embedded in the human NOTCH2NLC exon 1 sequence and fused these repeats to GFP in the three possible frames (Figures 1A and S1A). These frames were called polyG, polyalanine (polyA), and polyarginine (polyR), according to the stretch of amino acids potentially encoded by the repeats. Cell transfection followed by immunoblotting against GFP or observation of GFP fluorescence indicated that NOTCH2NLC GGC repeats were translated predominantly in the glycine frame (Figures 1B and 1C). qRT-PCR quantification indicated that GGC repeats cloned in the three frames are similarly transcribed (Figure S1B). Furthermore, cell transfection of expanded GGC repeats cloned downstream of an artificial ATG start codon in the alanine or arginine frames and fused to GFP results in expression of ATG-driven polyA and polyR GFP-tagged proteins (Figures S1C and S1D). These controls indicate that the limited expression of polyA or polyR from the NOTCH2NLC 5′ UTR cannot be accounted for by lack of their expression at the RNA level or a technical bias impairing their detection and that the NOTCH2NLC GGC repeats are mainly translated into a polyG-containing protein, which was called uN2CpolyG (upstream of N2C polyG-containing protein).

To better characterize this protein, GGC repeats embedded in human NOTCH2NLC exon 1 and fused to GFP in the glycine frame were expressed in HEK293 cells, GFP immunoprecipitated, trypsin digested, and analyzed by mass spectrometry. Liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis confirmed that NOTCH2NLC GGC repeats are translated into a glycine stretch (Figures 1D and S1E). As a further control, treatment with lysostaphin, a glycylglycine endopeptidase, eliminated the N-terminal glycine stretch of uN2CpolyG-GFP, confirming its polyG nature (Figure 1E). Importantly, MS also revealed that GGC repeats are translated through initiation to a canonical AUG start codon located 15 nt before the repeats (Figures 1D and S1E). Deletion of this ATG codon abolished uN2CpolyG expression (Figure 1F). These data suggest

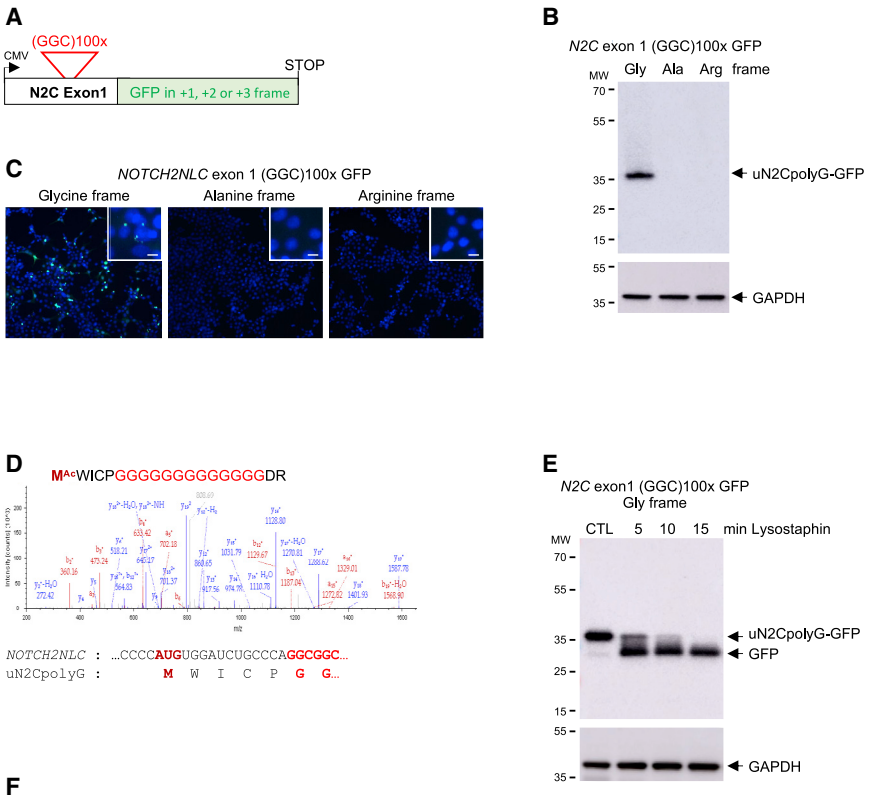


Figure 1. *NOTCH2NLC* GGC repeats are translated in a polyG-containing protein

(A) Schematic of NOTCH2NLC exon 1 with GGC repeats fused to GFP in the glycine, alanine, or arginine frame.

(B and C) Immunoblot against GFP (B) or direct fluorescence (C) of HEK293 cells transfected for 24 h with GGC repeats embedded in NOTCH2NLC exon 1 and fused to GFP in the three possible frames.

(D) Top panel: N-terminal sequence and corresponding LC-MS/MS spectra of GFP-immunoprecipitated and trypsin-digested protein expressed from uN2C-GFP-transfected HEK293 cells. Bottom panel: nucleotide and amino acid sequences corresponding to the NOTCH2NLC upstream ORF (uN2C) N terminus.

(E) Immunoblot against GFP or the glyceraldehyde 3-phosphate dehydrogenase (GAPDH) of lysostaphin-digested proteins extracted from uN2CpolyG-GFP-transfected HEK293 cells.

(F) Immunoblot against GFP or the GAPDH of proteins extracted from wild-type or mutant (ΔATG) uN2CpolyG-GFP-transfected HEK293 cells.

(G) Top panel: schematic of NOTCH2NLC exons 1 and 2. Bottom panel: amino acid sequence of the NOTCH2NLC upstream ORF (uN2C).

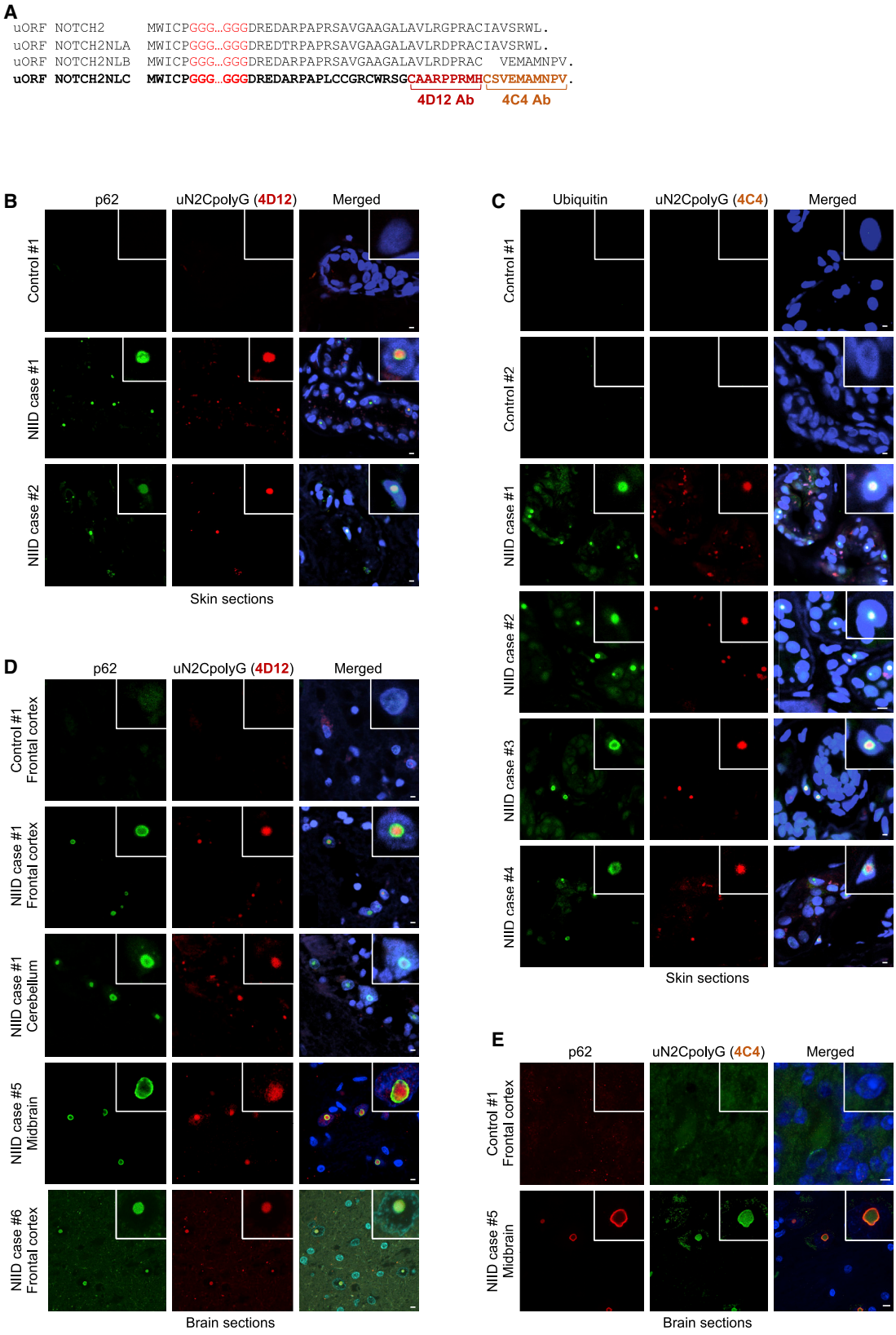
See also Figure S1.

that the NOTCH2NLC 5′ UTR contains a small upstream ORF, uN2C, that spans exons 1 and 2 and ends 8 nt before the ATG start site of the main NOTCH2NLC (N2C) ORF (Figures 1G and S1F). Consequently, the uN2CpolyG protein is composed of a short 5-amino-acid N terminus, a variable central glycine stretch corresponding to the number of GGC repeats, and a 38-amino-acid C-terminal part (Figure 1G).

The uN2polyG protein is present in NIID intranuclear inclusions

To confirm that NOTCH2NLC GGC repeats are translated into a polyG-containing protein in NIID, we developed two antibodies (4D12 and 4C4) directed against the uN2C C terminus, whose amino acid sequence differs from NOTCH2 and NOTCH2NLA predicted upstream ORFs (Figures 2A and S2A). Antibody specificity was confirmed by immunoblot and immunofluorescence (Figures S2B–S2D). Importantly, immunofluorescence using these antibodies revealed the presence of uN2CpolyG in the characteristic p62- and ubiquitin-positive intranuclear inclusions in skin and brain sections of individuals with NIID, carriers of an expansion of GGC repeats in NOTCH2NLC (N2C-NIID), but no staining in control age-matched individuals (Figures 2B–2E and S2E). As a control for uN2CpolyG staining specificity, identical results were observed with 4C4 and 4D12 antibodies, which are directed against different epitopes of the uN2C protein. Moreover, neither 4D12 nor 4C4 antibodies labeled p62-positive intranuclear inclusions of FMRpolyG in brain sections of FXTAS, a neurodegenerative disease caused, like NIID, by expanded GGC repeats translated into a polyG-containing protein (Figure S2F). Finally, quantification indicated quasi-complete co-localization of uN2CpolyG with p62-positive intranuclear inclusions in NIID (Figure S2G). These results demonstrate that NOTCH2NLC GGC repeat expansions are translated into a novel polyG-containing protein that is found in the typical intranuclear inclusions in individuals with N2C NIID.

During this analysis, we noted that various European individuals with NIID with typical p62-positive intranuclear inclusions were nevertheless negative for uN2CpolyG staining (Figure S2H; data are summarized in Table S1). Genetic analyses revealed that these individuals are negative for the NOTCH2NLC GGC expansion, which is consistent with a recent report showing that most Europeans with NIID are negative for this mutation



(legend on next page)

(Chen et al., 2020b). These unexpected results suggest that NIID is a heterogenous syndrome and that other mutations causing subtypes of this disorder remain to be identified, notably in European individuals with NIID.

Expression of uN2CpolyG is pathogenic in cells

To further study the uN2CpolyG protein, we cloned the full NOTCH2NLC uORF, from its ATG start codon in exon 1 to its penultimate codon in exon 2, containing its 38 C-terminal amino acids, with a control (12×) or an expanded (100×) size of GGC repeats, and fused this uN2C ORF to GFP or a hemagglutinin (HA) tag (Figure S3A). Cell transfection of the GFP-tagged N2C uORF with 12 or 100 GGC repeats indicated that their expression occurs independent of repeat size (Figure S3B). In contrast, fusion of the N2C uORF to a smaller HA tag (~1 kDa) resulted in detection of uN2CpolyG with 100 GGC repeats but limited expression with 12 repeats (Figures S3C and S3D). This is characteristic of short upstream ORFs that are translated into small and often unstable peptides that can be stabilized and detected when fused with a large tag such as GFP (Aspden et al., 2014). In support of this hypothesis, cell treatment with Bafilomycin A1, which inhibits the autophagy degradation pathway, increased expression of the small uN2C-HA protein with 12 GGC repeats (Figure S3D). These results suggest that translation of the N2C uORF occurs independent of the length of the GGC stretch. However, in absence of an expansion, the uN2C protein, like most uORF proteins, is small and unstable and, thus, expressed at low levels in control individuals, whereas uN2C is stabilized by the polyG expansion and accumulates in individuals with N2C NIID.

Next we investigated whether the uN2CpolyG protein could be responsible for the typical inclusions and the neuronal cell dysfunctions observed in NIID. Importantly, immunofluorescence analyses indicated that sole expression of the uN2CpolyG protein in embryonic mouse cortical neuronal cell cultures was sufficient to form cytoplasmic and intranuclear inclusions, which are p62-positive (Figure 3A). Inclusions formation is likely driven by the polyG expansion because expression of a pure polyG stretch (ATG polyG-GFP), deleted of any uN2C sequence, also forms p62-positive cellular inclusions (Figure 3A). Identical results were obtained in immortalized cell lines (Figures S3E–S3G). Further analyses indicated that accretion of uN2CpolyG into cellular aggregates is progressive and parallels its accumulation in the insoluble fraction of transfected cell lysates (Figures S3F and S3G). Because protein aggregates can be degraded by macroautophagy, we tested whether compounds known to activate this catabolic process may prevent accumulation of the uN2CpolyG protein. Interestingly, inhibition of the mammalian target of rapamycin (mTOR) kinase decreases uN2CpolyG accumulation (Figure 3B). Finally, expression of the uN2CpolyG protein is toxic and leads to neuronal cell death in embryonic mouse cortical neuronal cell cultures (Figure 3C) and in GT1-7 immortalized neuronal cells (Figure S3H). Importantly, expression of the uN2C protein with a control number of glycine repeats (GGC 12x) was not overtly toxic, whereas expression of a polyG stretch in isolation and deleted of any uN2C sequence was sufficient to cause cell death (Figures 3C and S3H). These data indicate that expression of uN2CpolyG is pathogenic in cell culture, with its polyG stretch driving aggregation and toxicity.

To further investigate the mechanism of uN2CpolyG toxicity, we searched for potential proteins interacting with it. Cell transfection of uN2CpolyG-GFP followed by GFP immunoprecipitation and MS identified various associated proteins with clear enrichment for the Ku70 and Ku80 proteins (Figure S4A; Table S2). Ku70 and Ku80 form a heterodimeric ring that initiates nonhomologous DNA end joining (NHEJ) repair at DNA double-strand breaks (Grundy et al., 2014; Fell and Schild-Poulter, 2015). Co-immunoprecipitation confirmed that the control uN2C and uN2CpolyG proteins interact with the endogenous Ku proteins (Figure S4B). However, this interaction is independent of the glycine stretch because ku70 and Ku80 interact neither with the polyG-containing protein expressed in FXTAS (FMRpolyG) nor with a pure polyG stretch (ATG polyG-GFP) (Figure S4B). Mutation analyses of the uN2C ORF revealed that a central amino acid sequence that is absent from NOTCH2, NOTCH2NLA, and NOTCH2NLB putative upstream ORFs is essential for the interaction of uN2C with the Ku proteins (Figure S4C). Next we investigated localization of the uN2C and uN2CpolyG proteins upon induction of DNA damage. Interestingly, most of uN2CpolyG was immobilized in inclusions, with little protein recruited to sites of DNA damage (Figure S4D). In contrast, the control uN2C protein was recruited within second to sites of laser microirradiation-induced DNA damage. Mutation analyses indicated that this recruitment is dependent on the uN2C amino acid sequence required to interact with the Ku proteins (Figure S4D). These data suggest that the 5′ UTR of NOTCH2NLC contains an upstream ORF that encodes a small protein, uN2C, which binds to the Ku proteins and is recruited at DNA damage sites. In contrast, upon GGC repeat expansion, uN2C encodes a mutant polyG-containing protein, uN2CpolyG, which is immobilized in aggregates and, thus, cannot join DNA damage. These data question whether uN2C or uN2polyG may affect DNA repair. Immunoblotting and immunofluorescence against phospho-Ser139 H2AX (γH2AX), a marker of DNA double-strand breaks, indicated faster kinetics of DNA repair in irradiated cells expressing the uN2C control protein with 12 glycine as compared to control cells (Figures S4E and S4F). In contrast, the polyG expansion in uN2C leads to an inactive protein that is unable to promote DNA repair (Figures S4E and S4F). We thus tested whether this impaired activity may translate into any

Figure 2. The uN2polyG protein is present in NIID intranuclear inclusions

(A) Alignment of NOTCH2 and NOTCH2NLA, NOTCH2NLB, and NOTCH2NLC putative upstream ORFs. Brackets indicate the amino acid sequences against which the 4D12 and 4C4 antibodies are directed.

(B–E) Immunofluorescence against uN2CpolyG using 4D12 (B and D) or 4C4 (C and E) antibody and p62 (B, D, and E) or ubiquitin (C) on skin (B and C) and brain (D and E) sections of carriers of the *NOTCH2NLC*-GGC repeat expansion (N2C NIID) or age-matched control individuals. Scale bars, 10 μm. Nuclei were counterstained with DAPI.

See also Figure S2 and Table S1.

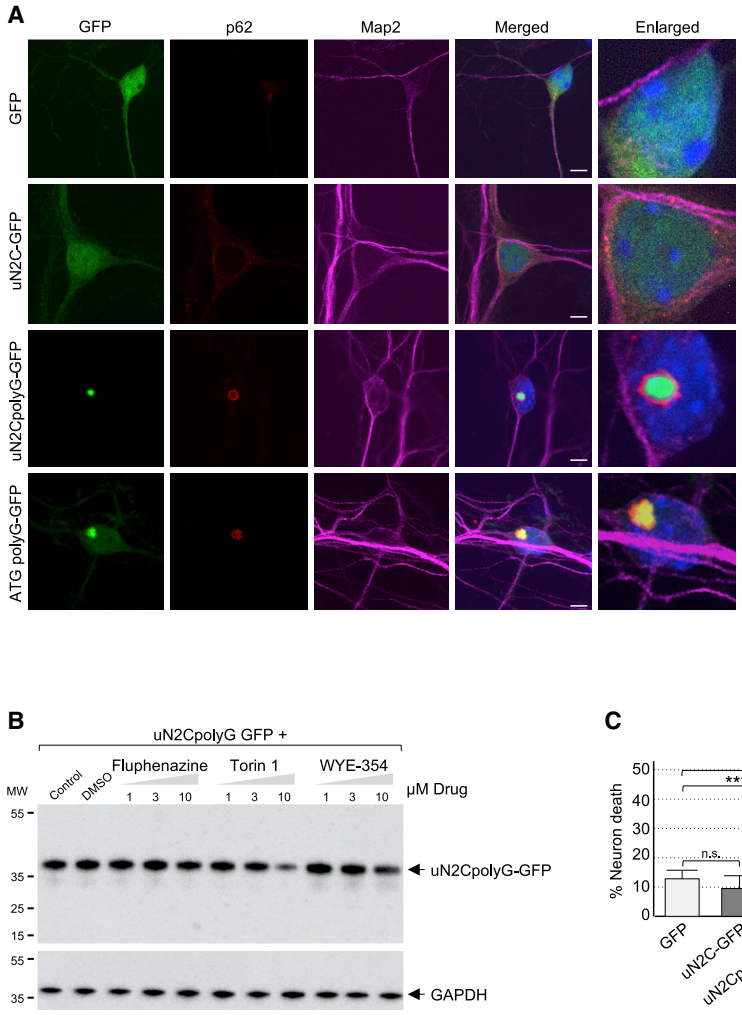


Figure 3. Expression of uN2CpolyG is toxic in cell culture

(A) GFP fluorescence and immunofluorescence against p62 and Map2 of cortical neuronal cell cultures from mouse embryos transduced for 24 h with GFP, uN2C-GFP (12 GGC), uN2CpolyG-GFP (100 GGC), or ATG polyG-GFP (70 GGC) AAV2/PHP.eB. Scale bars, 10 μm. Nuclei were counterstained with DAPI.

(B) Immunoblot against GFP or the GAPDH of proteins extracted from uN2CpolyG-GFP-transfected HEK293 cells treated for 15 h with the indicated drugs.

(C) Cell viability of cortical neuronal cell cultures from mouse embryos transduced with GFP, uN2C-GFP, uN2CpolyG-GFP, or ATG polyG-GFP AAV2/PHP.eB. Error bars indicate SEM. Student's t test, ***$p < 0.001$.

See also Figures S3, S4, and Table S2.

pathological consequences in NIID. However, immunofluorescence analyses revealed no overt mislocalization of the Ku proteins or any evident accumulation of γH2AX-positive DNA damage in uN2CpolyG-expressing cells or in skin and brain sections of individuals with adult-onset N2C NIID. These results indicate that the small uN2C protein is potentially a novel regulator of DNA damage response, whereas in NIID, a polyG expansion alters uN2C localization and DNA repair activity, but this reduced function is likely compensated by the second NOTCH2NLC allele and is not sufficient to drive overt DNA repair alterations in individuals with NIID.

Expression of uN2CpolyG is pathogenic in animals

To evaluate the pathological consequences of uN2polyG expression in animals, we injected mice with adeno-associated virus (AAV) particles expressing GFP-tagged uN2C with a control (12×) or pathogenic (100×) size of polyG. To target the central nervous system, we used an AAV serotype (PHP.eB) that crosses the blood-brain barrier upon intravenous injection. Furthermore, to focus on polyG protein toxicity and exclude any potential RNA gain-of-function mechanism, the GGC repeat sequence was modified to include alternative codons (GGA, GGT, and GGG) that also encode for glycine and, thus, avoid production of a pure GGC RNA hairpin. qRT-PCR indicated similar levels of expression between AAV GFP-, control uN2C-GFP-, and uN2CpolyG-GFP-injected mice (Figure S5A). To determine the consequences of uN2CpolyG production in mice, we conducted a series of locomotor assays. Three months after AAV injection, we noted a progressive alteration of motor performance and coordination of uN2CpolyG-expressing mice. First, AAV uN2CpolyG-injected mice were unable to sustain hindlimb extension under tail suspension compared with control AAV GFP- or uN2C-injected mice (Figure 4A). Furthermore, uN2CpolyG-expressing mice had an increased likelihood of falling from a rotarod (Figure 4B), showing increased foot faults and slips on a notched bar (Figure 4C; Videos S1, S2, and S3), and were hyperactive in the open field arena (Figure 4C) compared with control mice expressing the GFP or N2C uORF with a normal size of GGC repeats. These alterations likely originate from specific neuronal dysfunction and not from global health deterioration because animals expressing the uN2CpolyG protein were overall healthy with normal grip strength and normal weight (Figures S5B and S5C). We noted none of these locomotor changes 1.5 months after AAV injection, demonstrating that these alterations are progressive. Finally, expression of the polyG protein is deleterious because mice expressing uN2CpolyG die around 4–6 months after injection, whereas GFP- and control uN2C-injected mice exhibit normal longevity and are indistinguishable from control non-injected mice (Figure 4E). Importantly, expression of uN2CpolyG in mice leads to widespread formation of intranuclear inclusions, which are p62-, ubiquitin-, and sumo-positives (Figures 4F, 4G, and S5D–S5F), reproducing the characteristic histopathological features of NIID. Finally, we observed

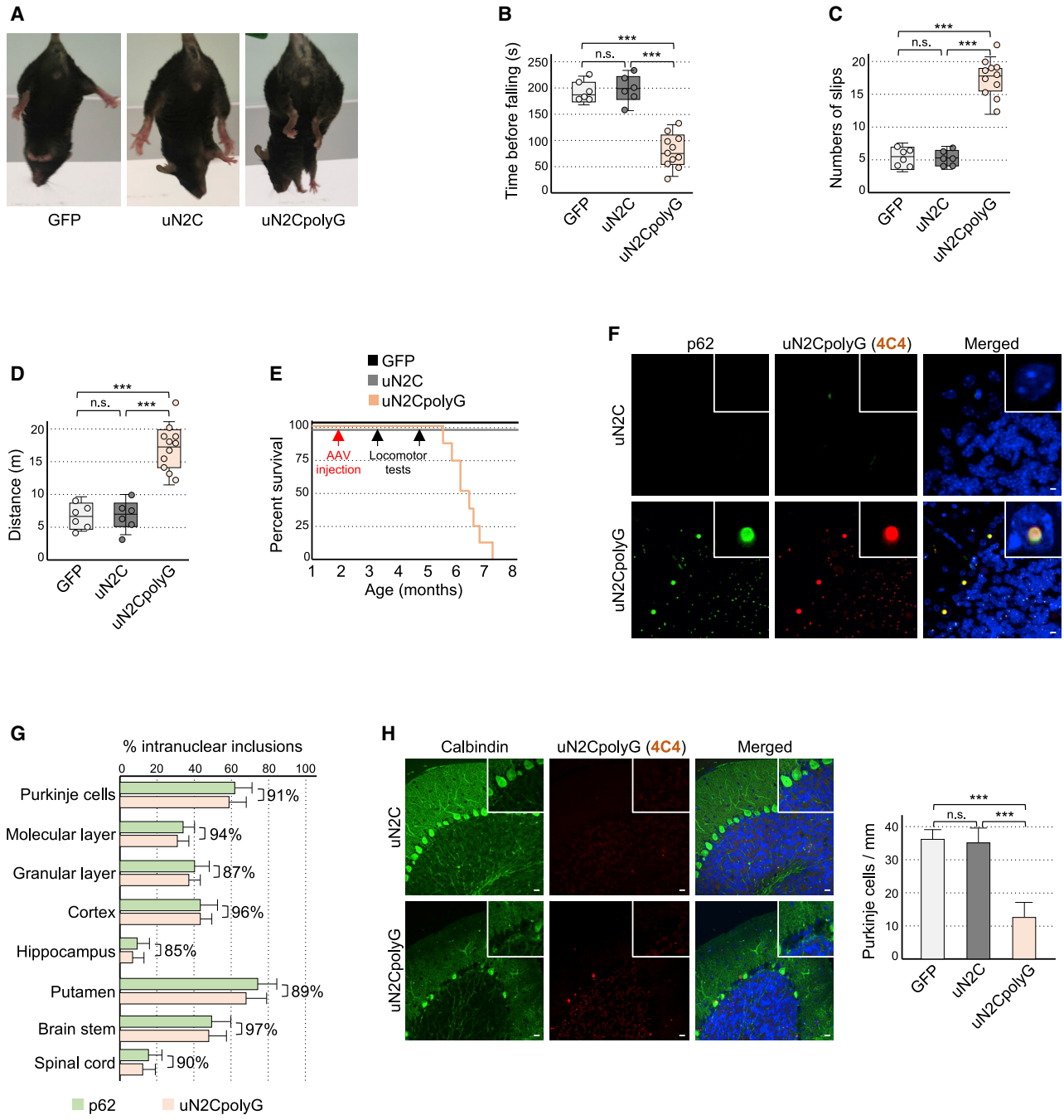


Figure 4. Expression of uN2CpolyG is pathogenic in animals

(A) 60-s tail suspension test shows hind limb clasping in uN2CpolyG-GFP-expressing mice compared with GFP or uN2C-GFP control animals.

(B–D) Time before falling from a rotating rod (B), numbers of paw slips and errors in the notched bar test (C), and maximal distance traveled during 30 min in an open field (D) for AAV2/PHP.eB GFP-injected (n = 6), uN2C-GFP-injected (n = 6), and uN2CpolyG-GFP-injected (n = 11) male mice tested 3 months after injection.

(E) Kaplan-Meier survival curve of AAV2/PHP.eB GFP-injected (n = 6), uN2C-GFP-injected (n = 6), and uN2CpolyG-GFP-injected (n = 8) male mice. Dates of AAV injection and locomotor tests are indicated by arrows.

(F) Immunofluorescence against uN2CpolyG and p62 on cerebellum areas of uN2C-GFP- and uN2CpolyG-GFP-expressing mice sacrificed 2 months after AAV injection. Scale bars, 10 μm.

(G) Quantification of p62- or uN2CpolyG-positive intranuclear inclusions in different brain regions of uN2CpolyG-GFP-expressing mice. Brackets indicate the percentage of co-localization between p62- and uN2polyG-positive intranuclear inclusions. N = 3 mice; at least 200 nuclei were counted per brain region and per animal.

(legend continued on next page)

neuronal cell death, notably loss of Purkinje cells, in mice expressing the uN2CpolyG protein (Figures 4H and S5G), which is consistent with the progressive loss of motor balance and coordination observed in these animals (Figures 4B and 4C; Videos S1, S2, and S3). Consistent with neurodegeneration, Gfap staining indicated increased neuroinflammation in brain sections of uN2CpolyG-injected mice (Figure S5H). Overall, these data suggest that expression of the uN2CpolyG protein in animals is toxic and is sufficient to generate the typical NIID intranuclear inclusions (Figure S6).

DISCUSSION

NIID is a rare neurodegenerative disorder caused by an expansion of GGC repeats located in the 5′ UTR of the NOTCH2NLC gene (Sone et al., 2019; Ishiura et al., 2019; Deng et al., 2019; Tian et al., 2019). The NOTCH2NLA, NOTCH2NLB, and NOTCH2NLC genes reside in the 1q21.1 locus and result from partial duplication of the NOTCH2 N-terminal part, which encodes six epidermal growth factor (EGF)-like domains but excludes NOTCH2 peptide signal, transmembrane, and cytoplasmic domains. Important for hominid brain size evolution, NOTCH2NL proteins regulate Notch signaling and expand human cortical progenitors (Fiddes et al., 2018; Suzuki et al., 2018). Furthermore, genomic duplications or deletions encompassing the NOTCH2NLA and/ or NOTCH2NLB genes lead to neurodevelopmental syndromes, whereas a GGC expansion in the NOTCH2NLC gene causes a neurodegenerative disease. These results highlight the importance of the *NOTCH2NL* gene for neuronal cells in humans but question how a repeat expansion embedded in a sequence predicted to be "non-coding" can be pathogenic.

Here we found that the NOTCH2NLC GGC repeats are embedded in a small upstream ORF, resulting in their translation into a polyG-containing protein that forms intranuclear inclusions and is toxic in cell and animal models (Figure S6). Immunoblotting and direct GFP observation indicate that NOTCH2NLC GGC repeats are mainly translated in the glycine frame through canonical initiation at an AUG start codon, but these assays may not be sensitive enough to detect low levels of polyA or polyR-containing proteins translated from non-canonical translation initiation in the repeats; thus, a contribution of RAN translation in NIID cannot be formally excluded. Translation of a small upstream ORF containing an expansion of GGC repeats into a toxic polyG protein is reminiscent of another neurodegenerative disorder, FXTAS, which is caused by an expansion of 70–200 CGG repeats located in the 5′ UTR of the *FMR1* gene (Hagerman et al., 2001). These repeats are embedded in a short upstream ORF that is translated through canonical initiation at an ACG near-cognate start codon into a polyG-containing protein, FMRpolyG, which, like uN2CpolyG, forms p62-positive intranuclear inclusions and causes neuronal cell death in cell and animal models (Todd et al., 2013; Hukema et al., 2015; Sellier et al., 2017). NIID and FXTAS share some common clinical features and quasi-indistinguishable histopathological characteristics with remarkably similar intranuclear inclusions (Gelpi et al., 2017; Lim et al., 2020). Thus, like neurodegenerative polyQ diseases, which are caused by CAG repeat expansions embedded in diverse ORFs that are translated into toxic polyglutamine-containing proteins, we propose that NIID and FXTAS represent a novel class of disorders, polyG diseases, where expansions of GGC repeats embedded in diverse upstream ORFs are translated into toxic polyG-containing proteins (Figure S6). Of interest, GGC repeat expansions located in the 5′ UTRs of the *LRP12*, *GIPC1*, and *NUTM2E* (also known as FAM22E) genes have been recently identified as the cause of the OPDM and OPML neuromuscular and neurodegenerative disorders (Ishiura et al., 2019; Deng et al., 2020; Xi et al., 2021). OPDM and OPML share some overlapping symptoms and similar histopathological features with NIID and FXTAS. Furthermore, GGC repeat expansions in NOTCH2NLC have been identified recently in individuals with oculopharyngodistal myopathy with neurological symptoms, highlighting the overlap between these diseases (Ogasawara et al., 2020; Yu et al., 2021). Whether these newly identified diseases caused by GGC repeats are pathogenic through a common mechanism, translation of ill-defined short upstream ORFs into toxic polyG-containing proteins that form intranuclear inclusions, is an exciting question for the future (Figure S6).

Another topic of discussion is the relationship between GGC repeat length and clinical features. In polyQ diseases, a clear correlation exists between the size of the CAG repeat expansion and the age of onset and disease severity. In NIID, it is noteworthy that intermediate sizes (40–80) of NOTCH2NLC GGC repeats increased the susceptibility to develop the parkinsonism-dominant form of NIID and were found recently in individuals with Parkinson's disease (Ma et al., 2020), whereas carriers of longer GGC expansions are more at risk to develop the dementia- and muscle weakness-dominant forms of NIID (Sone et al., 2019; Ishiura et al., 2019; Deng et al., 2019; Tian et al., 2019). However, an unambiguous correlation between repeat size and clinical severity might be limited to a narrow interval for 5′ UTR-embedded GGC repeat diseases because large expansions may trigger a potential "protective" mechanism because of their potential methylation and close proximity with promoters. Indeed, expansions over 200 CGG repeats in the 5′ UTR of the X-linked *FMR1* gene promote epigenetic DNA changes that silence *FMR1* promoter, resulting in decreased expression of the uORF encoding the toxic FMRpolyG protein but also in decreased expression of the main ORF encoding the synaptically important FMRP protein, whose loss cannot be compensated by a second allele in males, ultimately causing the

(H) Left panel: immunofluorescence against uN2CpolyG and calbindin on the cerebellum of uN2C-GFP- and uN2CpolyG-GFP-expressing mice sacrificed 4 months after AAV injection. Scale bars, 20 μm. Right panel: quantification of Purkinje cell numbers in GFP-expressing (n = 4), uN2C-GFP-expressing (n = 4), or uN2CpolyG-GFP-expressing (n = 4) mice.

In box-and-whisker plots, box upper and lower limits represent the 25th and 75th percentiles, whiskers represent minimum and maximum values, and a horizontal line across the box represents the median. Bar graphs indicate standard error of the mean (SEM). Student's t test, ***$p < 0.001$. Nuclei were counterstained with DAPI.

See also Figures S5 and S6 and Videos S1, S2, and S3.

neurodevelopmental fragile X syndrome (FXS). In NIID, if large GGC repeat expansions promote similar epigenetic changes, then this would result in silencing of the NOTCH2NLC allele carrying the GGC expansion, abolishing expression of the toxic uN2CpolyG protein, whereas decreased expression of the main ORF encoding the N2C protein would be likely non-pathogenic, as compensated by the second NOTCH2NLC allele. In short, because of a potential "protective" transcriptional silencing mechanism promoted by a high number of repeats, it is possible that 5′ UTR-embedded GC-rich repeat diseases (FXTAS, NIID, OPDM, and OPML) may not systematically follow a classic autosomal dominant inheritance. This potential promoter silencing mechanism may also explain why these diseases do not show evident anticipation (increased severity and/or younger age of onset with increased number of repeats) over a certain threshold of repeats.

Regarding the potential mechanism of toxicity of these polyG proteins, how they cause cell death is unclear. We found that the control uN2C protein interacts with the Ku70 and Ku80 proteins and activates DNA repair. Expansion of the polyG stretch impairs uN2C localization and interaction with the Ku proteins, resulting in a NHEJ-inactive uN2CpolyG protein. Alterations of DNA repair mechanisms can lead to various neurodegenerative syndromes. However, potential haploinsufficiency of uN2C in NIID is unlikely because NOTCH2NLC mRNA levels are unaltered in NIID (Sone et al., 2019; Ishiura et al., 2019; Tian et al., 2019). Furthermore, individuals with NIID lack the typical sensitivity to ionizing radiation, immunodeficiency, and microcephaly observed in individuals with radiosensitivity with severe combined immunodeficiency (RS-SCID), who are carriers of loss-of-function mutations in components of the NHEJ repair mechanism (Woodbine et al., 2014). Finally, we do not observe overt mislocalization of the Ku proteins or accumulation of DNA damage in cell and animal models of NIID, nor in skin and brain sections of adult individuals with NIID. Similarly, we found neither alterations of the Ku proteins nor increased DNA damage in cells or mice expressing the FXTAS polyG-containing protein FMRpolyG. Thus, we believe that potential alteration of DNA repair is most likely not the main cause of neuronal cell death in NIID. In contrast, uN2CpolyG and FMRpolyG possess a common polyG stretch that, when expressed in isolation, is sufficient to form p62-positive inclusions and causes neuronal cell death. Formation of cellular inclusions is consistent with the known *in vitro* self-aggregation properties of polyG homopolypeptides, which form amyloid-like fibrils (Lorusso et al., 2011; Plumley et al., 2011), but the molecular and cellular mechanisms by which these polyG proteins move into cell nuclei and drive neuronal cell dysfunctions remain unknown.

These results unveil a novel class of genetic disorders, polyG diseases, in which expanded GGC repeats embedded in upstream ORFs are translated into toxic polyG-containing proteins. Pharmaceutical compounds modulating autophagy could be of therapeutic interest to prevent toxic accumulation of these proteins.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Human samples
  - Mice
  - Cell cultures and models
- METHOD DETAILS
  - Constructs
  - Cell transfection and treatments
  - Mass spectrometry analysis
  - Antibody production
  - AAV production and retro-orbital injection
  - Animal phenotyping
  - X-ray and UVA-laser irradiation
  - Immunofluoresence and immunchemistry
  - Lysostaphin treatment
  - Co-immunoprecipitation assay
  - Western blotting
  - Quantitative real time RT-PCR
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.neuron.2021.03.038.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (81571219, 82071409, and U20A20356 to Z.W.); the Double Thousand Talents Program of Jiangxi Province (to D.H.); the Peking University Medicine Fund of Fostering Young Scholars' Scientific & Technological Innovation (to J.D.); the Japan Society for the Promotion of Science (KAKENHI JP19H03577) and the MHLW FC Program (JPMH19189624) (to J.S.); ERC-2012-StG 310659, ANR-18-CE16-0019, and FRM EQU202103012936 (to N.C.B.); and ANR-10-LABX-0030-INRT and ANR-10-IDEX-0002-02 (to I.G.B.M.C.).

AUTHOR CONTRIBUTIONS

Experiments were performed by M.B., J.D., J.S., V.P., B.M., F. Ruffenach, F. Riet, E.G., P.K., H.J., and M.O.-A. Control and NIID cases originated from K.M., W.A.C., A.A.D., D.H., H.M., Y.I., J.S., and Z.W. Data were collected and analyzed by M.B., J.D., J.S., Z.W., W.A.C., A.A.D., E.G., L.N., and N.C.-B. The study was designed, coordinated, and written by N.C.-B. with input from all authors.

DECLARATION OF INTEREST

The authors declare no competing interests.

Received: August 8, 2020
Revised: January 8, 2021
Accepted: March 30, 2021
Published: April 21, 2021

SUPPORTING CITATIONS

The following references appear in the Supplemental information: Cupidi et al. (2019), McFadden et al. (2005).

REFERENCES

Aspden, J.L., Eyre-Walker, Y.C., Phillips, R.J., Amin, U., Mumtaz, M.A., Brocard, M., and Couso, J.P. (2014). Extensive translation of small Open Reading Frames revealed by Poly-Ribo-Seq. Elife *3*, e03528.

Chen, H., Lu, L., Wang, B., Cui, G., Wang, X., Wang, Y., Raza, H.K., Min, Y., Li, K., Cui, Y., et al. (2020a). Re-defining the clinicopathological spectrum of neuronal intranuclear inclusion disease. Ann. Clin. Transl. Neurol. *7*, 1930–1941.

Chen, Z., Yan Yau, W., Jaunmuktane, Z., Tucci, A., Sivakumar, P., Gagliano Taliun, S.A., Turner, C., Efthymiou, S., Ibáñez, K., Sullivan, R., et al.; Genomics England Research Consortium (2020b). Hardy J, Ryten M, Vandrovcova J, Houlden H. Neuronal intranuclear inclusion disease is genetically heterogeneous. Ann. Clin. Transl. Neurol. *7*, 1716–1725.

Cortese, A., Simone, R., Sullivan, R., Vandrovcova, J., Tariq, H., Yau, W.Y., Humphrey, J., Jaunmuktane, Z., Sivakumar, P., Polke, J., et al. (2019). Biallelic expansion of an intronic repeat in RFC1 is a common cause of late-onset ataxia. Nat. Genet. *51*, 649–658.

Cupidi, C., Dijkstra, A.A., Melhem, S., Vernooij, M.W., Severijnen, L.A., Hukema, R.K., Rozemuller, A.J.M., Neumann, M., van Swieten, J.C., and Seelaar, H. (2019). Refining the Spectrum of Neuronal Intranuclear Inclusion Disease: A Case Report. J. Neuropathol. Exp. Neurol. *78*, 665–670.

Deng, J., Gu, M., Miao, Y., Yao, S., Zhu, M., Fang, P., Yu, X., Li, P., Su, Y., Huang, J., et al. (2019). Long-read sequencing identified repeat expansions in the 5'UTR of the *NOTCH2NLC* gene from Chinese patients with neuronal intranuclear inclusion disease. J. Med. Genet. *56*, 758–764.

Deng, J., Yu, J., Li, P., Luan, X., Cao, L., Zhao, J., Yu, M., Zhang, W., Lv, H., Xie, Z., et al. (2020). Expansion of GGC Repeat in GIPC1 Is Associated with Oculopharyngodistal Myopathy. Am. J. Hum. Genet. *106*, 793–804.

Fang, P., Yu, Y., Yao, S., Chen, S., Zhu, M., Chen, Y., Zou, K., Wang, L., Wang, H., Xin, L., et al. (2020). Repeat expansion scanning of the NOTCH2NLC gene in patients with multiple system atrophy. Ann. Clin. Transl. Neurol. *7*, 517–526.

Fell, V.L., and Schild-Poulter, C. (2015). The Ku heterodimer: function in DNA repair and beyond. Mutat. Res. Rev. Mutat. Res. *763*, 15–29.

Fiddes, I.T., Lodewijk, G.A., Mooring, M., Bosworth, C.M., Ewing, A.D., Mantalas, G.L., Novak, A.M., van den Bout, A., Bishara, A., Rosenkrantz, J.L., et al. (2018). Human-Specific NOTCH2NL Genes Affect Notch Signaling and Cortical Neurogenesis. Cell *173*, 1356–1369.e22.

Gao, F.B., and Richter, J.D. (2017). Microsatellite Expansion Diseases: Repeat Toxicity Found in Translation. Neuron *93*, 249–251.

Gelpi, E., Botta-Orfila, T., Bodi, L., Marti, S., Kovacs, G., Grau-Rivera, O., Lozano, M., Sánchez-Valle, R., Muñoz, E., Valldeoriola, F., et al. (2017). Neuronal intranuclear (hyaline) inclusion disease and fragile X-associated tremor/ataxia syndrome: a morphological and molecular dilemma. Brain *140*, e51.

Grundy, G.J., Moulding, H.A., Caldecott, K.W., and Rulten, S.L. (2014). One ring to bring them all–the role of Ku in mammalian non-homologous end joining. DNA Repair (Amst.) *17*, 30–38.

Hagerman, R.J., Leehey, M., Heinrichs, W., Tassone, F., Wilson, R., Hills, J., Grigsby, J., Gage, B., and Hagerman, P.J. (2001). Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. Neurology *57*, 127–130.

Hukema, R.K., Buijsen, R.A., Schonewille, M., Raske, C., Severijnen, L.A., Nieuwenhuizen-Bakker, I., Verhagen, R.F., van Dessel, L., Maas, A., Charlet-Berguerand, N., et al. (2015). Reversibility of neuropathology and motor deficits in an inducible mouse model for FXTAS. Hum. Mol. Genet. *24*, 4948–4957.

Ishiura, H., and Tsuji, S. (2020). Advances in repeat expansion diseases and a new concept of repeat motif-phenotype correlation. Curr. Opin. Genet. Dev. *65*, 176–185.

Ishiura, H., Shibata, S., Yoshimura, J., Suzuki, Y., Qu, W., Doi, K., Almansour, M.A., Kikuchi, J.K., Taira, M., Mitsui, J., et al. (2019). Noncoding CGG repeat expansions in neuronal intranuclear inclusion disease, oculopharyngodistal myopathy and an overlapping disease. Nat. Genet. *51*, 1222–1232.

Li, M., Li, K., Li, X., Tian, Y., Shen, L., Wu, G., Zhang, Z., and Chen, W. (2020). Multiple reversible encephalitic attacks: a rare manifestation of neuronal intranuclear inclusion disease. BMC Neurol. *20*, 125.

Lim, S.Y., Ishiura, H., Ramli, N., Shibata, S., Almansour, M.A., Tan, A.H., Houlden, H., Lang, A.E., and Tsuji, S. (2020). Adult-onset neuronal intranuclear inclusion disease mimicking Fragile X-associated tremor-ataxia syndrome in ethnic Chinese patients. Parkinsonism Relat. Disord. *74*, 25–27.

Lindenberg, R., Rubinstein, L.J., Herman, M.M., and Haydon, G.B. (1968). A light and electron microscopy study of an unusual widespread nuclear inclusion body disease. A possible residuum of an old herpesvirus infection. Acta Neuropathol. *10*, 54–73.

Liu, Y., Mimuro, M., Yoshida, M., Hashizume, Y., Niwa, H., Miyao, S., Ujihira, N., and Akatsu, H. (2008). Inclusion-positive cell types in adult-onset intranuclear inclusion body disease: implications for clinical diagnosis. Acta Neuropathol. *116*, 615–623.

Lorusso, M., Pepe, A., Ibris, N., and Bochicchio, B. (2011). Molecular and supramolecular studies on polyglycine and poly-l-proline. Soft Matter *7*, 6327.

Ma, D., Tan, Y.J., Ng, A.S.L., Ong, H.L., Sim, W., Lim, W.K., Teo, J.X., Ng, E.Y.L., Lim, E.C., Lim, E.W., et al. (2020). Association of NOTCH2NLC Repeat Expansions With Parkinson Disease. JAMA Neurol. *77*, 1–5.

McFadden, K., Hamilton, R.L., Insalaco, S.J., Lavine, L., Al-Mateen, M., Wang, G., and Wiley, C.A. (2005). Neuronal intranuclear inclusion disease without polyglutamine inclusions in a child. J. Neuropathol. Exp. Neurol. *64*, 545–552.

Mori, F., Tanji, K., Odagiri, S., Hattori, M., Hoshikawa, Y., Kono, C., Yasui, K., Yokoi, S., Hasegawa, Y., Kamitani, T., et al. (2012). Ubiquitin-related proteins in neuronal and glial intranuclear inclusions in intranuclear inclusion body disease. Pathol. Int. *62*, 407–411.

Munoz-Garcia, D., and Ludwin, S.K. (1986). Adult-onset neuronal intranuclear hyaline inclusion disease. Neurology *36*, 785–790.

Nakamura, M., Murray, M.E., Lin, W.L., Kusaka, H., and Dickson, D.W. (2014). Optineurin immunoreactivity in neuronal and glial intranuclear inclusions in adult-onset neuronal intranuclear inclusion disease. Am. J. Neurodegener. Dis. *3*, 93–102.

Nelson, D.L., Orr, H.T., and Warren, S.T. (2013). The unstable repeats–three evolving faces of neurological disease. Neuron *77*, 825–843.

Ogasawara, M., Iida, A., Kumutpongpanich, T., Ozaki, A., Oya, Y., Konishi, H., Nakamura, A., Abe, R., Takai, H., Hanajima, R., et al. (2020). CGG expansion in NOTCH2NLC is associated with oculopharyngodistal myopathy with neurological manifestations. Acta Neuropathol. Commun. *8*, 204.

Plumley, J.A., Tsai, M.I., and Dannenberg, J.J. (2011). Aggregation of capped hexaglycine strands into hydrogen-bonding motifs representative of pleated and rippled β-sheets, collagen, and polyglycine I and II crystal structures. A density functional theory study. J. Phys. Chem. B *115*, 1562–1570.

Pountney, D.L., Huang, Y., Burns, R.J., Haan, E., Thompson, P.D., Blumbergs, P.C., and Gai, W.P. (2003). SUMO-1 marks the nuclear inclusions in familial neuronal intranuclear inclusion disease. Exp. Neurol. *184*, 436–446.

Rafehi, H., Szmulewicz, D.J., Bennett, M.F., Sobreira, N.L.M., Pope, K., Smith, K.R., Gillies, G., Diakumis, P., Dolzhenko, E., Eberle, M.A., et al. (2019). Bioinformatics-Based Identification of Expanded Repeats: A Non-reference Intronic Pentamer Expansion in RFC1 Causes CANVAS. Am. J. Hum. Genet. *105*, 151–165.

Rodriguez, C.M., and Todd, P.K. (2019). New pathologic mechanisms in nucleotide repeat expansion disorders. Neurobiol. Dis. *130*, 104515.

Sellier, C., Buijsen, R.A.M., He, F., Natla, S., Jung, L., Tropel, P., Gaucherot, A., Jacobs, H., Meziane, H., Vincent, A., et al. (2017). Translation of Expanded CGG Repeats into FMRpolyG Is Pathogenic and May Contribute to Fragile X Tremor Ataxia Syndrome. Neuron *93*, 331–347.

Sone, J., Hishikawa, N., Koike, H., Hattori, N., Hirayama, M., Nagamatsu, M., Yamamoto, M., Tanaka, F., Yoshida, M., Hashizume, Y., et al. (2005). Neuronal intranuclear hyaline inclusion disease showing motor-sensory and autonomic neuropathy. Neurology *65*, 1538–1543.

Sone, J., Tanaka, F., Koike, H., Inukai, A., Katsuno, M., Yoshida, M., Watanabe, H., and Sobue, G. (2011). Skin biopsy is useful for the antemortem

diagnosis of neuronal intranuclear inclusion disease. Neurology *76*, 1372–1376.

Sone, J., Kitagawa, N., Sugawara, E., Iguchi, M., Nakamura, R., Koike, H., Iwasaki, Y., Yoshida, M., Takahashi, T., Chiba, S., et al. (2014). Neuronal intranuclear inclusion disease cases with leukoencephalopathy diagnosed via skin biopsy. J. Neurol. Neurosurg. Psychiatry *85*, 354–356.

Sone, J., Mori, K., Inagaki, T., Katsumata, R., Takagi, S., Yokoi, S., Araki, K., Kato, T., Nakamura, T., Koike, H., et al. (2016). Clinicopathological features of adult-onset neuronal intranuclear inclusion disease. Brain *139*, 3170–3186.

Sone, J., Mitsuhashi, S., Fujita, A., Mizuguchi, T., Hamanaka, K., Mori, K., Koike, H., Hashiguchi, A., Takashima, H., Sugiyama, H., et al. (2019). Long-read sequencing identifies GGC repeat expansions in NOTCH2NLC associated with neuronal intranuclear inclusion disease. Nat. Genet. *51*, 1215–1221.

Sun, Q.Y., Xu, Q., Tian, Y., Hu, Z.M., Qin, L.X., Yang, J.X., Huang, W., Xue, J., Li, J.C., Zeng, S., et al. (2020). Expansion of GGC repeat in the human-specific NOTCH2NLC gene is associated with essential tremor. Brain *143*, 222–233.

Suzuki, I.K., Gacquer, D., Van Heurck, R., Kumar, D., Wojno, M., Bilheu, A., Herpoel, A., Lambert, N., Cheron, J., Polleux, F., et al. (2018). Human-Specific NOTCH2NL Genes Expand Cortical Neurogenesis through Delta/Notch Regulation. Cell *173*, 1370–1384.e16.

Takahashi-Fujigasaki, J. (2003). Neuronal intranuclear hyaline inclusion disease. Neuropathology *23*, 351–359.

Takahashi-Fujigasaki, J., Nakano, Y., Uchino, A., and Murayama, S. (2016). Adult-onset neuronal intranuclear hyaline inclusion disease is not rare in older adults. Geriatr. Gerontol. Int. *16* (*Suppl 1*), 51–56.

Tian, Y., Wang, J.L., Huang, W., Zeng, S., Jiao, B., Liu, Z., Chen, Z., Li, Y., Wang, Y., Min, H.X., et al. (2019). Expansion of Human-Specific GGC Repeat in Neuronal Intranuclear Inclusion Disease-Related Disorders. Am. J. Hum. Genet. *105*, 166–176.

Todd, P.K., Oh, S.Y., Krans, A., He, F., Sellier, C., Frazer, M., Renoux, A.J., Chen, K.C., Scaglione, K.M., Basrur, V., et al. (2013). CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. Neuron *78*, 440–455.

Woodbine, L., Gennery, A.R., and Jeggo, P.A. (2014). The clinical impact of deficiency in DNA non-homologous end-joining. DNA Repair (Amst.) *16*, 84–96.

Xi, J., Wang, X., Yue, D., Dou, T., Wu, Q., Lu, J., Liu, Y., Yu, W., Qiao, K., Lin, J., et al. (2021). 5′ UTR CGG repeat expansion in GIPC1 is associated with oculopharyngodistal myopathy. Brain *144*, 601–614.

Yu, J., Deng, J., Guo, X., Shan, J., Luan, X., Cao, L., Zhao, J., Yu, M., Zhang, W., Lv, H., et al. (2021). The GGC repeat expansion in NOTCH2NLC is associated with oculopharyngodistal myopathy type 3. Brain. Published online March 9, 2021. https://doi.org/10.1093/brain/awab077.

Yuan, Y., Liu, Z., Hou, X., Li, W., Ni, J., Huang, L., Hu, Y., Liu, P., Hou, X., Xue, J., et al. (2020). Identification of GGC repeat expansion in the *NOTCH2NLC* gene in amyotrophic lateral sclerosis. Neurology *95*, e3394–e3405.

Zu, T., Gibbens, B., Doty, N.S., Gomes-Pereira, M., Huguet, A., Stone, M.D., Margolis, J., Peterson, M., Markowski, T.W., Ingram, M.A., et al. (2011). Non-ATG-initiated translation directed by microsatellite expansions. Proc. Natl. Acad. Sci. USA *108*, 260–265.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-uN2CpolyG (SVEMAMNPV), Mouse Monoclonal Clone 4C4	This paper	N/A
Anti-uN2CpolyG (AARPPRMH), Mouse Monoclonal Clone 4D12	This paper	N/A
Anti-FLAG (DYKDDDDK) tag, Rabbit Polyclonal,	Thermo Fisher Scientific	Cat# PA1-984B; RRID: AB_347227
Anti-HA (YPYDVPDYA) tag, Mouse Monoclonal Clone 16B12	Abcam	Cat# ab130275; RRID: AB_11156884
Anti-GFP, Rabbit Polyclonal,	Abcam	Cat# ab290; RRID: AB_303395
Anti-GAPDH, Mouse Monoclonal Clone GA1R	Abcam	Cat# ab125247; RRID: AB_11129118
Anti-Calbindin, Mouse Monoclonal Clone CB-955	Abcam	Cat# ab82812; RRID: AB_1658451
Anti-Ubiquitin, Rabbit Monoclonal Recombinant EPR8830	Abcam	Cat# ab134953; RRID: AB_2801561
Anti-Sumo 2 and 3, Rabbit Polyclonal	Abcam	Cat# ab3742; RRID: AB_304041
Anti-p62 (*SQSTM1*), Rabbit Monoclonal Recombinant EPR4844	Abcam	Cat# ab109012; RRID: AB_2810880
Anti-MAP2, Mouse Monoclonal Clone MT-01	Abcam	Cat# ab7756; RRID: AB_306050
Anti-Histone H2A.X phospho S139, Rabbit Monoclonal Recombinant EP854(2)Y	Abcam	Cat# ab81299; RRID: AB_1640564
Anti-Histone H2A.X, Rabbit Monoclonal Recombinant EPR22820-23	Abcam	Cat# ab229914
Anti-Ku80 (*XRCC5*), Mouse Monoclonal Clone 5C5	Abcam	Cat# ab119935; RRID: AB_10899161
Anti-Ku-70 (*XRCC6*), Mouse Monoclonal Clone N3H10	Santa Cruz Biotechnology	Cat# sc-56129; RRID: AB_794205
Anti-GFAP, Rabbit Polyclonal,	Proteintech	Cat# 16825-1-AP; RRID: AB_2109646
Bacterial and virus strains		
AAV2/PHP.eB CAG GFP	This paper	N/A
AAV2/PHP.eB CAG uN2C-GFP	This paper	N/A
AAV2/PHP.eB CAG uN2CpolyG-GFP	This paper	N/A
One Shot Stbl3 Chemically Competent *E. coli*	Invitrogen	Cat# C737303
One Shot TOP10 Chemically Competent *E. coli*	Invitrogen	Cat# C404003
Biological samples		
Human skin and brain tissues	Described in Table S1	N/A
Chemicals, peptides, and recombinant proteins		
Lysostaphin Enzyme Protein Recombinant	ProSpec	Cat# ENZ-269
Histosol Plus	LifeScience Products	Cat# HS-100-5GL
TO-PRO-3 Iodide (642/661) - 1 mM Solution in DMSO	Fisher Scientific	Cat# T-3605
Lipofectamine 2000 Transfection reagent	Fisher Scientific	Cat# 11668019
Complete Mini EDTA-free Protease Inhibitor Cocktail	Roche	Cat# 11836170001
Bafilomycin A1	Sigma	Cat# B1793
MG132	Sigma	Cat# M8699
Doxycycline hydrochloride	Sigma	Cat# D3447
Critical commercial assays		
Anti-GFP Magnetic beads	Abcam	Cat# ab193983
Anti-HA Magnetic Beads	Pierce	Cat# 88837
EC Prime Western Blotting Detection Reagent	Sigma	Cat# GERPN2236
DAB Substrate Kit	Abcam	Cat# ab64238
TRI Reagent (Guanidine Thiocyanate & Phenol)	Merck	Cat# T9424-25ML
Transcriptor High Fidelity cDNA Synthesis Kit	Sigma	Cat# 5081955001
LightCycler 480 SYBR Green I Master	Roche	Cat# 04707516001

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models: Cell lines		
Human: U2OS cells	ATCC	Cat# 300364/p489_U-2_OS; RRID: CVCL_0042
Human: HEK293 cells	ATCC	Cat# 300192/p777_HEK293; RRID: CVCL_0045
Mouse: GT1-7 cells	ATCC	Cat# SCC116; RRID: CVCL_0281
Human: U2OS T-REx TO-uN2C-GFP	This paper	N/A
Human: U2OS T-REx TO-uN2CpolyG-GFP	This paper	N/A
Experimental models: Organisms/strains		
Wild type Mouse: C57BL/6J *Mus musculus*	The Jackson, Laboratory	JAX:000664, RRID: IMSR_JAX:000664
Oligonucleotides		
Primer RT-qPCR GFP Forward ACGTAAACGGCCACAAGTTC	This paper	N/A
Primer RT-qPCR GFP Reverse AAGTCGTGCTGCTTCATGTG	This paper	N/A
Primer RT-qPCR *RPLP0* Forward GAAGTCACTGTGCCAGCCCA	This paper	N/A
Primer RT-qPCR *RPLP0* Reverse GAAGGTGTAATCCGTCTCCA	This paper	N/A
Recombinant DNA		
Plasmid: pcDNA3-TetOn uN2C-GFP (12 GLY)	This paper	N/A
Plasmid: pcDNA3-TetOn uN2CpolyG-GFP (100 GLY)	This paper	N/A
Plasmid: pAAV2-CAG uN2C-GFP (12 GLY)	This paper	N/A
Plasmid: pAAV2-CAG uN2CpolyG-GFP (100 GLY)	This paper	N/A
Plasmid: pcDNA3 *NOTCH2NLC* Exon1-GFP Gly frame	This paper	N/A
Plasmid: pcDNA3 *NOTCH2NLC* Exon1-GFP Ala frame	This paper	N/A
Plasmid: pcDNA3 *NOTCH2NLC* Exon1-GFP Arg frame	This paper	N/A
Plasmid: pcDNA3 ATG GLY 70x-GFP	This paper	N/A
Plasmid: pcDNA3 ATG ALA 100x-GFP	This paper	N/A
Plasmid: pcDNA3 ATG ARG 100x-GFP	This paper	N/A
Software and algorithms		
Fiji (ImageJ)	NIH	RRID: SCR_002285
Prism	GraphPad	RRID: SCR_005375

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Nicolas Charlet-Berguerand (ncharlet@igbmc.fr).

Materials availability

All unique reagents (e.g., uN2CpolyG constructs, 4C4 and 4D12 antibodies, etc.) generated in this study are available from the Lead Contact under MTA subject to restrictions from commercial source.

Data and code availability

Proteomics datasets related to Figure S4 in the paper are available in Table S2. Complete proteomics source data are available from the corresponding author upon reasonable request. No further unique datasets or codes were generated in this study.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human samples

Patient available information is described in Table S1. Human skin and brain samples were sampled with the informed consent of individuals and families and approved by the Institutional Review Board of the Peking University First Hospital and Suzuka National Hospital.

Mice

All animal work was performed with approval from the IGBMC/ICS Animal Care Committee and of the French agency for research on animal (DGRI) authorization number APAFIS#11459-2017092208308760. C57BL/6 wild-type male mice were retro-orbitally AAV-injected at 2 months and then housed for 6 to 8 months in a temperature-controlled room (19–22°C) with a 12:12-hours light/dark cycle and free access to food and water. Mice were sacrificed by cervical dislocation in order to dissect the brain and spinal cord which were subsequently frozen for molecular biology or PFA-fixed and embedded in paraffin for histology.

Cell cultures and models

Mouse primary cortical neurons were prepared from E18 C57BL/6 wild-type mice embryos. Cortical regions were dissected and digested in 1X HBSS (ThermoFisher) with 0.25% trypsin (ThermoFisher) at 37°C for 10 minutes. Trypsin digestion was stopped by addition of DMEM (ThermoFisher), 10% horse serum (Life Technologies) and 1x GlutaMAX (ThermoFisher). Then, cortical neurons were dissociated, centrifuged, resuspended and plated on 0.1 mg/ml poly L-lysine (Sigma) pre-coated 24-well plates in Neurobasal Medium (ThermoFisher) supplemented with 1x B27, 0.5 mM L-glutamine and 100 IU/ml penicillin/streptomycin (ThermoFisher) at 37°C with 5% CO2. Neuronal GT1-7 cells were grown in DMEM 4.5 g/l glucose with 10% fetal calf serum and 100 IU/ml penicillin/streptomycin at 37°C in 5% CO2. U2OS and HEK293 cells were grown in DMEM 1 g/l glucose with 10% FCS and gentamycin at 37°C in 5% CO2. U2OS T-REx cells (ThermoFisher) stably expressing uN2C-GFP or uN2CpolyG-GFP were Lipofectamine-transfected with Pci1-linearized pcDNA3-TetOn expressing uN2C with 12 or 100 optimized GGC fused to the GFP and sectioned for neomycin resistance for two weeks.

METHOD DETAILS

Constructs

Human NOTCH2NLC exon 1 sequence containing 166 nts upstream of the GGC repeats was cloned into pcDNA3.1 fused to a the GFP deleted of its ATG and in all three frames. Mutations of the repeats size, ATG start codon or within the uN2C ORF were achieved by inverse PCR or by oligonucleotide ligations. NOTCH2NLC upstream ORF fused to the GFP with either 12 or 100 optimized GGN repeats was cloned into the pcDNA3-TetOn and pAAV2-CAG vector. To insure stability of repeat expansions, all GGC repeat-containing plasmids were transformed into STBL3 bacterial strain (Invitrogen). All constructs were confirmed by sanger sequencing.

Cell transfection and treatments

For AAV transduction of mouse primary cortical neurons, 4 to 5 days differentiated mouse primary cortical neurons were incubated with 5x10EXP10 vg/ml of AAV2/PHP.eB expressing the GFP, uN2C-GFP or uN2CpolyG-GFP, half of the media was changed every 48 hours and cells were let differentiated for 4 to 6 more days. For transient transfection, cells were plated in DMEM and 0.1% fetal bovine serum and transfected for 24 hours using Lipofectamine 2000 (Fisher Scientific). After 1 to 6 days post transient transfection, induction with 1 μg/mL doxycycline (Sigma) or AAV transduction, cells were analyzed by immunofluorescence or western blotting. For cell viability, mouse cortical neurons primary cultures were incubated 10 min with 1 μg/ml of propidium iodide and analyzed by microscopy. For GT1-7 cell viability, cells were detached by trypsin and resuspended in PBS with 20 nM TO-PRO-3 iodide (Fisher scientific) and FACS analyzed. Cells were treated with 100 nM of Bafilomycin A1, 1 μM MG132 or the indicated concentration of drug (Sigma) during 15 hours before analysis.

Mass spectrometry analysis

HEK293 cells were transfected with uN2C-GFP plasmid using Lipofectamine 2000 (Fisher Scientific) for 24 hours. Proteins were purified by GFP-immunoprecipitation, trypsin digested and the peptides were extracted twice with acetonitrile/water/formic acid-45/45/10-v/v/v followed by a final extraction with acetonitrile /formic acid (FA)-95/05-v/v. Extracted peptides were then analyzed using an Ultimate 3000 nano-RSLC (Thermo Scientific) coupled in line with an Orbitrap ELITE (Thermo Scientific). Peptides were separated on a C18 nano-column with a linear gradient of acetonitrile and analyzed with in a Top 20 collision-induced dissociation data-dependent mass spectrometry with an inclusion list. Data were processed by database searching using SequestHT (Thermo Fisher Scientific) with Proteome Discoverer 1.4 software (Thermo Fisher Scientific) or against a homemade database of all potential three frames translated proteins or peptides from the human NOTCH2NLC 5'UTR sequences. Precursor and fragment mass tolerance were set at 7 ppm and 0.5 Da respectively. Oxidation (M) and Nterminal Acetylation were set as variable modification, and Carbamidomethylation (C) as fixed modification. Peptides were filtered with the Fixed value node of Proteome Discoverer 1.4.

Antibody production

To generate monoclonal antibodies directed against uN2polyG, two months old female BALB/c mice were injected intraperitoneally with KLH conjugated peptides (4D12: CAARPPRMH, 4C4: CSVEMAMNPV) with 200 ug of poly(I/C) as adjuvant. Three injections were performed at 2 weeks intervals and four days prior to hybridoma fusion, mice with positively reacting sera were re-injected. Spleen cells were fused with Sp2/0.Agl4 myeloma cells. Supernatants of hybridoma cultures were tested at day 10 by ELISA for cross-reaction with peptides. Positive supernatants were then tested by Immunofluorescence and western blot on transfected HEK293 cells. Specific cultures were cloned twice on soft agar. Specific hybridomas were established and ascites fluid was prepared by injection of

2x106 hybridoma cells into Freund adjuvant-primed BALB/c mice. All animal experimental procedures were performed according to the French and European authority guidelines.

AAV production and retro-orbital injection

Recombinant AAV2/PHP.eB were generated by tri-transfection of HEK293 cells with pAAV2-GFP, -uN2C-GFP or -uN2CpolyG-GFP with pUCmini-iCAP-PHP.eB and pAD-DELTA-F16. Recombinant vectors were purified by double cesium chloride ultracentrifugation gradients from cell lysates, followed by dialysis and concentration against sterile PBS. Particles were quantified by real time PCR and vector titers were expressed as viral genomes per ml (vg/ml). 2 months old C57BL/6 male mice were injected retro-orbitally with 100 μL of sterile PBS with 1.5x10EXP13 vg/kg of AAV2/PHP.eB.

Animal phenotyping

Rotarod test (Bioseb, Chaville, France) was performed with three testing trials during which the rotation speed accelerated from 4 to 40 rpm in 5 min. Trials were separated by 10-15 min interval. The average latency was used as index of motor coordination performance. Grip test: this test measures the maximal muscle strength (g) using an isometric dynamometer connected to a grid (Bioseb). Mice were allowed to grip the grid with all its paws then they were pulled backward until they released it. Each mouse was submitted to 3 consecutive trials. The maximal strength developed by the mouse before releasing the grid was recorded and the average value of the three trials adjusted to body weight. Notched bar test: mice were tested under 100-lux lighting on a 2 cm-wide and 50 cm-long natural wooden piece notched bar comprising 12 platforms of 2 cm spaced by 13 gaps of 2 cm and bearing a 6 cm^2 terminal platform. Animals had to cross the notched bar twice for training and 3 times for the test. Every instance of a back paw going through the gap was considered an error, and the global error percentage was calculated. Open field test: mice were tested in automated open fields (Panlab, Barcelona, Spain), each virtually divided into central and peripheral regions. The open fields were placed in a room homogeneously illuminated at 120 Lux. Each mouse was placed in the periphery of the open field and allowed to explore freely the apparatus for 30 min, with the experimenter out of the animal's sight. The distance traveled, the number of rears, and time spent in the central and peripheral regions were recorded over the test session. The number of entries and the percent time spent in center area are used as index of emotionality/anxiety

X-ray and UVA-laser irradiation

U2OS cells expressing uN2C with 12 or 100 glycine GFP-tagged were X-ray irradiated for 10 minutes at 10 Gy in a CellRad (Precision). UV-laser micro-irradiation of U2OS cells expressing GFP-tagged U2C with 12 or 100 glycine seeded onto glass-bottom dishes was performed with a spinning disk Yokogawa X1 equipped with a Nikon microscope. For scanning irradiation, we use UV laser 266 nm pulse width < 1 ns power 15 mW and one repetition for each irradiation area. For GFP signal, we use time-lapse spinning disk acquisition with 491 nm 100 mW diode laser, the complete system is customized by Gataga System (Massy-France) and controlled with Metamorph Software.

Immunofluoresence and immunchemistry

For immunofluorescence, mouse or human brain sections were deparaffinized for 10 min in Histosol Plus (LifeScienceProducts) and dehydrated as follows: ethanol 100% (10 min), ethanol 90% (5 min), ethanol 70% (5 min), and rinsed in water. Antigen retrieval was performed in pressure cooker in 10 mM Tris PH9, 1 mM EDTA followed by blocking 1 h with PBS, 0,5% Triton X-100 and 5% Horse Serum. Glass coverslips containing plated cells or brain sections treated as described above were fixed for 15 min in PBS with 4% paraformaldehyde, washed with PBS and incubated in PBS plus 0.5% Triton X-100 during 10 min. The cells were washed three times with PBS and the coverslips were incubated during one hour with primary antibody against Calbindin (Abcam ab82812, 1/200), γH2AX (Abcam ab81299, 1/300), ubiquitin (Abcam ab134953, 1/100), Sumo2/3 (Abcam ab3742, 1/100), p62/Sqstm1 (Abcam ab109012, 1/1000), GFAP (Proteintech 16825-1-AP, 1/250), MAP2 (Abcam ab7756, 1/200) and uN2C 4C4 or 4D12 (mouse monoclonal homemade, 1/100). After washing with PBS, the coverslips were incubated with goat anti mouse secondary antibody conjugated with Alexa 488 or CY3 (Interchim SA) for one hour, washed twice with PBS and incubated for 3 min in PBS/DAPI (1/10 000 dilution). Coverslips were rinsed twice before mounting in Pro-Long media (Molecular Probes) and were examined using a Leica microscope. For immunochemistry, mouse brain sections were deparaffinized 10 min in Histosol Plus (LifeScienceProducts) and dehydrated as follows: ethanol 100% (10 min), ethanol 90% (5 min), ethanol 70% (5 min), and rinsed in water. Antigen retrieval was performed in pressure cooker in 10 mM Tris PH9, 1 mM EDTA followed with 10 μg/ml protein kinase treatment for 20 min at 37°c. Endogenous peroxidase activity was blocked, and blocking 1 h with PBS, 0,5% Triton X-100 and 5% Horse Serum and immunostaining was performed overnight at 4°C using antibody against Calbindin (Abcam ab82812, 1/50), Gfap (Proteintech, 16825-1-AP, 1/100), Sumo2/3 (Abcam ab3742, 1/100) and uN2C 4C4 or 4D12 (mouse monoclonal homemade, 1/100). Antigen–antibody complexes were visualized by incubation with DAB substrate (Abcam) and slides were counterstained with hematoxylin and eosin.

Lysostaphin treatment

HEK293 cells transfected with a uN2CpolyGFP construct were scrapped in PBS 1X and centrifuged during 10 min at 3000 rpm at 4°C. The cell pellet was resuspended in 400 μl of RIPA and 16 μl of cell extract was incubated with 1 μg of lysostaphin (Prospec, ENZ-269) during 5 to 15 minutes at 37°C. Laemmli buffer was add to the mix and proteins were analyze by western blot.

Co-immunoprecipitation assay

24 h after transfection of HEK293 cells with 1 μg of uN2C-GFP constructs in Lipofectamine 2000 (Invitrogen), cells were lysed in RIPA buffer (50 mM Tris-HCl pH 7.5, 0.15 M NaCl, 0.5% Triton X-100) supplemented with protease inhibitor cocktail (Roche) and clarified by centrifugation at 14000 rpm for 10 min. Immunoprecipitations were performed at 4°C for 1 h using pre-washed Anti-GFP (Abcam ab193983) or anti-HA (Pierce 88837) Magnetic Beads in RIPA buffer, washed three time, then bound proteins were eluted by 3 min denaturation step at 95°C with Laemmli buffer followed by mass spectrometry or western blot.

Western blotting

Proteins were denatured 3 min at 95°C, separated on 4%–12% bis-Tris Gel (NuPAGE), transferred on nitrocellulose membranes (Whatman Protan), blocked with 5% non-fat dry milk in Tris Buffer Saline buffer (TBS), incubated with anti-GFP (Abcam ab290), HA (Abcam ab130275), FLAG (Thermofisher PA1-984B), γH2AX (Abcam ab81299), H2AX (Abcam ab229914), Ku70 (SantaCruz sc-56129), Ku80 (Abcam ab 119935), GAPDH (Abcam ab125247) in TBS plus 5% non-fat dry milk, washed 3 times and incubated with anti-rabbit or mouse Peroxidase antibody (1:10,000, Cell Signaling) 1 hour in TBS, followed by washing and ECL Prime chemiluminescence revelation kit (Sigma).

Quantitative real time RT-PCR

Total RNAs from mouse tissues or cells were isolated by TriReagent (Merck). cDNAs were generated using the Transcriptor High Fidelity cDNA synthesis kit (Sigma) for quantification of mRNAs. qPCR were realized using the LightCycler 480 SYBR Green I Master (Roche) in a Lightcycler 480 (Roche) with 15 min at 94°C followed by 50 cycles of 15 s at 94°C, 20 s at 58°C and 20 s at 72°C. *RPLPO* mRNA was used as standard and data were analyzed using the Lightcycler 480 analysis software (2ΔCt method).

QUANTIFICATION AND STATISTICAL ANALYSIS

To eliminate bias, image or animal analyzes were either completely automated or blinded. All statistical analyses were performed using Excel (Microsoft) and Prism (GraphPad). Experiments are represented as either mean value ± Standard Error of Mean (SEM) or box-and-whisker plots with box upper and lower limits representing the 25th and 75th quartiles, respectively, the whiskers depicting the lowest and highest data points and the horizontal line through the box represent the median. The statistical tests used are two-tailed paired Student's t test or ANOVA. Significance was set as $*p < 0.05$; $**p < 0.005$ and $***p < 0.001$. Sample-sizes were determined based on past experiments and to minimize the number of mice used. No statistical method was used to determine whether data meet assumptions of the statistical approach. Detailed statistical information, including the statistical test, measures, number "n" of animals, cells and/ or experiments are indicated in the figures and their respective legends.

Neuron, Volume *109*

Supplemental information

Translation of GGC repeat expansions into a toxic polyglycine protein in NIID defines a novel class of human genetic disorders: The polyG diseases

Manon Boivin, Jianwen Deng, Véronique Pfister, Erwan Grandgirard, Mustapha Oulad-Abdelghani, Bastien Morlet, Frank Ruffenach, Luc Negroni, Pascale Koebel, Hugues Jacob, Fabrice Riet, Anke A. Dijkstra, Kathryn McFadden, Wiley A. Clayton, Daojun Hong, Hiroaki Miyahara, Yasushi Iwasaki, Jun Sone, Zhaoxia Wang, and Nicolas Charlet-Berguerand

Supplemental Information

Translation of GGC repeat expansions into a toxic polyglycine protein in NIID defines a novel class of human genetic disorders: the polyG diseases.

1 - Figure S1: *NOTCH2NLC* GGC repeats are translated into polyglycine.

2 - Figure S2: uN2polyG is present in NIID intranuclear inclusions.

3 - Figure S3: Expression of uN2CpolyG is toxic in cell culture.

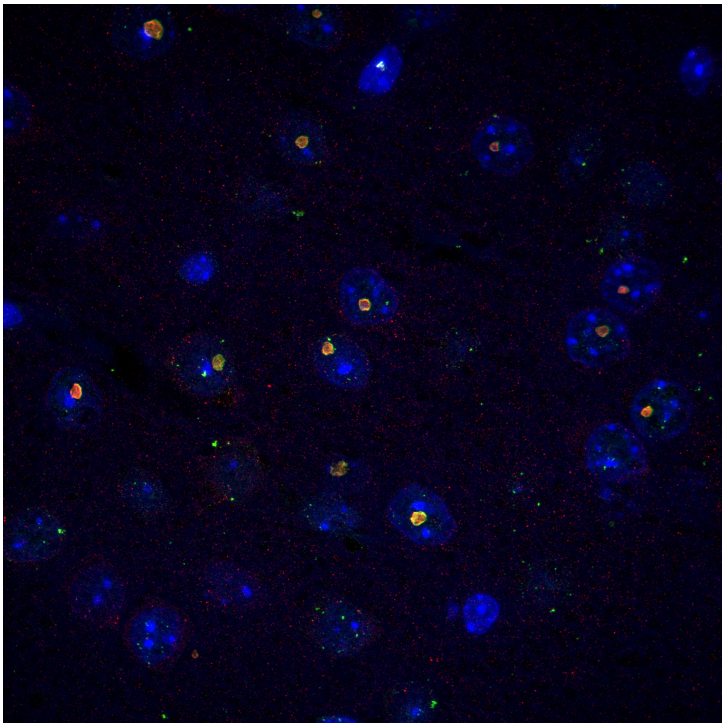
4 - Figure S4: The uN2C control protein promotes DNA repair.

5 - Figure S5: Expression of uN2CpolyG is pathogenic in animal.

6 - Figure S6: Model of polyG diseases.

7 - Table S1: Patient information.

8 - Table S2: Proteins interacting with uN2CpolyG.



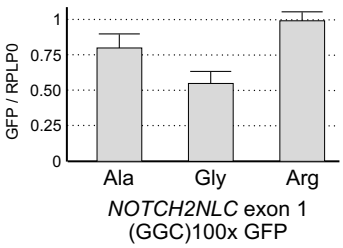
P62 (green) and uN2CpolyG (red)
staining in cortical area of NIID mouse model

A

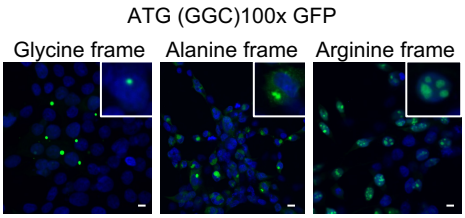
NOTCH2NLC (N2C) exon1 (GGC)100x *BamHI* GFP (+1 frame, glycine frame):

GCTGAGGCGGCGGCCGAGGAGCGGCGGACTCGGGGCGCGGGGAGTCGAGGCATTTGCGCCTGTGCTTCGGACCGTAGCGCCAGGGCCTGAGCCTTTGAAGCAGGAGGAG
GGGAGGAGAGAGTGGGGCTCCTCTATCGGGACCCCCTCCCC**ATG**TGGATCTGCCCA**GGCGGCGGC...GGCGGCGGC**GACCGAGAAGATGCCCGCCCTGCGCCGCTCTGCT
GTGGGCGCTGCTGGCGCTCTGGCTGTGCTGCGCGACCCCCGCGCATGggatccGTCAGCAAAGGCGAGGAGCTTTTCACCGGGGTTGTTCCCATCCTCGTCGAGCTCGACGGC
GACGTAAACGGCCACAAATTCAGCGTTTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTCACCCTCAAGTTCATCTGCACCACCGGCAAGCTCCCCGTCCCCTGGC
CCACCCTCGTCACCACCCTCACCTACGGCGTCCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGA
GCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAG
GACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAA
CATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCT
GAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAG

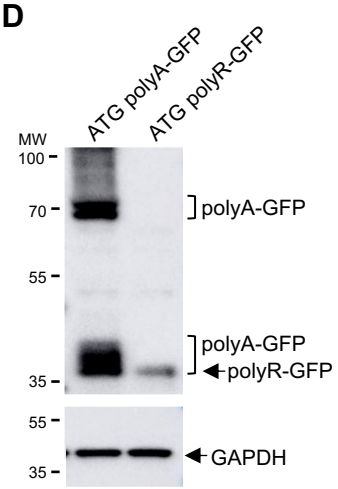
B



C



D



E

Ion Series | Neutral Losses | Precursor Ions

#1	a+	a2+	b+	b2+	Seq.	y+	y2+	#2
1	146.06341	73.53534	174.05833	87.53280	M-Acetyl			20
2	332.14272	166.57500	360.13764	180.57246	W	1587.67171	794.33949	19
3	445.22679	223.11703	473.22170	237.11449	I	1401.59240	701.29984	18
4	605.25744	303.13236	633.25235	317.12981	C-Carbami...	1288.50834	644.75781	17
5	702.31020	351.65874	730.30512	365.65620	P	1128.47769	564.74248	16
6	759.33166	380.16947	787.32658	394.16693	G	1031.42492	516.21610	15
7	816.35313	408.68020	844.34804	422.67766	G	974.40346	487.70537	14
8	873.37459	437.19093	901.36951	451.18839	G	917.38200	459.19464	13
9	930.39606	465.70167	958.39097	479.69912	G	860.36053	430.68390	12
10	987.41752	494.21240	1015.41243	508.20986	G	803.33907	402.17317	11
11	1044.43898	522.72313	1072.43390	536.72059	G	746.31760	373.66244	10
12	1101.46045	551.23386	1129.45536	565.23132	G	689.29614	345.15171	9
13	1158.48191	579.74459	1186.47682	593.74205	G	632.27468	316.64098	8
14	1215.50337	608.25533	1243.49829	622.25278	G	575.25321	288.13025	7
15	1272.52484	636.76606	1300.51975	650.76351	G	518.23175	259.61951	6
16	1329.54630	665.27679	1357.54122	679.27425	G	461.21029	231.10878	5
17	1386.56776	693.78752	1414.56268	707.78498	G	404.18882	202.59805	4
18	1443.58923	722.29825	1471.58414	736.29571	G	347.16736	174.08732	3
19	1558.61617	779.81172	1586.61109	793.80918	D	290.14590	145.57659	2
20					R	175.11895	88.06311	1

M^{Ac}WICPGGGGGGGGGGGGGDR

F

NOTCH2NLC (N2C) exons 1 to 2:

CCAAACTTCGGGCGGCGGCTGAGGCGGCGGCCGAGGAGCGGCGGACTCGGGGCGCGGGGAGTCGAGGCATTTGCGCCTGT
GCTTCGGACCGTAGCGCCAGGGCCTGAGCCTTTGAAGCAGGAGGAGGGGAGGAGAGAGTGGGGCTCCTCTATCGGGACCCC
CTCCCC**ATG**TGGATCTGCCCA**GGCGGCGGC...GGCGGCGGC**GGAGGAGGCGGCGACCGAGAAGATGCCCGCCCTGCGCCG
CTCTGCTGTGGGCGCTGCTGGCGCTCTGGCTGTGCTGCGCGACCCCCGCGCATGgtgagta…intron1…atttttagCATTGCAGTGTC
GAGATGGCTATGAACCCTGTG**TAA**ATGAAGGA**ATG**TGTGTTACCTACCACAATGGCACAGGATACTGCAAgtaagttt…intron2

ATG N2C

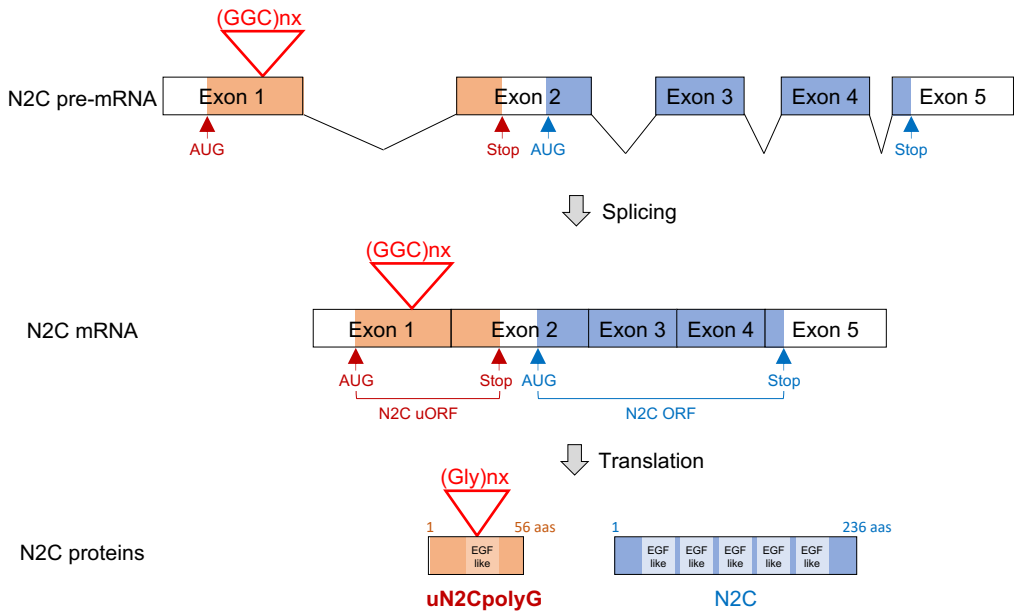


Figure S1. Related to Figure 1.

Figure S1. Related to Figure 1.
***NOTCH2NLC* GGC repeats are translated into polyglycine.**

(**A**) Sequence of the human *NOTCH2NLC* exon 1 with GGC repeats fused to the GFP in the glycine frame and cloned into pcDNA3.

(**B**) RT-qPCR of the GFP and *RPLP0* mRNA expression of HEK293 cells transfected for 24 hours with GGC repeats embedded in the *NOTCH2NLC* exon 1 and fused to the GFP in either the glycine, alanine or arginine frame.

(**C**) GFP fluorescence of HEK293 cells transfected for 24 hours with 100 GGC repeats cloned downstream of an artificial ATG start codon and fused to the GFP in the three possible frames.

(**D**) Immunoblot against the GFP or the GAPDH of HEK293 cells transfected for 24 hours with 100 GGC repeats cloned downstream of an artificial ATG start codon and fused to the GFP in either the alanine or the arginine frame, encoding either a polyalanine or a polyarginine-containing protein, respectively.

(**E**) LC-MS/MS table analysis of the N-terminal peptide from GFP-immunoprecipitated and trypsin digested protein expressed from HEK293 cells transfected with *NOTCH2NLC* exon 1 with GGC repeats and fused to the GFP in the glycine frame.

(**F**) Upper panel, sequence of the human *NOTCH2NLC* exons 1 to 2. Exon sequences are in upper cases, intronic sequences are in grey lower cases. Lower panel, scheme of the *NOTCH2NLC* transcript encoding the upstream (uN2C) and main (N2C) ORFs, which contain 1 and 5 potential EGF-like domains, respectively. Upstream (uN2C) and main (N2C) NOTCH2NLC ORFs are indicated in orange and blue, respectively.

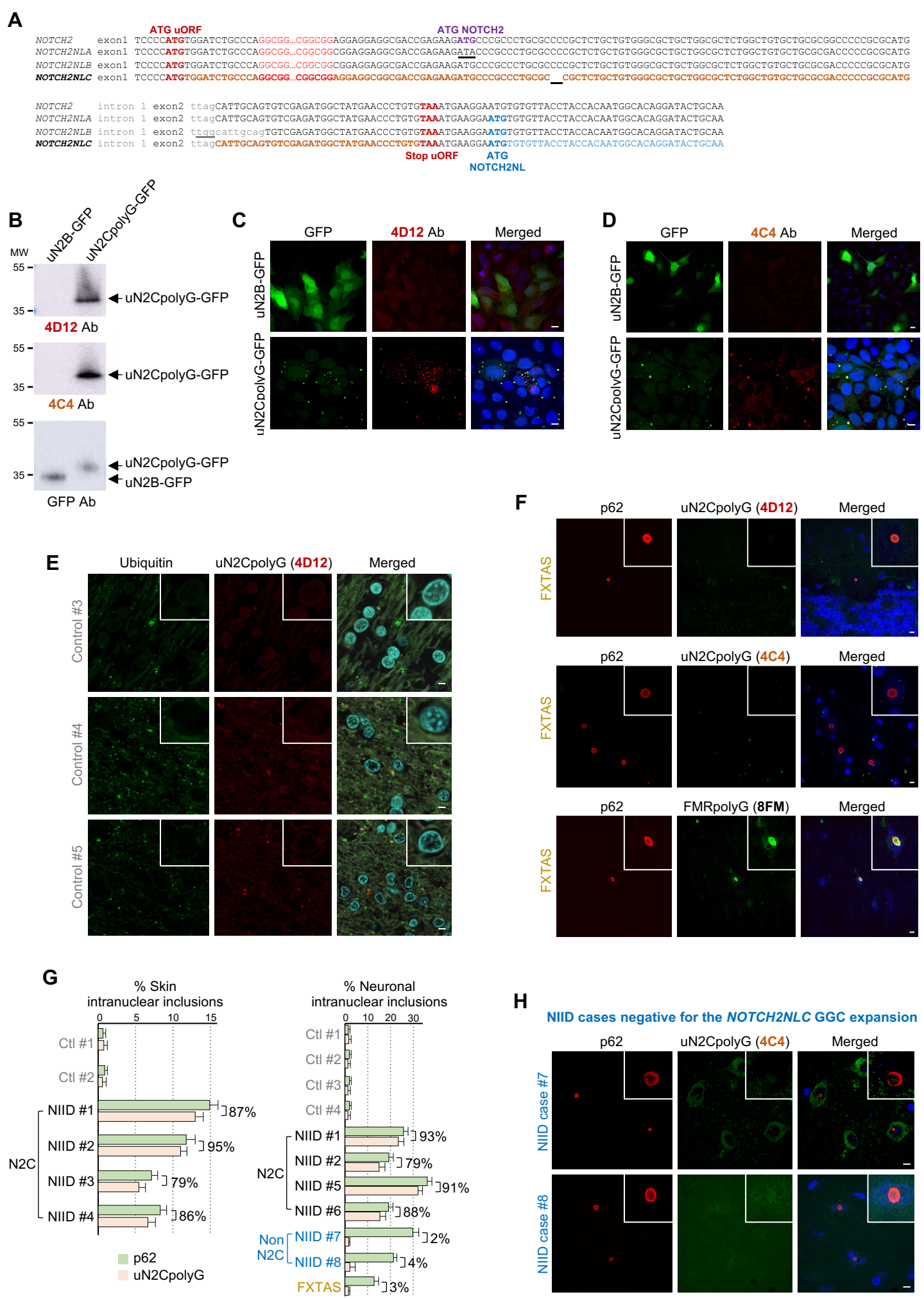


Figure S2. Related to Figure 2.

Figure S2. Related to Figure 2.
uN2polyG is present in NIID nuclear inclusions.

(**A**) Alignment of *NOTCH2* and *NOTCH2NLA*, B and C exons 1 and 2 sequences. Exon sequences are in upper cases, intronic sequences are in grey lower cases. Sequence encoding the *NOTCH2NLC* upstream (uN2C) and main (N2C) ORFs are indicated in orange and blue, respectively. ATG start and stop codons of the potential upstream ORFS are indicated in bold red, while ATG start codons of the NOTCH2NL main ORFs are indicated in bold blue. Of interest, NOTCH2NL proteins are likely cytosolic, as they start in exon 2 and thus lack NOTCH2 peptide signal sequence encoded by the exon 1. Absence of peptide signal in NOTCH2NL proteins is caused by different mechanisms, as the corresponding *NOTCH2* ATG start codon in exon 1 is absent in *NOTCH2NLA* and out of frame in *NOTCH2NLB* and *NOTCH2NLC* due to either mutation of the 3' splice site of *NOTCH2NLB* intron 1 or deletion of 2 nucleotides in *NOTCH2NLC* exon 1. Mutations of *NOTCH2NLA* ATG and *NOTCH2NLB* 3' splice site as well as the 2-nts deletion within *NOTCH2NLC* exon 1 are underlined.

(**B**) Immunoblot of uN2B-GFP or uN2CpolyG-GFP transfected HEK293 cells with 4C4, 4D12 and anti-GFP antibodies.

(**C**, **D**) GFP fluorescence and immunofluorescence of uN2B-GFP and uN2CpolyG-GFP transfected HEK293 cells using either 4D12 (**C**) or 4C4 (**D**) antibody.

(**E**) Control immunofluorescence against uN2CpolyG (4D12 antibody) and ubiquitin on brain sections of control individuals.

(**F**) Immunofluorescence against p62 and either uN2CpolyG (4D12 or 4C4 antibody) or FMRpolyG (8FM antibody) on brain sections of a FXTAS case, carrier of an expansion of CGG repeats within the 5'UTR of the *FMR1* gene.

(**G**) Quantification of p62 or uN2CpolyG-positive intranuclear inclusions per 100 nuclei in skin (left panel) or brain (right panel) sections of control, FXTAS or NIID (carrier or not of an expansion of GGC repeats within *NOTCH2NLC*) individuals. Brackets indicate the percentage of co-localization between p62- and uN2polyG-positive intranuclear inclusions. At least 100 nuclei were counted per individual.

(**H**) Immunofluorescence against p62 and uN2CpolyG (4C4 antibody) on brain sections of two European NIID cases negative for the *NOTCH2NLC* GGC mutation.

Scale bars, 10 µm. Nuclei were counterstained with DAPI.

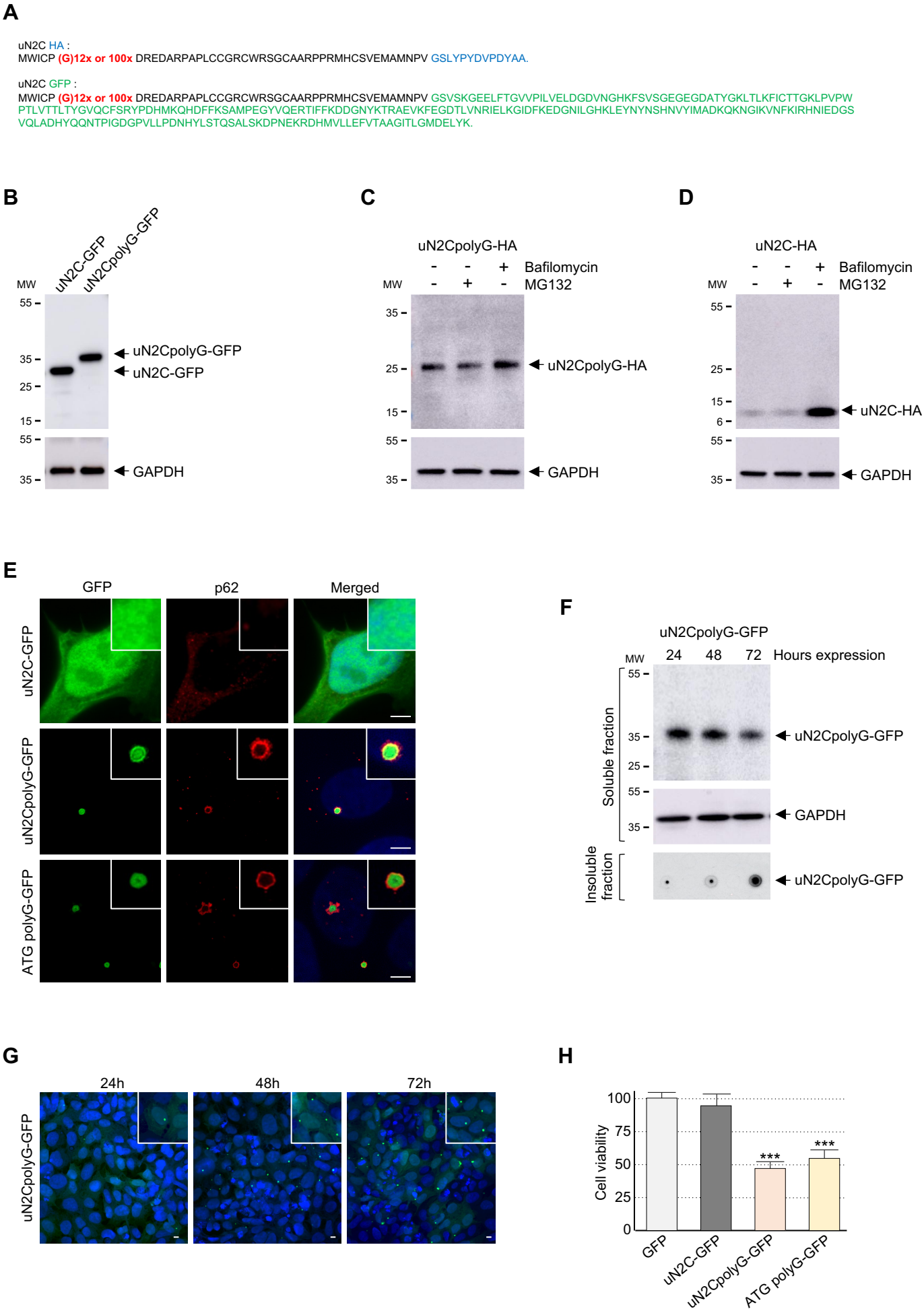


Figure S3. Related to Figure 3.

Figure S3. Related to Figure 3.
Expression of uN2CpolyG is toxic in cell culture.

(**A**) Sequences of the uN2C and uN2CpolyG-HA or -GFP tagged proteins.

(**B**) Immunoblot against the GFP or the GAPDH of proteins extracted from HEK293 cells transfected for 24 hours with uN2C-GFP or uN2CpolyG-GFP.

(**C, D**) Immunoblot against the HA tag or the GAPDH of proteins extracted from HEK293 cells transfected for 24 hours with either uN2CpolyG-HA (C) or uN2C-HA (D) and treated with or without MG132 and/or Bafilomycin A1 for 15 hours.

(**E**) GFP fluorescence and immunofluorescence against p62 of U2OS cells transfected with constructs expressing either the GFP, uN2C-GFP, uN2CpolyG-GFP or ATG polyG-GFP. Scale bars, 5 μm. Nuclei were counterstained with DAPI.

(**F**) Immunoblot on whole cell extract (soluble fraction) or dot blot on urea-treated pellet (insoluble fraction) against the GFP of HEK293 cells transfected with uN2CpolyG-GFP for 24, 48 and 72 hours.

(**G**) GFP fluorescence of U2OS cells expressing uN2CpolyG-GFP for 24, 48 and 72 hours. Scale bars, 10 μm. Nuclei were counterstained with DAPI.

(**H**) Cell viability of GT1-7 neuronal cells transfected for 24 hours with either the GFP, uN2C-GFP, uN2CpolyG-GFP or ATG polyG-GFP. Error bars indicate SEM. Student t-test, *** indicates $p<0.001$.

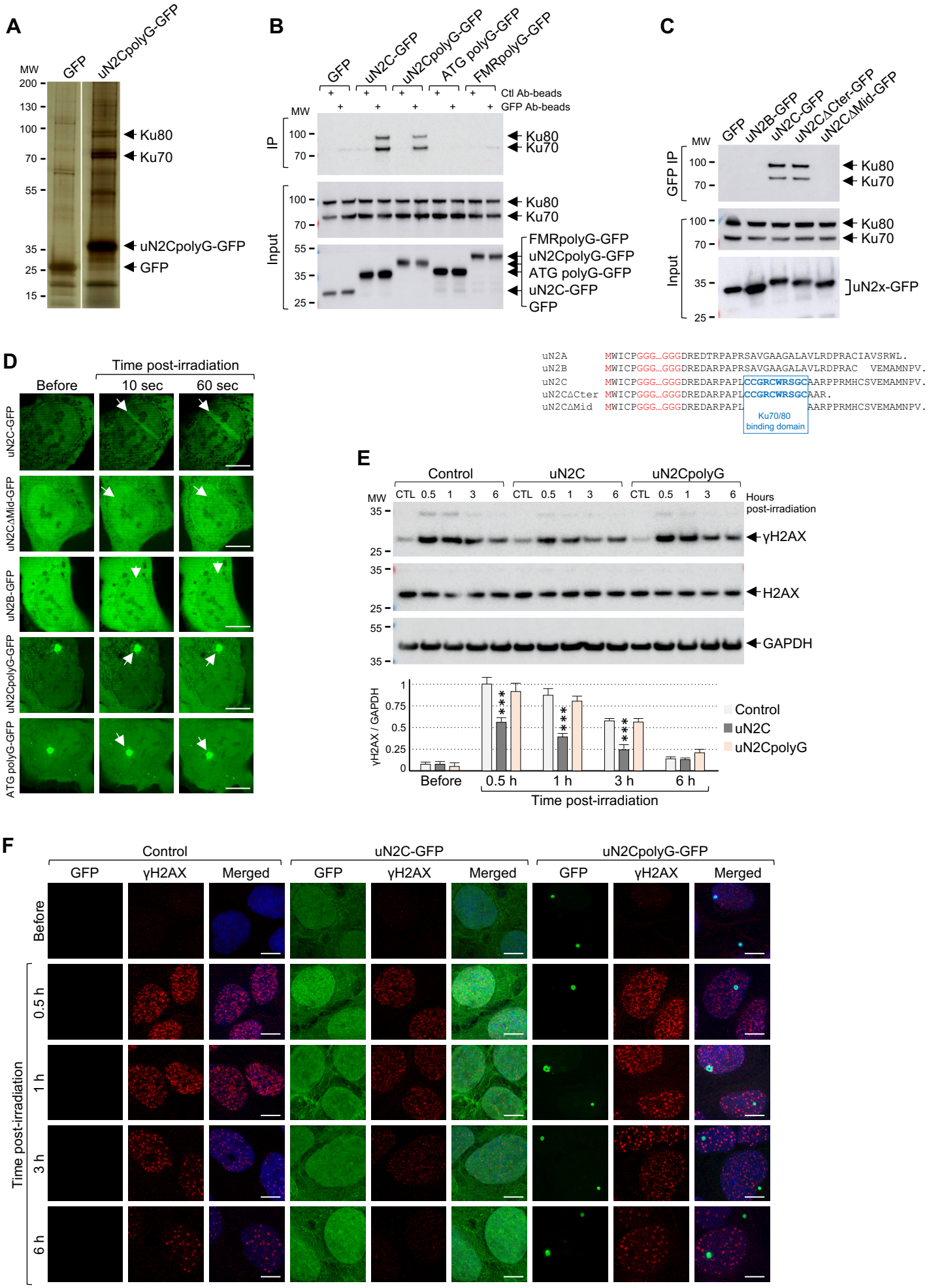


Figure S4. Related to Figure 3.

Figure S4. Related to Figure 3.
The uN2C control protein promotes DNA repair

(**A**) Silver staining of the proteins captured by GFP immunoprecipitation from HEK293 cells transfected for 24 hours with either GFP or uN2CpolyG-GFP.

(**B**) Immunoblot against endogenous Ku70 and Ku80 of either control antibody (anti-HA) or GFP immunoprecipitated proteins (IP) or whole cell lysate (Input) from HEK293 cells transfected for 24 hours with either GFP, uN2C-GFP, uN2CpolyG-GFP, ATG polyG-GFP or FMRpolyG-GFP.

(**C**) Upper panel, immunoblot against endogenous Ku70 and Ku80 of GFP immunoprecipitated proteins or whole cell lysate from HEK293 cells transfected for 24 hours with either GFP, uN2B-GFP or mutants of uN2C-GFP. Lower panel, sequence alignment of the upstream ORFs of *NOTCH2NLA*, *B* and C wild type or mutant constructs. The uN2C amino acid sequence important for binding to Ku proteins is indicated in bold blue.

(**D**) U2OS cells transfected 24 hours with either uN2C-GFP, mutant uN2CΔMid-GFP, uN2B-GFP, uN2CpolyG-GFP or ATG polyG-GFP were subjected to laser micro-irradiation. White arrows indicate laser stripes. Scale bars, 10 μm. Nuclei were counterstained with DAPI.

(**E, F**) Immunoblot against H2AX, γH2AX or the GAPDH (E) or GFP fluorescence and immunofluorescence against γH2AX (F) of control cells or U2OS stably expressing either uN2C-GFP or uN2CpolyG-GFP, X-ray irradiated for 10 minutes, and let to recover for 30 minutes, 1, 3 or 6 hours. Lower panel, quantification of γH2AX / GAPDH ratio before and after X-ray irradiation. Error bars indicate SEM. Student t-test, *** indicates $p<0.001$. Scale bars, 10 μm. Nuclei were counterstained with DAPI.

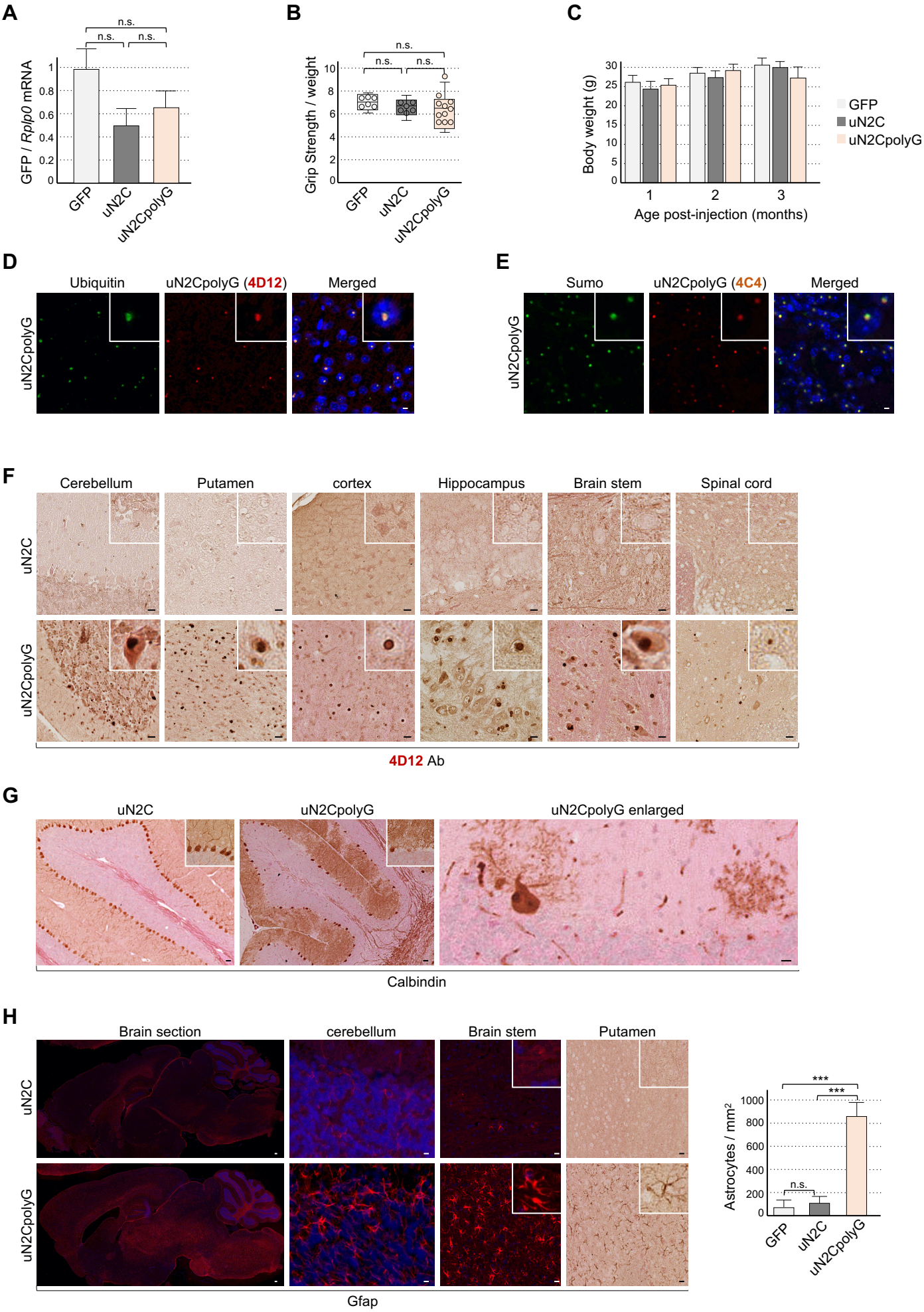


Figure S5. Related to Figure 4.

Figure S5. Related to Figure 4.
Expression of uN2CpolyG is pathogenic in animal.

(**A**) RT-qPCR analysis of GFP expression relative to *Rplp0* mRNA in brain of AAV2/PHP.eB GFP (n=3), uN2C-GFP (n=3) or uN2CpolyG-GFP (n=3) injected mice.

(**B**) Grip test. Maximal force relative to mouse body weight exerted to releases AAV2/PHP.eB GFP (n=6), uN2C-GFP (n=6) and uN2CpolyG-GFP (n=11) injected male mice holding a grid with their forepaws and tested 3 months post-injection. Box-and-whisker plot, box upper and lower limits represent 25th and 75th percentiles, whiskers represent minimum and maximum values and the horizontal line across the box represents the median.

(**C**) Body weight at 1, 2 and 3 months post-injection of AAV2/PHP.eB GFP (n=6), uN2C-GFP (n=6) and uN2CpolyG-GFP (n=11) injected male mice.

(**D**, **E**) Immunofluorescence against uN2CpolyG using either the 4D12 (D) or 4C4 (E) antibody and ubiquitin (D) or sumo (E) on putamen areas of uN2CpolyG-GFP expressing mice scarified 2 months after AAV injection. Scale bars, 10 µm. Nuclei were counterstained with DAPI.

(**F**) Immunohistochemistry against uN2CpolyG of cerebellum, putamen, cortical, hippocampal, brain stem and spinal cord areas of uN2C-GFP and uN2CpolyG-GFP expressing mice scarified 3 months post AAV injection. Sections were counterstained with hematoxylin and eosin. Scale bars, 20 µm.

(**G**) Calbindin immunohistochemistry of cerebellum areas of uN2C-GFP and uN2CpolyG-GFP expressing mice scarified 4 months post AAV injection. Scale bars, 10 µm.

(**H**) Left panel, immunofluorescence or immunohistochemistry against GFAP of brain sections of uN2C-GFP and uN2CpolyG-GFP expressing mice scarified 3.5 months post AAV injection. IF, nuclei were counterstained with DAPI. IHC, sections were counterstained with hematoxylin and eosin. Scale bars, 10 µm. Right panel, quantification of Gfap-positive cells in the putamen of GFP (n=4), uN2C-GFP (n=4) or uN2CpolyG-GFP (n=4) expressing mice. Bar graphs indicate standard error of the mean (SEM). Student t-test, *** indicates $p<0.001$.

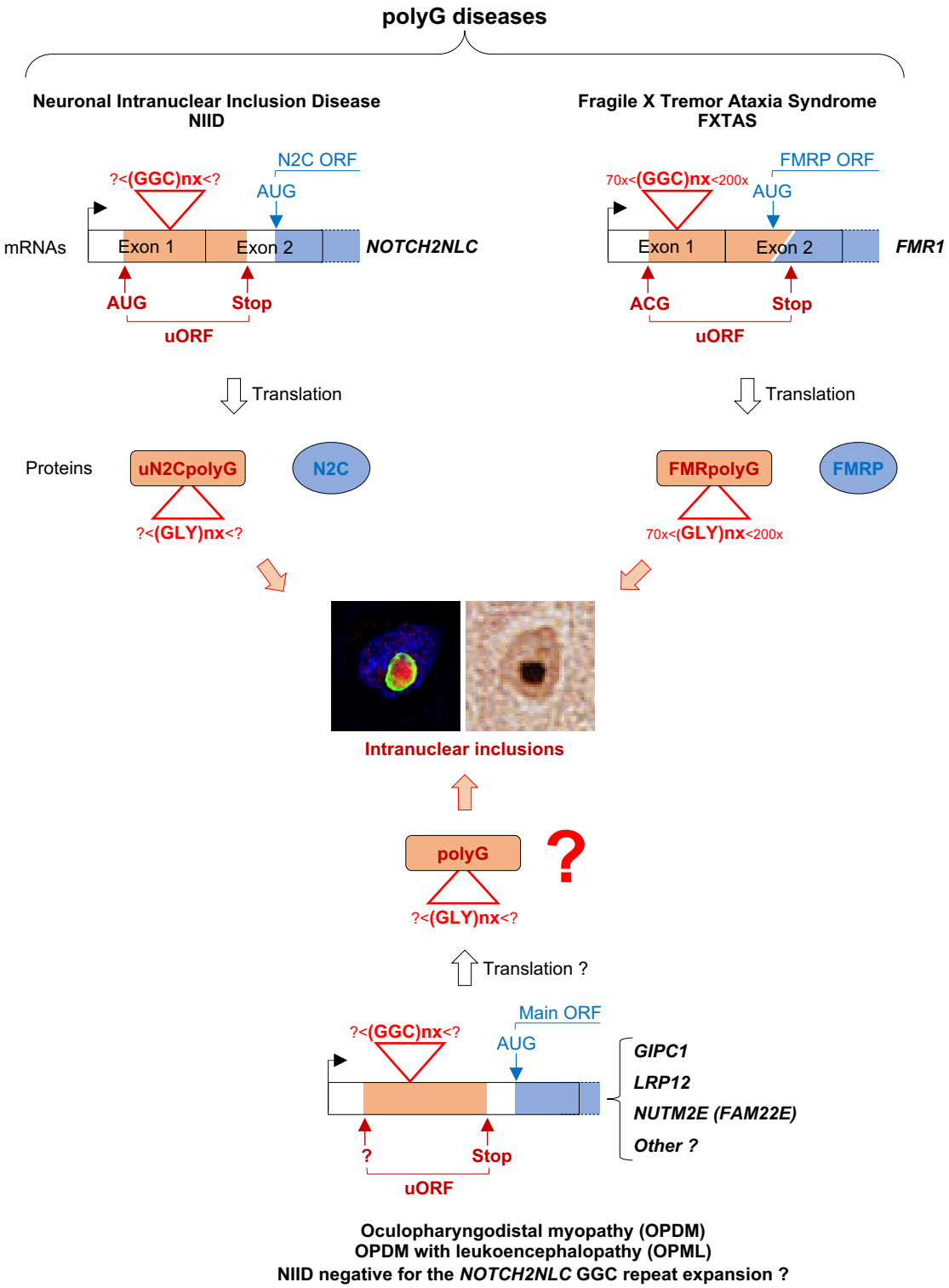


Figure S6. Related to Figure 4.
Model of polyG diseases.

Expanded GGC repeats embedded within upstream ORFs located into *FMR1* and *NOTCH2NLC* 5'UTR are translated into polyglycine containing proteins, FMRpolyG and uN2CpolyG, which are toxic and form p62-positive intranuclear inclusions in FXTAS and NIID, respectively. A similar GGC repeat translation mechanism into yet to discover polyglycine proteins may occur in OPDM and OPML.

Patient	Gender	Age at onset	Duration (months)	Family history	Initial symptom	GGC repeats,	Reference
NIID #1	Female	68	<1	Sporadic	Dizziness & vomiting	102 *NOTCH2NLC*	-
NIID #2	Male	60	12	Sporadic	Paroxysmal encephalopathy	142 *NOTCH2NLC*	Deng et al., 2019
NIID #3	Female	62	60	Familial	Tremor	n.a.	-
NIID #4	Male	63	24	Sporadic	Urinary incontinence	115 *NOTCH2NLC*	-
NIID #5	Male	7	72	Sporadic	Tremor, ataxia,	n.a.	McFadden et al., 2005
NIID #6	Male	30	37	Familial	Muscle weakness	162 *NOTCH2NLC*	Sone et al., 2019
NIID #7	Female	65	3	Sporadic	Memory problems	No expansions *FMR1, NOTCH2NLC*	Cupidi et al., 2019
NIID #8	Male	50	16	Sporadic	Behavioral changes	No expansions *FMR1, NOTCH2NLC*	-
FXTAS	Male	43	22	Familial	Memory problems, aphasia	77 *FMR1*	-
CTL #1	Female	53	-	-	Memory decline	-	-
CTL #2	Female	51	-	-	Urinary difficulties	-	-
CTL #3	Male	66	-	-	Sudden death	-	-
CTL #4	Male	70	-	-	Interstitial pneumonia	-	-
CTL #5	Female	67	-	-	polymyositis	-	-

n.a. Not available

Table S1. Related to Figure 2. Patient information.

Accession	Description	MW [kDa]	Coverage [%]	# Peptides	# PSMs	PSM GFP	PSM uN2CpolyG	uN2cpolyG / GFP
P12956	X-ray repair cross-complementing protein 6 OS=Homo sapiens OX=9606 GN=XRCC6 PE=1 S	70	54	34	460	1	94	94
P13010	X-ray repair cross-complementing protein 5 OS=Homo sapiens OX=9606 GN=XRCC5 PE=1 S	83	76	43	522	1	123	92
P78527	DNA-dependent protein kinase catalytic subunit OS=Homo sapiens OX=9606 GN=PRKDC PE	469	9	24	48	1	36	36
Q16531	DNA damage-binding protein 1 OS=Homo sapiens OX=9606 GN=DDB1 PE=1 SV=1	127	28	23	72	1	41	31
O14744	Protein arginine N-methyltransferase 5 OS=Homo sapiens OX=9606 GN=PRMT5 PE=1 SV=	73	54	26	148	3	32	12
Q9BS26	Endoplasmic reticulum resident protein 44 OS=Homo sapiens OX=9606 GN=ERP44 PE=1 SV	47	48	17	115	3	27	10
P27708	CAD protein OS=Homo sapiens OX=9606 GN=CAD PE=1 SV=3	243	28	42	131	3	26	10
P13639	Elongation factor 2 OS=Homo sapiens OX=9606 GN=EEF2 PE=1 SV=4	95	57	41	211	6	53	9
Q99714	3-hydroxyacyl-CoA dehydrogenase type-2 OS=Homo sapiens OX=9606 GN=HSD17B10 PE=1	27	50	8	39	1	9	9
P61962	DDB1- and CUL4-associated factor 7 OS=Homo sapiens OX=9606 GN=DCAF7 PE=1 SV=1	39	29	8	46	1	12	9
P11586	C-1-tetrahydrofolate synthase, cytoplasmic OS=Homo sapiens OX=9606 GN=MTHFD1 PE=1	102	12	10	34	1	9	9
P30153	Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform OS=H	65	42	20	77	2	20	9
P49411	Elongation factor Tu, mitochondrial OS=Homo sapiens OX=9606 GN=TUFM PE=1 SV=2	50	37	13	53	2	14	8
P35579	Myosin-9 OS=Homo sapiens OX=9606 GN=MYH9 PE=1 SV=4	226	12	17	54	1	11	8
Q92945	Far upstream element-binding protein 2 OS=Homo sapiens OX=9606 GN=KHSRP PE=1 SV=	73	37	20	88	2	19	8
Q9H845	Acyl-CoA dehydrogenase family member 9, mitochondrial OS=Homo sapiens OX=9606 GN=	69	16	8	24	1	8	8
Q9BQA1	Methylosome protein 50 OS=Homo sapiens OX=9606 GN=WDR77 PE=1 SV=1	37	33	10	61	2	13	8
P07237	Protein disulfide-isomerase OS=Homo sapiens OX=9606 GN=P4HB PE=1 SV=3	57	23	11	44	1	11	8
P22102	Trifunctional purine biosynthetic protein adenosine-3 OS=Homo sapiens OX=9606 GN=GAR	108	14	9	29	1	8	8
P30048	Thioredoxin-dependent peroxide reductase, mitochondrial OS=Homo sapiens OX=9606 GN	28	51	8	59	2	16	8
O75534	Cold shock domain-containing protein E1 OS=Homo sapiens OX=9606 GN=CSDE1 PE=1 SV=	89	25	15	50	2	13	8
Q9H857	5'-nucleotidase domain-containing protein 2 OS=Homo sapiens OX=9606 GN=NT5DC2 PE=	61	21	10	31	1	10	8
P63151	Serine/threonine-protein phosphatase 2A 55 kDa regulatory subunit B alpha isoform OS=H	52	53	19	74	2	15	7
Q92973	Transportin-1 OS=Homo sapiens OX=9606 GN=TNPO1 PE=1 SV=2	102	13	8	22	1	7	7
P54136	Arginine--tRNA ligase, cytoplasmic OS=Homo sapiens OX=9606 GN=RARS PE=1 SV=2	75	24	11	46	1	10	7
P04843	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1 OS=Homo sapien	69	37	15	43	2	14	7
Q13162	Peroxiredoxin-4 OS=Homo sapiens OX=9606 GN=PRDX4 PE=1 SV=1	31	58	13	82	3	19	7
P42771	Cyclin-dependent kinase inhibitor 2A OS=Homo sapiens OX=9606 GN=CDKN2A PE=1 SV=2	17	70	7	33	1	9	7
P31689	DnaJ homolog subfamily A member 1 OS=Homo sapiens OX=9606 GN=DNAJA1 PE=1 SV=2	45	36	9	47	2	12	7
Q5H9R7	Serine/threonine-protein phosphatase 6 regulatory subunit 3 OS=Homo sapiens OX=9606 (	98	28	19	54	2	16	7
P49915	GMP synthase [glutamine-hydrolyzing] OS=Homo sapiens OX=9606 GN=GMPS PE=1 SV=1	77	19	12	29	1	9	7
P63244	Receptor of activated protein C kinase 1 OS=Homo sapiens OX=9606 GN=RACK1 PE=1 SV=	35	50	9	33	1	9	7
Q9Y265	RuvB-like 1 OS=Homo sapiens OX=9606 GN=RUVBL1 PE=1 SV=1	50	28	9	44	1	9	7
P53396	ATP-citrate synthase OS=Homo sapiens OX=9606 GN=ACLY PE=1 SV=3	121	10	8	24	1	7	7
P21333	Filamin-A OS=Homo sapiens OX=9606 GN=FLNA PE=1 SV=4	281	10	17	46	1	9	7
P28482	Mitogen-activated protein kinase 1 OS=Homo sapiens OX=9606 GN=MAPK1 PE=1 SV=3	41	53	15	61	2	13	6
Q93009	Ubiquitin carboxyl-terminal hydrolase 7 OS=Homo sapiens OX=9606 GN=USP7 PE=1 SV=2	128	12	10	25	1	6	6
Q9Y3F4	Serine-threonine kinase receptor-associated protein OS=Homo sapiens OX=9606 GN=STRA	38	46	12	58	2	13	6
O60884	DnaJ homolog subfamily A member 2 OS=Homo sapiens OX=9606 GN=DNAJA2 PE=1 SV=1	46	23	7	23	1	6	6
P67775	Serine/threonine-protein phosphatase 2A catalytic subunit alpha isoform OS=Homo sapien	36	56	12	61	2	15	6
P27361	Mitogen-activated protein kinase 3 OS=Homo sapiens OX=9606 GN=MAPK3 PE=1 SV=4	43	28	8	30	1	8	6
P45880	Voltage-dependent anion-selective channel protein 2 OS=Homo sapiens OX=9606 GN=VDA	32	43	8	31	1	8	6
P22234	Multifunctional protein ADE2 OS=Homo sapiens OX=9606 GN=PAICS PE=1 SV=3	47	40	14	76	3	16	6
P17812	CTP synthase 1 OS=Homo sapiens OX=9606 GN=CTPS1 PE=1 SV=2	67	26	12	48	2	12	6
Q9UK99	F-box only protein 3 OS=Homo sapiens OX=9606 GN=FBXO3 PE=1 SV=3	55	26	10	33	2	10	6
P13489	Ribonuclease inhibitor OS=Homo sapiens OX=9606 GN=RNH1 PE=1 SV=2	50	36	12	43	2	10	6
O00429	Dynamin-1-like protein OS=Homo sapiens OX=9606 GN=DNM1L PE=1 SV=2	82	19	10	24	1	8	6
P29966	Myristoylated alanine-rich C-kinase substrate OS=Homo sapiens OX=9606 GN=MARCKS PE	32	45	9	54	2	10	6
P49327	Fatty acid synthase OS=Homo sapiens OX=9606 GN=FASN PE=1 SV=3	273	32	50	206	8	46	6
P49366	Deoxyhypusine synthase OS=Homo sapiens OX=9606 GN=DHPS PE=1 SV=1	41	47	11	47	3	16	6
O95831	Apoptosis-inducing factor 1, mitochondrial OS=Homo sapiens OX=9606 GN=AIFM1 PE=1 SV	67	20	9	27	1	8	6
Q9UQ80	Proliferation-associated protein 2G4 OS=Homo sapiens OX=9606 GN=PA2G4 PE=1 SV=3	44	28	8	35	1	7	6
P42704	Leucine-rich PPR motif-containing protein, mitochondrial OS=Homo sapiens OX=9606 GN=	158	12	14	29	2	9	5
P31939	Bifunctional purine biosynthesis protein PURH OS=Homo sapiens OX=9606 GN=ATIC PE=1	65	19	7	16	1	5	5
Q15654	Thyroid receptor-interacting protein 6 OS=Homo sapiens OX=9606 GN=TRIP6 PE=1 SV=3	50	26	8	16	1	5	5
P22314	Ubiquitin-like modifier-activating enzyme 1 OS=Homo sapiens OX=9606 GN=UBA1 PE=1 S	118	34	25	110	5	24	5
Q13263	Transcription intermediary factor 1-beta OS=Homo sapiens OX=9606 GN=TRIM28 PE=1 SV	89	37	16	86	3	17	5
Q8IXB1	DnaJ homolog subfamily C member 10 OS=Homo sapiens OX=9606 GN=DNAJC10 PE=1 SV=	91	17	11	31	2	10	5
P25205	DNA replication licensing factor MCM3 OS=Homo sapiens OX=9606 GN=MCM3 PE=1 SV=3	91	22	15	57	2	10	5
Q96QK1	Vacuolar protein sorting-associated protein 35 OS=Homo sapiens OX=9606 GN=VPS35 PE=	92	28	18	58	3	13	5
P29590	Protein PML OS=Homo sapiens OX=9606 GN=PML PE=1 SV=3	98	9	7	16	1	5	5
P10599	Thioredoxin OS=Homo sapiens OX=9606 GN=TXN PE=1 SV=3	12	54	6	67	3	14	5
P55060	Exportin-2 OS=Homo sapiens OX=9606 GN=CSE1L PE=1 SV=3	110	8	7	20	1	6	5
Q96J01	THO complex subunit 3 OS=Homo sapiens OX=9606 GN=THOC3 PE=1 SV=1	39	33	8	22	1	6	5
P26639	Threonine--tRNA ligase, cytoplasmic OS=Homo sapiens OX=9606 GN=TARS PE=1 SV=3	83	19	10	30	2	9	5
P05141	ADP/ATP translocase 2 OS=Homo sapiens OX=9606 GN=SLC25A5 PE=1 SV=7	33	29	8	38	2	9	5
P05023	Sodium/potassium-transporting ATPase subunit alpha-1 OS=Homo sapiens OX=9606 GN=A	113	16	12	39	2	11	5
Q13185	Chromobox protein homolog 3 OS=Homo sapiens OX=9606 GN=CBX3 PE=1 SV=4	21	29	5	24	1	6	5
P31040	Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial OS=Homo sapie	73	31	12	42	3	12	5

Table S2. Related to Figure 3. Proteins interacting with uN2CpolyG.
Mass spectrometry analysis of GFP-trap immunoprecipitated proteins from HEK293 cells expressing either the GFP or uN2CpolyG-GFP.

RESULTATS

- OPDM2 -

Résultat sur l'OPDM2

La myopathie oculo-pharyngo-distale (OPDM) est une dystrophie rare des muscles de la face, du pharynx et des membres distaux. Cette pathologie est caractérisée par la présence d'inclusions intranucléaires et de vacuoles d'origine inconnue. L'OPDM est due à une expansion de répétitions de tri-nucléotides CGG située dans la partie 5'UTR de différents gènes, conduisant ainsi cette pathologie à être divisée en 3 sous-types selon le gène muté : *LRP12* pour l'OPDM1, *GIPC1* pour l'OPDM2 et *NOTCH2NLC* pour l'OPDM3. Mon travail de thèse a porté sur le mécanisme de toxicité des répétitions CGG situées dans la partie 5'UTR du gène *GIPC1*.

1. Traduction des répétitions CGG présentes dans la région 5'UTR du gène *GIPC1*

Compte tenu du nombre grandissant de maladies à microsatellites dues à la traduction en protéine toxiques d'expansions de répétitions situées dans des régions non codantes, nous avons testé si les répétitions CGG présentes dans la partie 5'UTR du gène *GIPC1* pouvaient éventuellement être traduites en une nouvelle protéine. Pour cela, la séquence de cette région comprenant 11 ou 100 répétitions CGG, reproduisant une situation non pathogénique ou l'OPDM2, a été fusionnée à l'ORF de la GFP dans les 3 cadres de lecture ouverts possibles : Glycine, Alanine ou Arginine (Figure 52A). De plus, afin d'éviter de possibles biais d'efficacité de transfection et / ou d'expression de ces différentes constructions, le plasmide contient aussi l'ADNc codant pour la mCHERRY cloné entre un promoteur et un site de polyadénylation différents de la casette 5'UTR GIPC1-GFP, assurant ainsi une expression de la mCHERRY indépendamment des répétitions CGG fusionnées à la GFP. La transfection de ces 3 constructions dans des cellules HEK293 suivie d'analyses par immunoblot révèle une traduction des répétitions CGG majoritairement dans le cadre de lecture glycine (Figure 52B). Ce résultat est confirmé par des expériences d'immunofluorescence (Figure 52C, 52D), et nous avons alors appelée cette nouvelle protéine uGIPC1-Gly pour « upstream of the GIPC1 ORF and translated in the Glycine frame » et uGIPC1polyG en présence d'une expansion de répétitions CGG. Afin de vérifier que l'absence de traduction dans les cadres de lecture alanine et arginine n'est pas causée par une faible expression de ces constructions au niveau ARN, une quantification par RT-qPCR des ARNm a été réalisé et ne révèle pas de différences d'expression entre ces différentes constructions (Figure 52E).

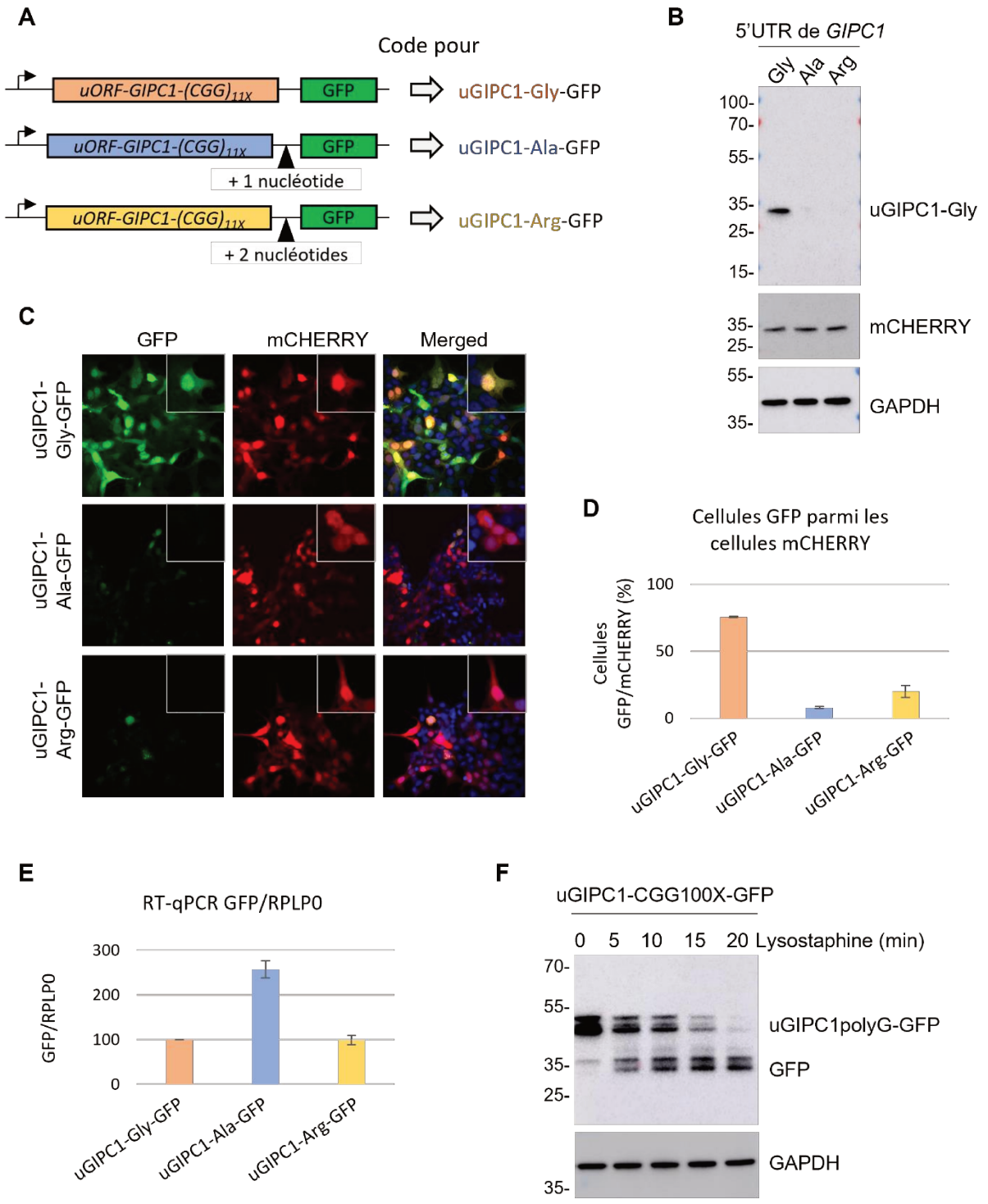


Figure 52. Traduction des expansions de répétitions CGG du gène *GIPC1*

(A) Schéma représentant les différentes constructions de la régions 5'UTR du gène *GIPC1* fusionnée à la GFP. (B-E) Transfection de ces constructions dans des cellules HEK293 suivie (B) d'un immunoblot contre la GFP, la mCHERRY et la GAPDH ; (C) d'une visualisation de la fluorescence et (D) sa quantification ou (E) d'une RT-qPCR. (F) Immunoblot dirigé contre la GFP et la GAPDH après transfection de cellules HEK293 avec une construction codant pour la protéine uGIPC1polyG-GFP et traitement du lysat protéique avec la lysostaphine.

Enfin, le traitement du lysat cellulaire par une enzyme clivant entre 2 résidus glycines successifs, la lysostaphine, conduit à une diminution de taille de la protéine uGIPC1polyG, confirmant ainsi que cette protéine contient une suite d'acides aminés glycines (Figure 52F). Ces résultats suggèrent donc que répétitions CGG situées dans la région 5'UTR du gène *GIPC1* sont potentiellement traduites en une protéine composée de polyGlycine, appelée uGIPC1polyG. Toutefois, il est à noter que ces répétitions CGG sont situées en amont de l'ATG initiateur de la protéine GIPC1, et que la région 5'UTR de *GIPC1* ne contient aucune séquence ATG, soulevant ainsi la question du mécanisme de traduction de la protéine uGIPC1polyG.

2. Mécanisme de traduction de la protéine uGIPC1polyG

Afin de mieux comprendre comment ces répétitions CGG sont potentiellement traduites en absence de codon d'initiation canonique de la traduction ATG, nous avons immunoprécipité la protéine uGIPC1-Gly-GFP puis cherché à identifier la séquence N-terminale de cette protéine par spectrométrie de masse. Cette analyse a révélé un peptide N-terminal de séquence M^{Ac}EFAEGR (Figure 53A) qui, une fois aligné à la région 5'UTR du gène *GIPC1* suggère une initiation de la traduction à un codon non-canonique CTG situé 51 nucléotides avant les répétitions CGG, dans le cadre de lecture glycine et situé dans une séquence Kozac convenable (GguCUGG, la séquence canonique étant RccAUGG) (Figure 53A). Nous avons alors muté ce codon CTG en codon CTT et des expériences d'immunoblot et d'immunofluorescence montrent que cette mutation abolit l'expression de la protéine uGIPC1-Gly-GFP (Figure 53B, 53C, 53D). Le contrôle de l'expression des ARNm uGIPC1, dosés par RT-qPCR, exclue un biais d'efficacité de transcription entre ces 2 constructions (Figure 53E). Enfin, la transfection de constructions avec différentes tailles d'expansion de répétitions CGG suivie d'une analyse par immunoblot révèle que la traduction de la protéine uGIPC1-Gly a lieu quelque soit la taille des répétitions CGG (Figure 53F). Néanmoins, la protéine uGIPC1 avec peu de résidus glycines est petite, instable et sujette à la dégradation. En effet, l'inhibition de la dégradation par le protéasome grâce à du MG132, mais pas de l'autophagie grâce à de la bafilomycine, conduit à une accumulation de cette protéine (Figure 53F).

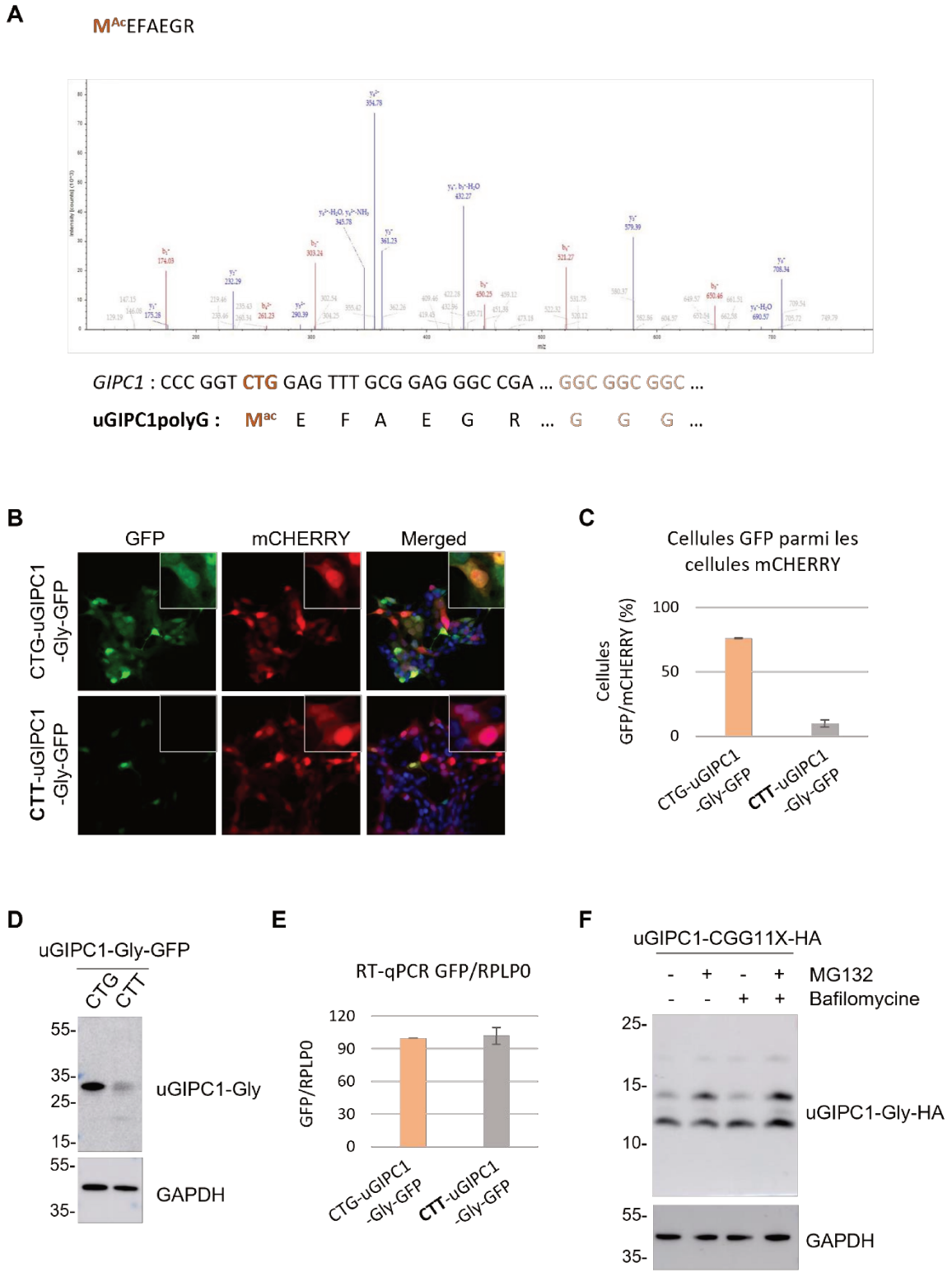


Figure 53. Traduction canonique des répétitions CGG dans le cadre de lecture glycine
(A) Analyse par spectrométrie de masse de la protéine uGIPC1-Gly-GFP après digestion à la trypsine. (B-F) Transfection des constructions dans des cellules HEK293 suivie (B) d'un immunoblot contre la GFP, la mCHERRY et la GAPDH ; (C) d'une visualisation de la fluorescence et (D) sa quantification ; (E) d'une RT-qPCR ou (F) immunoblot dirigé contre l'étiquette HA et la GAPDH après traitement des cellules avec 1 µM de MG132 ou 100 nM de bafilomycine pendant 16h.

Ces données suggèrent que la partie 5'UTR du gène *GIPC1* contient une petite uORF traduite via une initiation à un codon, ressemblant à un ATG, un codon CTG. Ce résultat n'est pas forcément étonnant car la plupart des gènes des mammifères contiennent des uORF dont l'initiation se fait à des codons non canoniques, le codon CTG étant le plus commun (Ingolia et al., 2011). Comme la plupart des autres uORF, cette uORF de *GIPC1* code pour une protéine de petite taille et donc instable et difficile à détecter en condition normale. Toutefois, cette uORF contient des répétitions CGG dont l'expansion dans l'OPDM2 conduit à l'expression d'une protéine contenant une suite d'acides aminés glycines, uGIPC1polyG, et donc de plus grande taille et probablement suffisamment stable pour s'accumuler et être potentiellement toxique. En conséquence, il reste à déterminer si cette nouvelle protéine uGIPC1polyG est observée chez les individus OPDM2, et si son expression peut récapituler les symptômes de cette maladie.

3. Détection de la protéine uGIPC1polyG

Dans le but de tester la présence de la protéine uGIPC1polyG chez les patients OPDM2, nous avons entrepris de développer des anticorps dirigés contre cette protéine. Nous avons purifié sous forme de protéine recombinante la protéine uGIPC1-Gly-GST-6His puis, le service d'anticorps monoclonaux de l'IGBMC l'a injectée dans des souris dans le but d'obtenir des hybridomes entre les lymphocytes B de ces souris et des cellules myéloïdes immortalisées Sp2/0.Agl4. Nous avons alors testé le surnageant de culture des différents clones obtenus contre la protéine uGIPC1polyG-GFP par immunofluorescence et, le clone 4H5 semble reconnaître spécifiquement les inclusions formées par cette protéine (Figure 54A). Nous souhaitons donc produire cet anticorps à plus grande échelle et collaborer avec des cliniciens spécialistes de l'OPDM afin de tester la présence de la protéine uGIPC1polyG dans des lames de muscles squelettiques de patients atteints de l'OPDM de type 2, en comparaison avec d'autres sous types de l'OPDM et/ou d'autre pathologies musculaires.

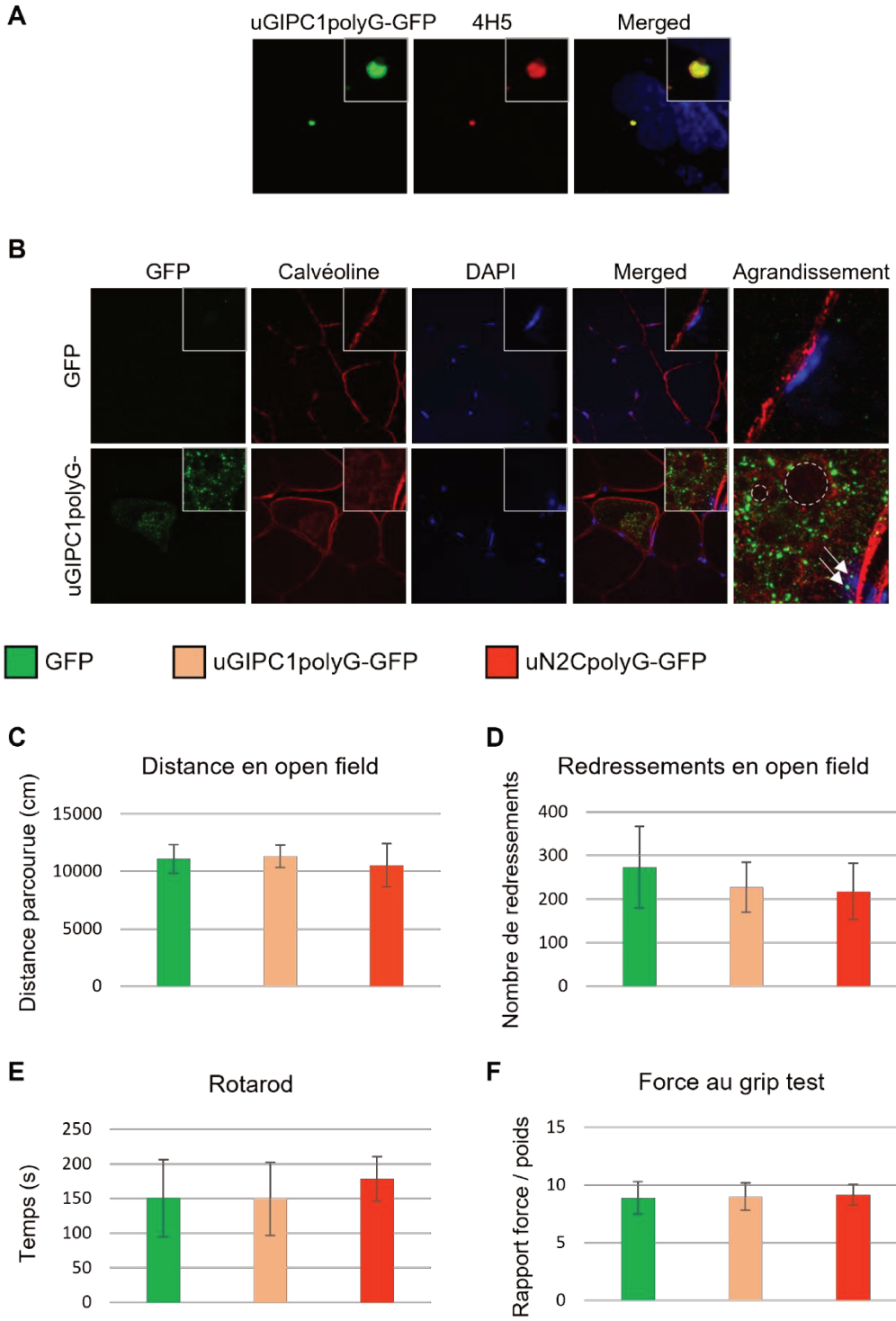


Figure 54. Impact de l'expression des protéines à polyGlycine

(A, B) Immunofluorescences (A) en cellules U2OS exprimant la protéine uGIPC1polyG-GFP et text de l'anticorps 4H5 et (B) sur des lames de muscles tibialis antérieur de souris injectées avec un virus AAV9 exprimant la GFP ou uGIPC1polyG-GFP et sacrifiées 3 mois post-injection. (**C**-**F**) Etude des capacités locomotrices des souris C57BL/6 WT injectées avec $1{,}5.10^{14}$ vg/kg de virus AAV9 codant pour la protéine GFP, uGIPC1polyG-GFP ou uN2CpolyG-GFP à 1 et 2 mois post-injection. (**C**, **D**) Résultats de l'open field avec (**C**) la distance parcourue et (**D**) le nombre de redressements. (**E**) Temps passé par les souris sur le rotarod avant la chute. (**F**) Force développé pendant le grip test.

4. Modèle murin exprimant la protéine uGIPC1polyG

Enfin, dans le but de tester la toxicité de la protéine uGIPC1polyG en modèle animal, nous avons développé avec l'aide de la plateforme de virus recombinant de l'IGBMC des AAV9 codant soit pour la GFP, soit pour la protéine uGIPC1polyG-GFP, soit pour la protéine uN2CpolyG (responsable de NIID et de l'OPDM3). Nous avons alors injecté $1{,}5.10^{14}$ vg/kg de virus par voie rétro-orbitale chez des souris C57BL/6 WT dans le but de transduire l'ensemble des muscles squelettiques de ces animaux. A 3 mois post-injection, nous avons observé par immunofluorescence que la protéine uGIPC1polyG-GFP forme des inclusions intranucléaires et cytoplasmiques dans les fibres musculaires de ces souris (Figure 54B). De manière intéressante, des vacuoles sont aussi visibles chez la souris exprimant la protéine uGIPC1polyG, mais non chez les souris exprimant la GFP (Figure 54B). Il est à noter que ces inclusions et vacuoles reproduisent l'histopathologie typique des patients OPDM. Enfin, un suivi du phénotype de ces souris GFP, uGIPC1polyG et uN2CpolyG est en cours. Un mois après l'injection virale, aucune altération locomotrice obtenue par les tests de l'open field, du rotarod et du grip test n'a été observée (Figure 54C, 54D, 54E, 54F), mais nous espérons observer des altérations locomotrices à des temps de suivi plus long.

En conclusion, nous avons montré que les répétitions CGG située dans la partie 5'UTR de *GIPC1* sont traduites en une nouvelle protéine, potentiellement toxique, et contenant une suite de polyGlycine, que nous avons appelé uGIPC1polyG. Cette protéine est traduite en absence de codon d'initiation ATG canonique et nos analyses montrent que l'initiation de la traduction de cette protéine se fait à un codon non-canonique CTG situé en amont des répétitions CGG. Il est à noter que ce mécanisme rappelle nos résultats sur la traduction des répétitions GGGGCC en protéine polyGlycine-Alanine dans la sclérose latérale amyotrophique (Boivin et al., 2020). De même, dans la maladie FXTAS, une expansion de répétitions CGG dans le 5'UTR du gène *FMR1* est traduite en une protéine polyGlycine grâce à une initiation de la traduction à un codon non canonique ACG situé avant les répétitions (Sellier et al., 2017). Ces données suggèrent que la traduction des répétitions en protéines toxiques à partir de codons d'initiation ressemblant à des ATG puisse être un nouveau mécanisme de pathogénicité dans les maladies génétiques à expansions.

RESULTATS

- ARTICLE 2 -

Reduced autophagy upon C9ORF72 loss synergizes with dipeptide repeat protein toxicity in G4C2 repeat expansion disorders

Article 2

Reduced autophagy upon C9ORF72 loss synergizes with dipeptide repeat protein toxicity in G4C2 repeat expansion disorders

Boivin M, Pfister V, Gaucherot A, Ruffenach F, Negroni L, Sellier C, Charlet-Berguerand N.

Publié dans *EMBO Journal 2020 Feb 17;39(4):e100574. doi: 10.15252/embj.2018100574*

La sclérose latérale amyotrophique (SLA) est la troisième maladie neurodégénérative la plus commune. Cette maladie dévastatrice est caractérisée par la mort des neurones moteurs supérieurs et inférieurs, conduisant à une atrophie et une paralysie musculaire fatale. La cause génétique la plus fréquente de SLA est une expansion de répétitions GGGGCC (G4C2) située dans le premier intron du gène *C9ORF72*. Ces répétitions induisent une diminution d'expression de la protéine C9ORF72, impliquée dans l'autophagie, et sont en parallèle traduites en protéines DPR toxiques, toutefois selon un mécanisme encore mal caractérisé.

Mon travail de thèse a porté sur la traduction de ces répétitions. Des expériences de transfection cellulaires et de spectrométrie de masse montrent que les répétitions sens GGGGCC sont traduites en une protéine polyGlycine-Alanine à partir d'un codon CTG situé en amont des répétitions. Au contraire, les répétitions antisenses CCCCGG sont, elles, traduites en une protéine composée de polyGlycine-Proline grâce à un codon ATG situé en amont des répétitions. Des expériences complémentaires montrent que ces protéines s'expriment faiblement car elles sont constamment dégradées par autophagie. La protéine C9ORF72 régulant l'autophagie (Sellier et al., 2016), nous avons donc testé s'il existait un lien entre le niveau d'expression de la protéine C9ORF72 et l'accumulation des protéines DPR. De façon intéressante, la déplétion par siARN de C9ORF72 conduit à une accumulation des protéines DPR polyGA et polyGP en modèle cellulaire. De plus, une mesure de la viabilité cellulaire montre une plus forte toxicité de ces protéines DPR lorsque la protéine C9ORF72 est déplétée. Enfin, l'utilisation de substances connues pour stimuler l'autophagie diminue la quantité de protéines DPR ainsi que leur toxicité, ouvrant ainsi une piste thérapeutique

Ces résultats suggèrent un mécanisme de toxicité synergique, où la déplétion de la protéine C9ORF72 conduit à une autophagie sous-optimale, ce qui aboutit à l'accumulation et à une plus forte toxicité des protéines DPR.

Article

TRANSPARENT PROCESS OPEN ACCESS

THE EMBO JOURNAL

Reduced autophagy upon C9ORF72 loss synergizes with dipeptide repeat protein toxicity in G4C2 repeat expansion disorders

Manon Boivin[1,2,3,4], Véronique Pfister[1,2,3,4], Angeline Gaucherot[1,2,3,4], Frank Ruffenach[1,2,3,4], Luc Negroni[1,2,3,4], Chantal Sellier[1,2,3,4,*] & Nicolas Charlet-Berguerand[1,2,3,4,**]

Abstract

Expansion of G4C2 repeats within the C9ORF72 gene is the most common cause of amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). Such repeats lead to decreased expression of the autophagy regulator C9ORF72 protein. Furthermore, sense and antisense repeats are translated into toxic dipeptide repeat (DPR) proteins. It is unclear how these repeats are translated, and in which way their translation and the reduced expression of C9ORF72 modulate repeat toxicity. Here, we found that sense and antisense repeats are translated upon initiation at canonical AUG or near-cognate start codons, resulting in polyGA-, polyPG-, and to a lesser degree polyGR-DPR proteins. However, accumulation of these proteins is prevented by autophagy. Importantly, reduced C9ORF72 levels lead to suboptimal autophagy, thereby impairing clearance of DPR proteins and causing their toxic accumulation, ultimately resulting in neuronal cell death. Of clinical importance, pharmacological compounds activating autophagy can prevent neuronal cell death caused by DPR proteins accumulation. These results suggest the existence of a double-hit pathogenic mechanism in ALS/FTD, whereby reduced expression of C9ORF72 synergizes with DPR protein accumulation and toxicity.

Keywords amyotrophic lateral sclerosis; autophagy; C9ORF72; neurodegeneration; RAN translation

Subject Categories Neuroscience; RNA Biology; Translation & Protein Quality

DOI 10.15252/embj.2018100574 | Received 27 August 2018 | Revised 28 November 2019 | Accepted 3 December 2019 | Published online 13 January 2020

The EMBO Journal (2020) 39: e100574

Introduction

With one individual affected in ~50,000 people, amyotrophic lateral sclerosis (ALS) is the third most common neurodegenerative disease. ALS is characterized by degeneration of both upper and spinal motor neurons that leads to progressive skeletal muscle paralysis, ultimately resulting in death from respiratory failure generally 3–5 years after age of onset. ALS patients can present significant clinical, genetic, and histopathological overlaps with frontotemporal dementia (FTD), a presenile dementia affecting frontal and temporal brain regions (Lomen-Hoerth *et al*, 2002; Ringholz *et al*, 2005; Neumann *et al*, 2006). Consequently, ALS and FTD are now considered as two extremes of a same disease continuum. This notion is emphasized by the identification of an expansion of GGGGCC (G4C2) repeats within the first intron of the *C9ORF72* gene as the most common inherited cause for both ALS and FTD in Caucasian populations of northern Europe and North America (DeJesus-Hernandez *et al*, 2011; Renton *et al*, 2011; Gijselinck *et al*, 2012; Majounie *et al*, 2012).

Expanded G4C2 repeats are pathogenic through three main non-exclusive mechanisms. First, transcripts containing expanded repeats accumulate in RNA foci that recruit various RNA-binding proteins, potentially altering their localization and functions (Almeida *et al*, 2013; Donnelly *et al*, 2013; Lagier-Tourenne *et al*, 2013; Lee *et al*, 2013; Mizielinska *et al*, 2013; Cooper-Knock *et al*, 2014; Haeusler *et al*, 2014; Conlon *et al*, 2016, 2018). Second, sense G4C2 and antisense C4G2 expanded repeats are repeat-associated non-AUG (RAN) translated into dipeptide repeat (DPR)-containing proteins, which form inclusions throughout the brain of patients with C9-ALS/FTD (Ash *et al*, 2013; Gendron *et al*, 2013; Mori *et al*, 2013; Zu *et al*, 2013), as well as in mice expressing expanded G4C2 repeats (Chew *et al*, 2015; O'Rourke *et al*, 2015; Peters *et al*, 2015; Jiang *et al*, 2016; Liu *et al*, 2016). Overexpression of DPR proteins using artificial ATG codons for their translation initiation leads to neurodegeneration in cell and animal models, notably through

1 Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Illkirch, France
2 Université de Strasbourg, Illkirch, France
3 Centre National de la Recherche Scientifique, UMR7104, Illkirch, France
4 Institut National de la Santé et de la Recherche Médicale, U1258, Illkirch, France
*Corresponding author. Tel: +33 388 653 309; Fax: +33 388 653 201; E-mail: sellier@igbmc.fr
**Corresponding author. Tel: +33 388 653 309; Fax: +33 388 653 201; E-mail: ncharlet@igbmc.fr

© 2020 The Authors. Published under the terms of the CC BY NC ND 4.0 license

alteration of mitochondria (Dafinca *et al*, 2016; Lopez-Gonzalez *et al*, 2016; Choi *et al*, 2019), DNA repair (Walker *et al*, 2017; Lopez-Gonzalez *et al*, 2019), nuclear and nucleolar organization (White *et al*, 2019; Zhang *et al*, 2019), and/or nucleocytoplasmic transport (Kwon *et al*, 2014; May *et al*, 2014; Mizielinska *et al*, 2014; Wen *et al*, 2014; Zhang *et al*, 2015, 2016; Freibaum *et al*, 2015; Jovičić *et al*, 2015; Tao *et al*, 2015; Boeynaems *et al*, 2016; Khosravi *et al*, 2017). Third, expanded G4C2 repeats promote DNA epigenetic changes that lead to decreased expression of C9ORF72 mRNA and protein levels in C9-ALS/FTD individuals (DeJesus-Hernandez *et al*, 2011; Gijselinck *et al*, 2012; Almeida *et al*, 2013; Waite *et al*, 2014; van Blitterswijk *et al*, 2015; Xiao *et al*, 2015; Saberi *et al*, 2017; Frick *et al*, 2018; Viodé *et al*, 2018). Interestingly, knockdown of C9ORF72 expression in zebrafish or in human motor neuron differentiated from iPS cells leads to some neuronal cell dysfunctions and cell death (Ciura *et al*, 2013; Shi *et al*, 2018). These results suggest that haploinsufficiency of C9ORF72 may play a role in the pathogenesis of ALS/FTD. However, the absence of neurodegeneration in mice knockout for C9orf72 argues against a sole loss of function of C9ORF72 as a cause of ALS/FTD (Lagier-Tourenne *et al*, 2013; Koppers *et al*, 2015; O'Rourke *et al*, 2016; Atanasio *et al*, 2016; Burberry *et al*, 2016; Jiang *et al*, 2016; Sudria-Lopez *et al*, 2016). Hence, the pathological consequences of C9ORF72 reduced expression in ALS/FTD remain to be clarified. Similarly, by which molecular mechanisms sense G4C2 and antisense C4G2 repeats are translated, and how such translation modulates repeat toxicity are key points that remain to be investigated.

Using constructs in which expanded repeats were embedded within the natural *C9ORF72* human sequence, we found that the G4C2 sense repeats are mostly translated into poly(glycine–alanine) (polyGA) and to a lesser extent into poly(glycine–arginine) (polyGR)-containing proteins through initiation to near-cognate start codons that are located upstream of the G4C2 repeats. Near-cognate initiation codons are codons differing from the AUG sequence by one base (CUG, GUG, UUG, ACG, AAG, etc.), but that can still initiate translation provided they are embedded in a correct Kozak consensus sequence (Kozak, 1989; Peabody, 1989). Moreover, we found that the antisense *C9ORF72* transcript containing expanded C4G2 repeats is mostly translated into a poly(proline–glycine) (polyPG)-containing protein using a canonical AUG start codon embedded in a poor Kozak consensus sequence. As a result of their non-optimal translation initiation, DPR proteins are expressed at low levels and consequently present limited toxicity in neuronal cells. Furthermore, accumulation of DPR proteins is prevented by their constant degradation through autophagy. As the C9ORF72 protein regulates autophagy (Amick *et al*, 2016; Sellier *et al*, 2016; Sullivan *et al*, 2016; Ugolino *et al*, 2016; Webster *et al*, 2016; Xiao *et al*, 2016; Yang *et al*, 2016; Jung *et al*, 2017; Chitiprolu *et al*, 2018; Ho *et al*, 2019), we tested whether C9ORF72 modulates DPR expression. Importantly, reduced expression of C9ORF72 increases the accumulation and toxicity of DPR proteins expressed under their natural sequences. Finally, we screened various pharmacological compounds known to activate autophagy and found that mTOR inhibitors and some phenothiazine derivatives, which are clinically approved antipsychotic drugs known to activate autophagy in neuronal cell culture (Tsvetkov *et al*, 2010), prevent the neuronal cell death caused by the synergic toxicity of DPR protein accumulation upon C9ORF72 reduced expression.

In conclusion, these results suggest an alternative molecular mechanism to RAN translation that may explain production of DPR proteins from *C9ORF72* sense and antisense repeats. Furthermore, these data support a double-hit mechanism in ALS/FTD, where the reduced expression of C9ORF72 synergizes the accumulation and toxicity of DPR proteins.

Results

Expanded sense G4C2 repeats are translated into polyGA and polyGR

To assess translation of the sense GGGGCC repeats, we cloned 80 G4C2 repeats embedded within the natural human *C9ORF72* sequence and fused these repeats to the GFP sequence in all three possible frames (Fig EV1A). These frames were named according to the DPR protein potentially encoded by the expanded G4C2 repeats, namely glycine–alanine (GA), glycine–proline (GP), and glycine–arginine (GR). Cell transfection and immunoblotting against the GFP indicated that expanded G4C2 repeats embedded in the natural *C9ORF72* sequence are predominantly translated into the GA frame and with a much lower efficiency in the GR frame (Fig 1A). In contrast, we found no or very limited translation of the G4C2 repeats into the GP frame. Direct observation of the GFP fluorescence confirmed marginal levels of polyPG (Fig EV1B), excluding a potential bias of aggregation and insolubility that would impair detection of this protein by immunoblotting. Furthermore, RT–qPCR quantification indicates similar levels of RNA expression for polyGA, polyGP, and polyGR constructs, indicating that the absence of polyGP at the protein level is not caused by a lack of its expression at the RNA level (Fig EV1C). As a further control, cell transfections of GFP-tagged polyGA, polyGP, and polyGR expressed under artificial ATG start codons demonstrate expression of all three DPR proteins by fluorescence (Fig EV1D) and immunoblot analyses (Fig EV1E). Expression of polyGP into a stable protein when using an artificial ATG start codon suggests that its absence when using the natural *C9ORF72* sequence is not due to a bias of insolubility and/or rapid degradation of this protein.

Next, we characterized the translation initiation sites of the polyGA and polyGR proteins. Expanded G4C2 repeats embedded in their natural *C9ORF72* sequence were modified so that a lysine, target of the LysC enzyme, is present downstream of three G4C2 repeats in the GA frame. Immunoprecipitation of the polyGA DPR protein followed by LysC digestion and LC-MS/MS analysis revealed a translation initiation to a CUG near-cognate codon embedded in a correct Kozak consensus sequence and located 24 nucleotides upstream of the G4C2 repeats (Fig 1B). This CUG start site of polyGA is identical to the one found in three other independent studies (Green *et al*, 2017; Sonobe *et al*, 2018; Tabet *et al*, 2018). Mass spectrometry indicates that the initial amino acid of the polyGA DPR protein is a methionine, suggesting that the CUG codon is decoded by an initiator Met-tRNA despite imperfect match. Mutation of this CTG codon into CTT abolished polyGA DPR protein expression (Fig EV1F), confirming the importance of this near-cognate codon for translation of the G4C2 repeats in the GA frame. In contrast, mutation of the CTG codon into a canonical

 © 2020 The Authors

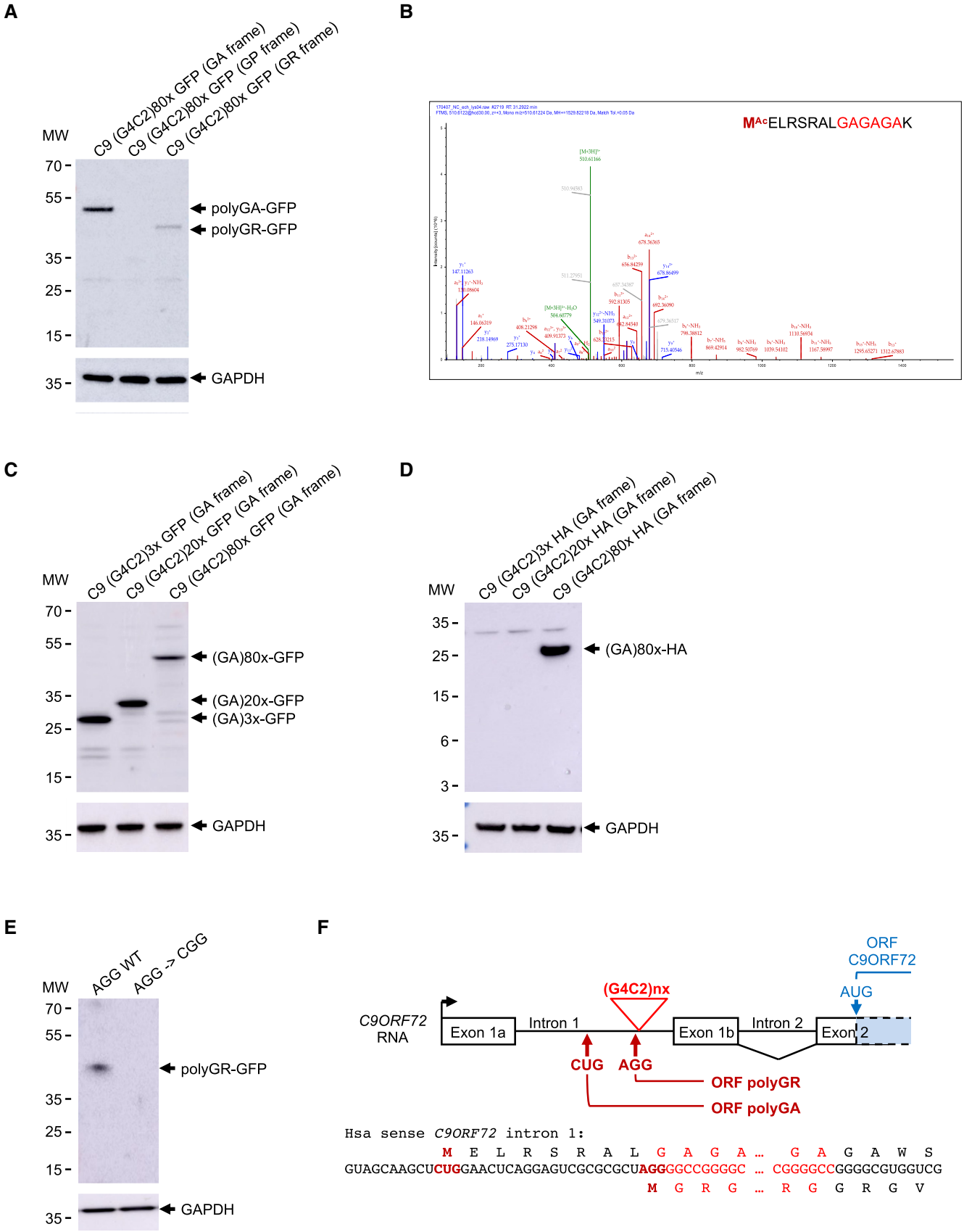

A
C9 (G4C2)80x GFP (GA frame)
C9 (G4C2)80x GFP (GP frame)
C9 (G4C2)80x GFP (GR frame)
MW
70
55
35
25
15
35
polyGA-GFP
polyGR-GFP
GAPDH
B
MAcELRSRALGAGAGAK
C
C9 (G4C2)3x GFP (GA frame)
C9 (G4C2)20x GFP (GA frame)
C9 (G4C2)80x GFP (GA frame)
MW
70
55
35
25
15
35
(GA)80x-GFP
(GA)20x-GFP
(GA)3x-GFP
GAPDH
D
C9 (G4C2)3x HA (GA frame)
C9 (G4C2)20x HA (GA frame)
C9 (G4C2)80x HA (GA frame)
MW
35
25
15
6
3
35
(GA)80x-HA
GAPDH
E
AGG WT
AGG -> CGG
MW
70
55
35
25
15
35
polyGR-GFP
GAPDH
F
ORF
C9ORF72
AUG
(G4C2)nx
C9ORF72
RNA
Exon 1a
Intron 1
Exon 1b
Intron 2
Exon 2
CUG
AGG
ORF polyGR
ORF polyGA
Hsa sense C9ORF72 intron 1:
M E L R S R A L G A G A ... G A G A W S
GUAGCAAGCUCUGGAACUCAGGAGUCGCGCGCUAGGGGCCGGGGC ... CGGGGCCGGGGCGUGGUCG
M G R G ... R G G R G V

Figure 1.

Figure 1. Expanded G4C2 repeats are translated into polyGA and polyGR.

A Immunoblotting against the GFP or the GAPDH of proteins extracted from HEK293 cells transfected for 24 h with 80 G4C2 repeats embedded in the human sense *C9ORF72* and fused in all three possible frames with the GFP deleted of its ATG.

B LC-MS/MS spectra of the N-terminal part of the GFP-immunoprecipitated and LysC-digested polyGA protein expressed as in (A).

C, D Immunoblotting against the GFP- (C) or the HA-tag (D) or the GAPDH of proteins extracted from HEK293 cells transfected for 24 h with 3, 20 or 80 G4C2 repeats embedded in the human sense *C9ORF72* sequence fused to the GFP in the GA frame.

E Immunoblotting against the GFP or the GAPDH of proteins extracted from HEK293 cells transfected with either a wild-type or a mutant (AGG into CGG) construct containing 80 G4C2 repeats embedded in sense *C9ORF72* fused to the GFP in the GR frame.

F Scheme of human *C9ORF72* sense transcript with intron 1 retained. Expanded G4C2 repeats, CUG and AGG near-cognate initiation codons, and polyGA and polyGR ORFs are indicated in red. C9ORF72 ORF is indicated in blue.

ATG start codon enhanced by fivefold to 10-fold expression of polyGA (Fig EV1G), highlighting the weak translation initiation efficiency of the natural *C9ORF72* sequence. As a control, RT–qPCR quantification shows that mutation of the CTG start codon does not modify polyGA expression at the RNA level (Fig EV1H), Next, cell transfection of various lengths of G4C2 repeats (3×, 20×, and 80×) fused to the GFP in the GA frame indicated that translation occurs with expanded G4C2 repeats of any lengths (Fig 1C). In contrast, fusion of G4C2 repeats to a smaller HA-tag (9 amino acids, ~1 kDa) in the GA frame resulted in detection of polyGA with expanded (80×) G4C2 repeats, but only trace of proteins with 3 or 20 G4C2 repeats (Fig 1D). This is likely caused by the rapid degradation of these small proteins, as inhibition of the proteasome and autophagy degradation pathways allowed detection of the polyGA protein with 20 G4C2 repeats (Fig EV1I). These results suggest that translation initiation occurs at a CUG near-cognate codon in the GA frame independently of the length of the G4C2 repeats. However, in absence of an expansion, the resulting protein is small and unstable, impairing its detection in normal conditions. In contrast, repeat expansion or fusion of the repeat to a large tag, such as the GFP protein, results in expression of a stable and detectable polyGA DPR protein.

Translation initiation of polyGA to a near-cognate start codon questions whether a similar mechanism may occur for polyGP and polyGR. However, in the GP frame a stop codon is located just before the G4C2 repeats, preventing expression of a polyGP protein from any upstream putative near-cognate initiation codons. Concerning the polyGR DPR protein, mass spectrometry failed to identify its N-terminus, but its expression is likely to be dependent of a near-cognate initiation mechanism as mutation of an AGG near-cognate codon located just upstream of the G4C2 repeats and in the GR frame impairs expression of this protein (Fig 1E). As a control, mutation of this AGG codon does not affect RNA expression (Fig EV1J). This AGG codon is located just before the repeats, leading to a methionine adjacent to the GR repeats, which may explain why we failed to identify it by mass spectrometry. Of interest, translation initiation at this AGG near-cognate codon is weak and results in limited expression of polyGR compared to polyGA (Fig 1A). To summarize this first part, translation initiation at either a CUG or an AGG near-cognate codon is predicted to result in a polyGA or a polyGR DPR protein, which is composed of a short N-terminus of either 8 or 1 amino acids, followed by a central repeated glycine–alanine or glycine–arginine stretch which length corresponds to the number of expanded G4C2 repeats and a C-terminus ending in either *C9ORF72* intron 2 or exon 2 according to splicing or retention of intron 2 (Fig 1F).

Expanded antisense C4G2 repeats are translated into polyPG

The *C9ORF72* gene is transcribed in both sense and antisense directions. To assess the translation of the antisense CCCCGG repeats, we cloned one hundred C4G2 repeats, embedded within the natural human *C9ORF72* antisense sequence, and fused to a HA-tag in all three possible frames, namely proline–alanine (PA), proline–glycine (PG), and proline–arginine (PR). Cell transfection and immunoblotting against the HA-tag indicated that expanded C4G2 repeats embedded in the antisense *C9ORF72* sequence are predominantly translated into the PG frame (Fig 2A). Fusion of the C4G2 repeats in all three frames to the GFP followed by cell transfection and direct observation of the GFP fluorescence confirmed that translation occurs mostly in the PG frame (Fig EV2A). As a control, RT–qPCR analysis indicates similar RNA levels of polyPA, polyPG, and polyPR (Fig EV2B). Furthermore, polyPA and polyPR proteins are correctly expressed when cloned downstream of artificial ATG start codons (Fig EV2C and D), excluding that the non-detection of these proteins is caused by a potential bias of protein insolubility and/or instability.

As C4G2 repeats appear to be translated mostly into polyPG, we next investigated the translation initiation mechanism of this protein. HA immunoprecipitation followed by LysC digestion and LC-MS/MS analysis revealed translation initiation to a canonical AUG start codon located 195 nucleotides upstream of the repeats (Fig 2B). Deletion of the *C9ORF72* antisense sequence containing this initiation codon abolished polyPG expression (Fig EV2E), demonstrating the importance of this initiation sequence for polyPG translation. As a control, RT–qPCR indicated that deletion of the sequence containing this ATG does not alter RNA expression (Fig EV2F). Of interest, this AUG start codon is located in a poor Kozak sequence and its mutation into a consensus Kozak sequence enhanced expression of the polyPG protein by threefold to fivefold (Fig EV2G). Consistent with translation starting before the repeats, expression of polyPG fused to the GFP is independent of the C4G2 expansion size (Fig 2C). In contrast, fusion of expanded (100×) or non-expanded (10×) C4G2 repeats to the small HA-tag resulted in a detectable polyPG protein only with 100 G4C2 repeats (Fig 2D). The absence of detection of the small polyPG protein with 10 C4G2 repeats fused to the HA-tag is likely due to its instability and rapid turnover as inhibition of the proteasome and autophagy degradation pathways allowed its accumulation and detection (Fig EV2H). These data indicate that the polyPG DPR protein is translated through initiation to a canonical AUG start codon, independently of the repeats. However, in the absence of a C4G2 expansion, the resulting polyPG protein is small and likely too unstable to be significantly observed in control conditions. In contrast, either expansion of the repeats or

 © 2020 The Authors

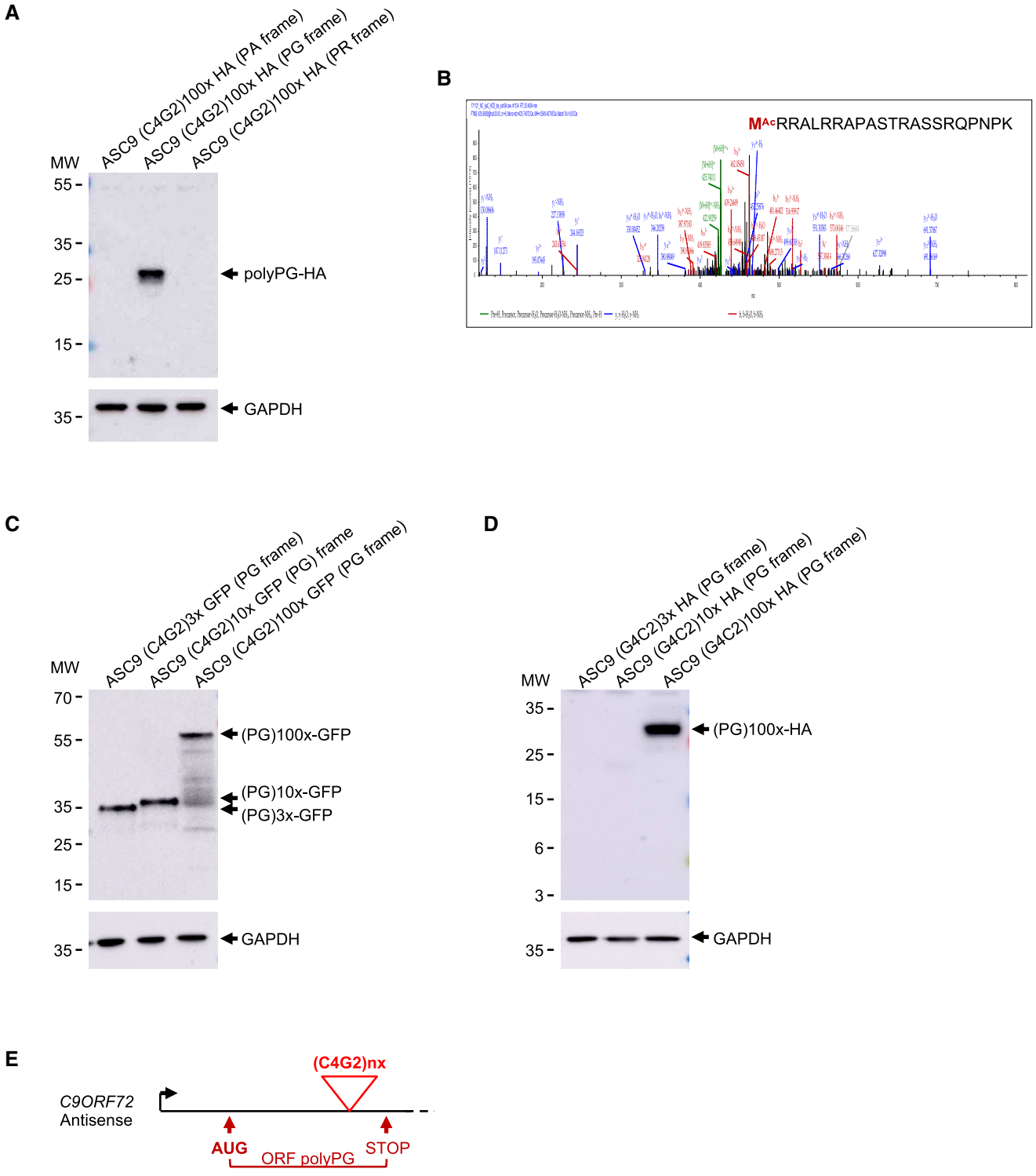

A
ASC9 (C4G2)100x HA (PA frame)
ASC9 (C4G2)100x HA (PG frame)
ASC9 (C4G2)100x HA (PR frame)
MW
55
35
25
15
35
polyPG-HA
GAPDH
B
MAcRRALRRAPASTRASSRQPNPK
C
ASC9 (C4G2)3x GFP (PG frame)
ASC9 (C4G2)10x GFP (PG) frame
ASC9 (C4G2)100x GFP (PG frame)
MW
70
55
35
25
15
35
(PG)100x-GFP
(PG)10x-GFP
(PG)3x-GFP
GAPDH
D
ASC9 (G4C2)3x HA (PG frame)
ASC9 (G4C2)10x HA (PG frame)
ASC9 (G4C2)100x HA (PG frame)
MW
35
25
15
6
3
35
(PG)100x-HA
GAPDH
E
(C4G2)nx
C9ORF72
Antisense
AUG
STOP
ORF polyPG
Hsa antisense C9ORF72:
AAAGUAAAAAUGCGUCGAGCUCUGAGGAGAGCCCCCGCUUCUACCCGCGCCUCUUCCCGGCAGCCGAACCCCAAACAGCCACCCGCCAGGAUGCCGCCU
M R R A L R R A P A S T R A S S R Q P N P K Q P P A R M P P
CCUCACUCACCCACUCGCCACCGCCUGCGCCTCCGCCGCCGCGGGCGCAGGCACCGCAACCGCAGCCCCGCCCCGGGCCCGCCCCCGGGCCCGCCCCGA
P H S P T R H R L R L R R R G R R H R N R S P A P G P P P G P P R
CCACGCCCCGGCCCCGGC … CCCCGGCCCCUAGCGCGCGACUCCUGA…
P R P G P G … P G P .

Figure 2.

Figure 2. Expanded C4G2 repeats are translated into polyPG.

A Immunoblotting against the HA-tag or the GAPDH of proteins extracted from HEK293 cells transfected for 24 h with 100 C4G2 repeats embedded in the human antisense *C9ORF72* and fused in all three possible frames with a HA-tag.

B LC-MS/MS spectra of the N-terminal part of the HA-immunoprecipitated and LysC-digested polyPG protein expressed as in (A).

C, D Immunoblotting against the GFP- (C) or the HA-tag (D) or the GAPDH of proteins extracted from HEK293 cells transfected for 24 h with 3, 10, or 100 C4G2 repeats embedded in the human antisense *C9ORF72* sequence fused to the GFP in the PG frame.

E Scheme of human antisense *C9ORF72* transcript. Expanded C4G2 repeats, AUG initiation codon, and polyPG ORF are indicated in red.

fusion to the GFP protein increases the size and the stability of this protein, enabling its detection. In summary, antisense C4G2 repeats are translated into a polyPG protein, which is predicted to be composed of a N-terminus of 65 amino acids rich in proline and arginine, a central repeated proline–glycine stretch which length corresponds to the number of expanded C4G2 repeats and a stop codon located just after the repeats (Fig 2E).

Decreased expression of C9ORF72 synergizes DPR protein toxicity

Next, we tested the toxicity of the polyGA, polyPG, and polyGR DPR proteins expressed under their natural sense or antisense *C9ORF72* human sequences. Consistent with their low expression, their transfection into neuronal cells resulted in low toxicity with only 10–20% of neuronal cell death (Fig 3A). This toxicity is likely due to translation of the repeats into DPR proteins and not caused by G4C2 or C2G4 RNA toxicity, as mutation or deletion of DPR protein initiation codons (CTG for polyGA, AGG for polyGR, and ATG for polyPG) resulted in negligible cell death (Fig 3A). As positive controls, transfection of polyGA, polyPG, and polyGR cloned downstream of canonical ATG start codons embedded in consensus Kozak sequences (gccATGg) resulted in higher DPR protein expression (Figs EV1C and EV2C) and consequently in higher neuronal cell toxicities, especially for polyGR (Fig 3A). These results are similar to the toxicity reported for DPR proteins driven by artificial ATG start codons and where polyGR are more toxic than polyGA or polyGP (May *et al*, 2014; Mizielinska *et al*, 2014; Yamakawa *et al*, 2015; Lopez-Gonzalez *et al*, 2016, 2019; Schludi *et al*, 2017; Zhang *et al*, 2018; Choi *et al*, 2019). DPR proteins form protein aggregates that are degraded by autophagy (Cristofani *et al*, 2018). Indeed, bafilomycin A1-mediated inhibition of autophagy enhances the accumulation of DPR proteins (Figs 3C and D, and EV3A and B). As the C9ORF72 protein regulates autophagy and as C9ORF72 protein expression is reduced in neuronal tissue from C9-positive ALS/FTD individuals, we tested whether decreased expression of C9ORF72 may modulate the expression and toxicity of DPR proteins. Immunoblotting indicated that siRNA-mediated depletion of C9ORF72 expression leads to an accumulation of polyGA, polyPG, and polyGR DPR proteins expressed under their natural start codons (Fig 3B–D). However, DPR accumulation upon partial decreased expression of C9ORF72 was milder compared to the complete block of autophagy induced by bafilomycin treatment (Fig 3C and D). These results are consistent with C9ORF72 decreased expression causing only a partial alteration of autophagy (Sellier *et al*, 2016). In that aspect, siRNA-mediated depletion of C9ORF72 impaired the recruitment of the P62/SQSTM1 autophagy adaptor protein to polyGA aggregates (Fig EV3C). As a control, decreased expression of C9ORF72 does not modify co-localization of ubiquitin with polyGA inclusions (Fig EV3D). These data suggest that C9ORF72 plays an early role in autophagy at the step of autophagosome formation. Furthermore, siRNA-mediated depletion of C9ORF72 leads to increased lysosomal labeling (Fig EV3E). As a control, endosome labeling is not altered by decreased expression of C9ORF72 (Fig EV3F). These observations are consistent with recent reports showing a role of the C9ORF72/SMCR8 complex at lysosomes (Amick *et al*, 2016; Sullivan *et al*, 2016; Corrionero & Horvitz, 2018; Lan *et al*, 2019), and suggest a complex function of C9ORF72 both at early and at late steps of the autophagic process. As decreased expression of C9ORF72 promotes DPR protein accumulation, we then tested whether C9ORF72 may modulate DPR toxicity. Importantly, reduced expression of C9ORF72 substantially increased the neuronal cell death induced by polyGA, polyPG, and polyGR DPR proteins expressed under their natural start codons (Fig 3E). Of technical interest, toxicity induced by expression of polyGA, polyPG, and polyGR proteins expressed under artificial ATG start codons was less sensitive to C9ORF72 loss, probably due to the already high expression and toxicity of these ATG-driven DPR proteins (Fig 3E). As negative controls, no toxicity was observed when the number of G4C2 repeats was reduced to control size or when DPR protein start codons (CUG for polyGA, AGG for polyGR, or AUG for polyPG) were deleted (Fig 3E). These controls indicate that G4C2 and C4G2 expanded repeats are likely not overly toxic at the RNA level. Furthermore, these results suggest a double-hit mechanism where the decreased expression of C9ORF72 synergizes the toxicity of DPR proteins expressed under their natural start codons.

Drugs activating autophagy reduce DPR protein accumulation and toxicity

Autophagy is a highly regulated cellular process amendable to drug treatment (Rubinsztein *et al*, 2012; Galluzzi *et al*, 2017). Thus, we tested whether drugs known to activate autophagy may bypass C9ORF72 decreased expression and correct DPR protein aggregation and toxicity. The compounds tested included Torin1, KU0063794, and rapamycin, which are well-characterized mTOR inhibitors and thus activators of autophagy, and various phenothiazine derivatives, which are clinically approved antipsychotic drugs known to activate autophagy in neuronal cell culture (Tsvetkov *et al*, 2010). Interestingly, addition of not only Torin1, but also to a lesser extent either promethazine, chlorpromazine, or fluphenazine, decreases the accumulation of polyGA and polyPG DPR proteins expressed under their natural start codons (Fig 4A and B). These compounds can also decrease accumulation of a polyGA-GFP-tagged protein expressed under an artificial ATG start codon (Fig EV4A and B). Direct observation of the GFP fluorescence confirmed that these compounds decrease the accumulation of polyGA protein aggregates (Fig EV4C). Next, we tested

 © 2020 The Authors

whether boosting autophagy may correct the toxicity of DPR protein expression. As inhibiting mTOR has many adverse consequences in animals (Zhang *et al*, 2011), we focused on phenothiazine derivatives, which are clinically approved drugs. Importantly, addition of promethazine alleviated neuronal cell death caused by the synergic toxicity caused by DPR protein accumulation upon C9ORF72 reduced expression (Fig 4C). Moreover, promethazine was also able to reduce cell death caused by expression of polyGA under an artificial ATG start codon (Fig EV4D). Overall, these results suggest that modulating autophagy could be of therapeutic interest in ALS/FTD to prevent the toxic accumulation of DPR proteins.

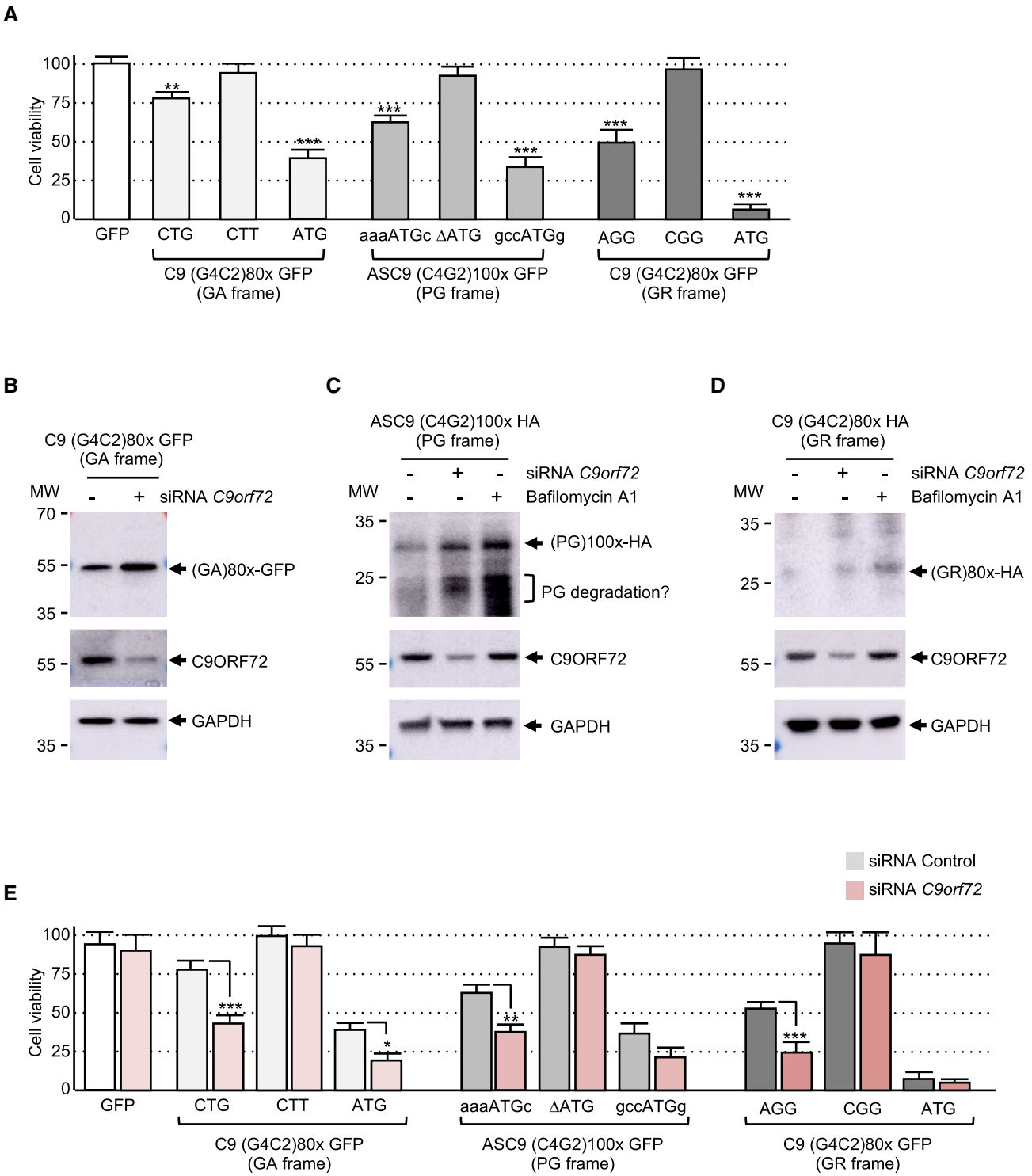


Figure 3.

Figure 3. Decreased expression of C9ORF72 synergizes DPR protein toxicity.

A Cell viability (TO-PRO-3 FACS staining) of GT1-7 neuronal cells transfected for 24 h with either wild-type or indicated mutant constructs containing either 80 G4C2 repeats or 100 C4G2 repeats embedded in sense or antisense *C9ORF72* fused to the GFP in the GA, PG, or GR frame.

B Immunoblotting against the GFP, endogenous C9ORF72, or the GAPDH of proteins extracted from Neuro2A cells transfected for 24 h with 80 G4C2 repeats embedded in the human sense *C9ORF72* sequence fused to the GFP in the GA frame and with either a control siRNA or a siRNA targeting *C9orf72* mRNA.

C Immunoblotting against the HA-tag, endogenous C9ORF72, or GAPDH of proteins extracted from Neuro2A cells transfected for 24 h with 100 C4G2 repeats embedded in the human antisense *C9ORF72* sequence fused to a HA-tag in the PG frame and with either a control siRNA or a siRNA targeting *C9orf72* mRNA or treated with bafilomycin A1 for 15 h.

D Immunoblotting against the HA-tag, endogenous C9ORF72, or GAPDH of proteins extracted from Neuro2A cells transfected for 24 h with 80 G4C2 repeats embedded in the human sense *C9ORF72* sequence fused to a HA-tag in the GR frame and with either a control siRNA or a siRNA targeting *C9orf72* mRNA or treated with bafilomycin A1 for 15 h.

E Cell viability (TO-PRO-3 FACS staining) of GT1-7 neuronal cells co-transfected for 24 h with either a control siRNA or a siRNA targeting *C9orf72* mRNA and wild-type or mutant constructs containing either 80 G4C2 repeats or 100 C4G2 repeats embedded in sense or antisense *C9ORF72* fused to the GFP in the GA, PG, or GR frame.

Data information: Error bars indicate s.e.m. Student's *t*-test, $*P < 0.05$, $**P < 0.01$, and $***P < 0.001$. $n = 5$ independent transfection.

Discussion

An expansion of G4C2 repeats within the first intron of the *C9ORF72* gene is the main genetic cause of ALS/FTD (DeJesus-Hernandez *et al*, 2011; Renton *et al*, 2011). Sense and antisense transcripts containing these repeats are RAN-translated into toxic DPR proteins (Ash *et al*, 2013; Gendron *et al*, 2013; Mori *et al*, 2013; Zu *et al*, 2013), which is a major pathogenic event in cell and animal models of ALS/FTD (Tran *et al*, 2015; Jiang *et al*, 2016; Liu *et al*, 2016; Moens *et al*, 2018). However, the molecular mechanisms by which expanded sense G4C2 and antisense C4G2 repeats are translated remain to be fully characterized.

Using repeats embedded in their natural human sequences, we found that expanded antisense C4G2 repeats are mostly translated into polyPG through initiation to a canonical AUG start codon located upstream of the repeats. In contrast, sense G4C2 repeats are translated mostly into polyGA and to a lesser extent into polyGR, through initiation to a CUG and an AGG near-cognate codon, respectively. In support of these results, translation initiation of polyGA from the exact same CUG near-cognate codon was independently identified in three other studies (Green *et al*, 2017; Sonobe *et al*, 2018; Tabet *et al*, 2018), and deletion of the sequence containing this CUG codon abolishes polyGA expression in neurons differentiated from C9-iPS cells with ~1,000 G4C2 repeats (Almeida *et al*, 2019). Of interest, expanded CGG repeats in the neurodegenerative disease fragile X-associated tremor/ataxia syndrome (FXTAS) are also translated into a toxic protein (FMRpolyGlycine) through initiation to a near-cognate start codon located upstream of the repeats (Todd *et al*, 2013; Kearse *et al*, 2016; Sellier *et al*, 2017). Near-cognate initiation codons are codons differing from the AUG sequence by one base, but that can still initiate translation provided they are embedded in a correct Kozak sequence (Kozak, 1989; Peabody, 1989). Ribosome profiling demonstrated that initiation to near-cognate codons is more common than previously thought and may occur in up to ~30% of mammalian genes (Ingolia *et al*, 2011; Fritsch *et al*, 2012). These data suggest that translation initiation at near-cognate codons located upstream of the repeats could be a common pathogenic mechanism in microsatellite expansion diseases, a hypothesis that may warrant further investigations (Gao *et al*, 2017). However, translation initiation at near-cognate codons should occur independently of the presence of downstream repeats, but DPR proteins are observed only in carriers of G4C2 repeat expansions and not in control individuals. Of interest, we noted that in the absence of an expansion, translation does indeed occur, but the resulting proteins are small and unstable and consequently hardly detectable. In contrast, repeat expansion increases the size and the stability of these DPR proteins, enabling their detection. As a note of caution, our experiments are based on cell transfection with constructs containing a limited number of repeats. Thus, it remains to test whether larger expansions would employ a different mechanism of translation, notably RAN translation initiating directly within the repeats and/or translation frameshifting. In that aspect, G4C2 repeat translation frameshift from the GA frame, initiated at the CUG near-cognate codon, to the GP and GR frames was recently reported in cell transfection assays (Tabet *et al*, 2018). However, deletion of this CUG codon abolishes polyGA translation, but does not affect polyGP and polyGR expression in C9-iPS cells (Almeida *et al*, 2019). Similarly, we found only trace amounts of polyGP expressed from the sense G4C2 repeats, suggesting a very low rate of frameshifting from the GA frame to the GP frame. However, this could be due to the lower sensitivity of our assay (immunoblotting) compared to the more sensitive nanoluciferase and *in vitro* translation assays used in Tabet *et al* In summary, we found that DPR proteins are produced through a classic mechanism of translation initiation at canonical AUG or near-cognate codons located upstream of the sense and antisense repeats, leading to expression of polyPG, polyGA, and, to a lesser degree, polyGR DPR proteins. Importantly, this expression pattern is similar to the one observed in patient tissues where inclusions of polyGA and polyGP/PG proteins are commonly observed, while fewer aggregates of polyGR are detected (Mackenzie *et al*, 2015; Davidson *et al*, 2016; Lee *et al*, 2017; Saberi *et al*, 2017).

A second important point of the present work is the synergic toxicity between C9ORF72 reduced expression and DPR protein expression (model presented in Fig 5). Indeed, we found that DPR proteins expressed under their natural sequences caused only limited toxicity, which is consistent with their limited expression. Mutation of their near-cognate codons into AUG canonical start codons enhanced both DPR protein expression and toxicity, especially for the polyGR DPR protein, as consistently reported in various recent studies (Lopez-Gonzalez *et al*, 2016, 2019; Zhang *et al*, 2018; Choi *et al*, 2019). Conversely, deletions of these DPR initiation codons abolished neuronal cell death, suggesting that expanded repeats cause little toxicity at the RNA level and that RAN translation initiating within the repeats is not sufficiently efficient to drive toxicity, at least with the limited number of repeats used in the

 © 2020 The Authors

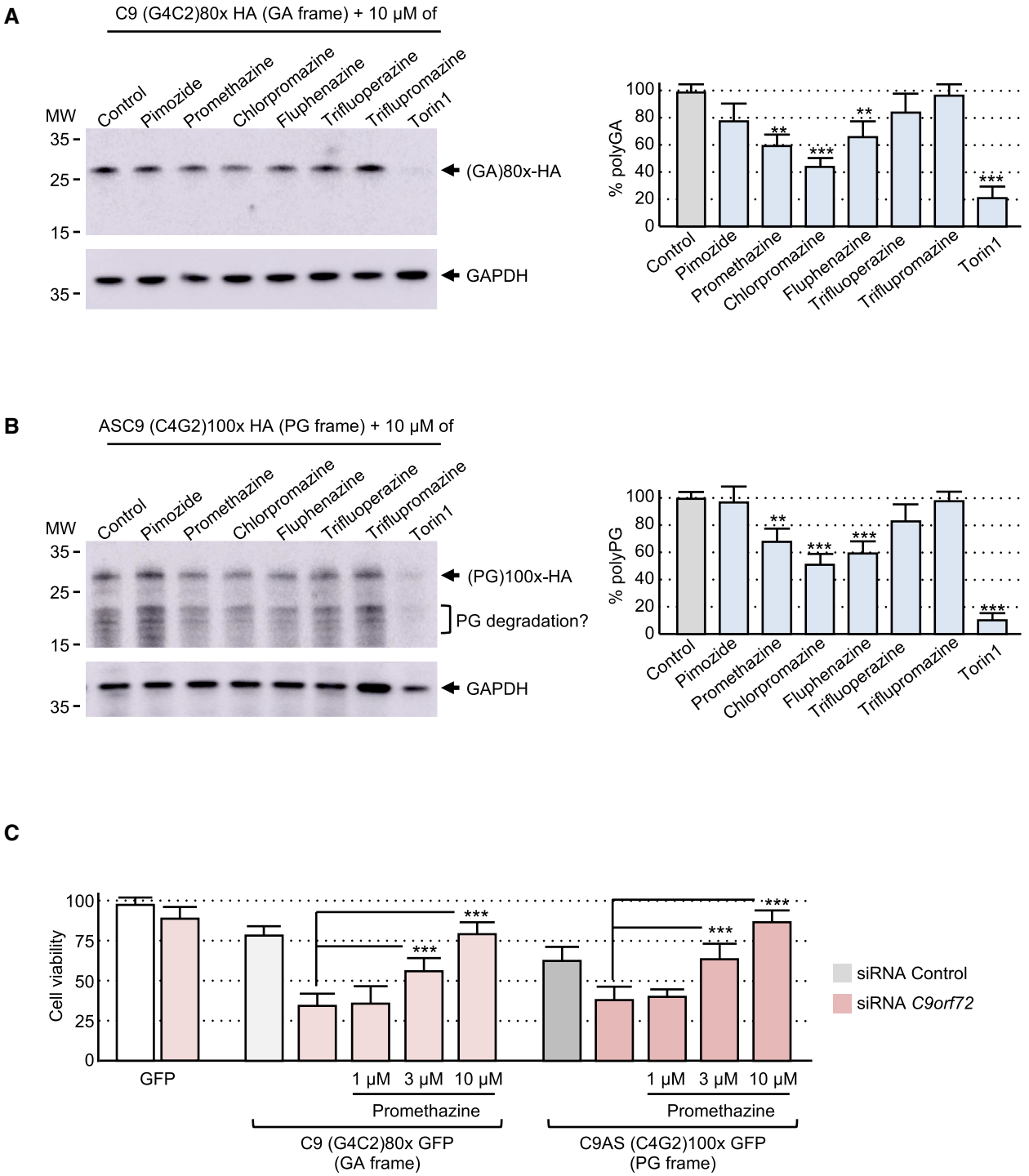


Figure 4. Promethazine reduces DPR protein accumulation and toxicity.

A Left panel, immunoblotting against the HA-tag or the GAPDH of proteins extracted from Neuro2A cells transfected for 24 h with a construct expressing 80 G4C2 repeats embedded in the human sense *C9ORF72* sequence fused to a HA in the GA frame and treated with 10 µM of the indicated compound for 15 h. Right panel, quantification of polyGA expression relative to the GAPDH.

B Left panel, immunoblotting against the HA-tag or the GAPDH of proteins extracted from Neuro2A cells transfected for 24 h with a construct expressing 100 C4G2 repeats embedded in the human antisense *C9ORF72* sequence fused to a HA-tag in the PG frame and treated with 10 µM of the indicated compound for 15 h. Right panel, quantification of polyPG expression relative to the GAPDH.

C Cell viability (TO-PRO-3 FACS staining) of GT1-7 neuronal cells treated with 1, 3, or 10 µM of promethazine and co-transfected for 24 h with either a control siRNA or a siRNA targeting *C9orf72* mRNA and a construct expressing either 80 G4C2 repeats or 100 C4G2 repeats embedded in sense or antisense *C9ORF72* fused to the GFP in the GA or PG frame.

Data information: Error bars indicate s.e.m. Student's *t*-test, $^{**}P < 0.01$, and $^{***}P < 0.001$. $n = 5$ independent transfection.

© 2020 The Authors

present study. Of interest, accumulation and thus toxicity of DPR proteins expressed under their weak natural initiation codons are prevented by their constant degradation, notably by autophagy, a pathway regulated by the C9ORF72 protein. As reduced expression of C9ORF72 leads to suboptimal autophagy and as expression of C9ORF72 is decreased in brain tissues of C9-positive ALS/FTD individuals, we tested whether reduced expression of C9ORF72 may modulate DPR protein accumulation and toxicity. Importantly, decreased expression of C9ORF72 increases the accumulation and thus the toxicity of polyGA, polyGR, and polyPG DPR proteins expressed under their native sequences. These data suggest that the sole reduction of C9ORF72 protein levels or the sole expression of DPR proteins under their native weak initiation codons may not be sufficient to drive pathogenicity in isolation. In contrast, expanded repeats in ALS/FTD may be pathogenic through a double-hit mechanism where the suboptimal autophagy due to C9ORF72 reduced expression may enhance the accumulation and toxicity of DPR proteins expressed under their natural sequences. These results are reminiscent of stress conditions that increase expression and toxicity of DPR proteins (Green *et al*, 2017; Cheng *et al*, 2018; Sonobe *et al*, 2018; Westergard *et al*, 2019). Our results are also consistent with the decreased expression of C9ORF72 that synergizes the toxicity of expanded G4C2 repeats in C9-BAC transgenic mice (Shao *et al*, 2019) or that enhances the toxicity of DPR proteins expressed under artificial ATG start codons in neurons differentiated from C9-iPS cells (Shi *et al*, 2018). Interestingly, a model where both loss of C9ORF72 and expression of DPR proteins are required to observe neuronal cell death is consistent with the absence of neurodegeneration in BAC transgenic mice expressing DPR proteins from their natural sequences but with normal levels of C9orf72 (O'Rourke *et al*, 2016; Peters *et al*, 2015), as well with the absence of a neuronal phenotype in mice knockout for C9orf72 but without DPR protein expression (Lagier-Tourenne *et al*, 2013; Koppers *et al*, 2015; O'Rourke *et al*, 2016; Atanasio *et al*, 2016; Burberry *et al*, 2016; Jiang *et al*, 2016; Sudria-Lopez *et al*, 2016). Furthermore, this double-hit mechanism is supported by the association of *C9ORF72* mRNA decreased levels with patient survival (van Blitterswijk *et al*, 2015), and by the absence of clinical symptoms in individuals with a limited (30 to 70 repeats) number of G4C2 repeats, who show inclusions of DPR proteins but normal levels of the C9ORF72 protein (Gami *et al*, 2015; McGoldrick *et al*, 2018).

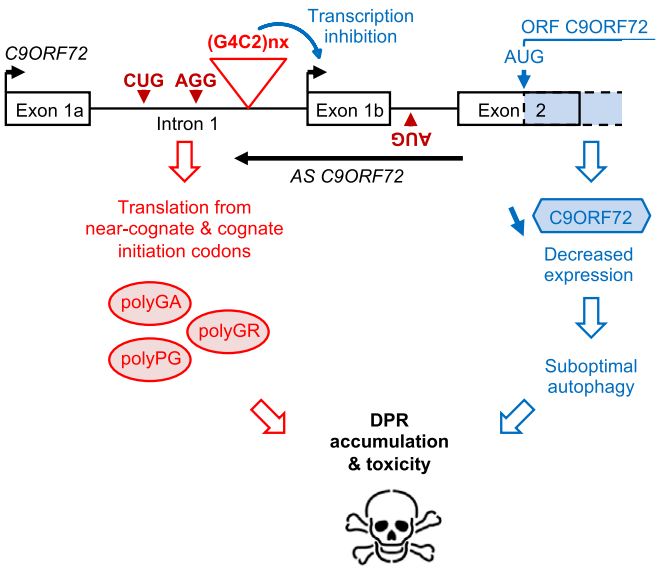


Figure 5. Model of C9ORF72 loss-of-function and DPR gain-of-function toxicity.

Expanded sense G4C2 and antisense C4G2 repeats are translated into polyGA, polyGR, and polyPG DPR proteins through initiation to near-cognate codons or a cognate ATG codon embedded in a poor Kozak sequence. Concomitantly, expanded G4C2 repeats promote epigenetic DNA changes that inhibit promoter 1b activity, ultimately resulting in decreased expression of the C9ORF72 protein. Reduced expression of C9ORF72 leads to suboptimal autophagy that promotes the toxic accumulation of polyGA, polyGR, and polyPG DPR proteins, ultimately resulting in neuronal cell death.

Finally, our study provides an additional proof of concept that modulating autophagy may be of therapeutic interest in ALS/FTD. Indeed, previous studies have shown that stimulating autophagy is beneficial in SOD1, FUS, and TDP-43 cell or animal models of ALS (Hetz *et al*, 2009; Crippa *et al*, 2010; Wang *et al*, 2012; Castillo *et al*, 2013; Barmada *et al*, 2014; Perera *et al*, 2017; Marrone *et al*, 2018, 2019). In agreement with these works, we found that compounds known to induce autophagy, notably promethazine, are able to bypass the reduced expression of C9ORF72 and can promote the clearance of DPR proteins, thus reducing their toxicity. Promethazine is a sedative and antihistamine drug with antipsychotic effects that belongs to a class of phenothiazine derivatives known to activate autophagy in neuronal cell cultures (Tsvetkov *et al*, 2010). Interestingly, these compounds ameliorate neuronal cell dysfunctions and prevent neuronal cell death in SOD1, TDP-43, and FUS cellular and animal models of ALS (Barmada *et al*, 2014; Patten *et al*, 2017; Marrone *et al*, 2018, 2019). Furthermore, pimozide, a phenothiazine derivative related to promethazine, displayed some promising effects in a short randomized clinical trial of sporadic ALS (Patten *et al*, 2017). These results support the need for further investigations of autophagy activators in ALS/FTD.

In conclusion, these results suggest a model where the sole loss of C9ORF72 or the sole expression of DPR proteins under their native sense or antisense sequences is not sufficient to be pathogenic in isolation. In contrast, neuronal cell loss in ALS/FTD may result from a double-hit mechanism where the suboptimal autophagy due to the reduced expression of C9ORF72 may synergize toxicity of other stress, notably the accumulation of toxic DPR proteins translated from their weak natural initiation codons (Fig 5). Of clinical importance, this double-hit mechanism can be prevented by pharmaceutical compounds enhancing autophagy.

Materials and Methods

Constructions

Human sense *C9ORF72* sequence containing 80 G4C2 repeats interrupted by NheI, HindIII, BamHI, and EcoRI restrictions sites was cloned into pcDNA3.1 fused to a HA-tag or the GFP deleted of its ATG in all three frames. Human antisense *C9ORF72* sequence containing 360 nts upstream of 100 C4G2 repeats was cloned into pcDNA3.1 fused to a HA-tag or the GFP deleted of its ATG in all three frames. Mutations of size of the repeats, of initiation codons,

 © 2020 The Authors

or of Kozak sequences were achieved by oligonucleotide ligations. To insure stability of the expanded repeats, all plasmids were transformed into STBL3 bacterial strain (Invitrogen) and grown at room temperature (22°C).

Immunofluorescence

Coverslip were incubated for 15 min in PBS with 4% paraformaldehyde, washed with PBS, and incubated in PBS plus 0.5% Triton X-100 during 10 min. The cells were washed three times with PBS, and the coverslips were incubated during 1 h with primary antibody against ubiquitin (ab134953, Abcam), EEA1 (ab109110, Abcam), M6PR (ab124767, Abcam), and P62/Sqstm1 (ab56416, Abcam). After washing with PBS, the coverslips were incubated with a goat anti-mouse or anti-rabbit secondary antibody conjugated with Alexa 488 or CY3 (Interchim SA) for 1 h, washed twice with PBS, and incubated for 3 min in PBS/DAPI (1/10,000 dilution). Coverslips were rinsed twice before mounting in Pro-Long media (Molecular Probes) and were examined using a Leica microscope.

Cell cultures, transfections, treatments, and viability assay

Neuronal GT1-7 cells were grown in DMEM 4.5 g/l glucose with 10% fetal calf serum, gentamicin, and penicillin at 37°C in 5% CO_2. HEK293 cells were growth in DMEM 1 g/l glucose with 10% FCS and gentamicin at 37°C in 5% CO_2. For transfection, cells were plated in DMEM and 0.1% fetal bovine serum and transfected for 24 h using Lipofectamine 2000 (Fisher Scientific) and/or RNAiMAX (Fisher Scientific) with either control siRNA or siRNA against C9orf72 (ON-TARGETplus, Dharmacon). After 1–3 days, neurons were analyzed by FACS analysis for viability or by immunofluorescence or Western blotting. Cells were treated with 100 nM of bafilomycin A1, 1 μM MG132, or the indicated concentration of drug (Sigma) during 15 h before analysis. For cell viability, cells were detached by scraping and resuspended in PBS. TO-PRO-3 iodide (Fisher Scientific, T-3605) was added at 20 nM to each sample and gently mixed just prior to analysis, and 30,000 cells were FACS-analyzed.

Western blotting

Proteins were denatured 3 min at 95°C, separated on 4–12% Bis-Tris Gel (NuPAGE), transferred on nitrocellulose membranes (Whatman Protran), blocked with 5% non-fat dry milk in Tris-buffered saline (TBS) buffer, incubated with anti-GFP (ab290, Abcam), HA (ab130275, Abcam), C9ORF72 (3H10, homemade), and GAPDH (ab125247, Abcam) in TBS plus 5% non-fat dry milk, washed three times, and incubated with anti-rabbit or mouse peroxidase antibody (1:10,000, Cell Signaling) 1 h in TBS, followed by washing and ECL chemiluminescence revelation (Amersham ECL Prime).

Mass spectrometry analysis

HEK293 cells were transfected with HA- or GFP-tagged plasmid using Lipofectamine 2000 (Fisher Scientific) for 24 h. Proteins were purified by immunoprecipitation against the HA or GFP, separated on 4–12% Bis-Tris Gel (NuPAGE), and visualized by Coomassie blue staining. Gel bands were excised and subjected to in-gel reduction in 10 mM DTT in 100 mM NH_4HCO_3 (Sigma-Aldrich) for 1 h at 57°C, alkylated for 45 min in the dark with 55 mM iodoacetamide in 100 mM NH_4HCO_3 (Sigma-Aldrich), washed in 25 mM NH_4HCO_3, dehydrated with acetonitrile, and dried in SpeedVac 5301 Concentrator (Eppendorf). Then, the gel pieces were rehydrated with 12.5 ng/μl LysC solution (Promega) in 50 mM NH4HCO3 and incubated overnight at 37°C. The peptides were extracted twice with acetonitrile/water/formic acid—45/45/10, v/v/v followed by a final extraction with acetonitrile/formic acid (FA)—95/05, v/v. Extracted peptides were then analyzed using an Ultimate 3000 nano-RSLC (Thermo Scientific) coupled in line with an Orbitrap ELITE (Thermo Scientific). Peptides were separated on a C18 nanocolumn with a linear gradient of acetonitrile and analyzed within a Top 20 collision-induced dissociation data-dependent mass spectrometry with an inclusion list. Data were processed by database searching using SequestHT (Thermo Fisher Scientific) with Proteome Discoverer 1.4 software (Thermo Fisher Scientific) against a homemade database of all potential three frames translated proteins or peptides from the human *C9ORF72* sense and antisense sequences. Precursor and fragment mass tolerance were set at 7 ppm and 0.5 Da, respectively. Oxidation (M) and N-terminal acetylation were set as variable modification, and carbamidomethylation (C), as fixed modification. Peptides were filtered with the fixed value node of Proteome Discoverer 1.4.

RT–qPCR

RT–qPCR analyses were performed using the Transcriptor Reverse Transcriptase Kit (Roche) and the QuantiTect SYBR Green PCR Kit (Qiagen) in a LightCycler 480 (Roche) with 15 min at 94°C followed by 50 cycles of 15 s at 94°C, 20 s at 58°C, and 20 s at 72°C using aagttcatctgcaccaccg (fwd) and ttctgctggtagtggtcggcg (rev) primers to quantify the GFP expression, with *RPLP0* mRNA as standard (fwd: gaagtcactgtgccagccca, rev: gaaggtgtaatccgtctcca). Data were analyzed using LightCycler 480 analysis software and the $2\Delta C_t$ method.

Statistical analysis

All cell experiments are represented as average ± standard error of mean (s.e.m.) with significance determined using Student's *t*-test.

Data availability

The data that support the findings of this study are available within the article and its supplementary information files or from the corresponding authors on reasonable request.

Expanded View for this article is available online.

Acknowledgments

We thank Prof. Mellon (UCSD, USA) for the gift of the GT1-7 cells and Prof. Suzuki for the gift of the ATG DPR plasmids (Yamakawa *et al*, 2015). This work was supported by Fondation de France Thierry Latran #57486 "Model-ALS" and #88187 "DRUG-C9DPR" (NCB), AFM #18605 "Role of C9ORF72 in ALS" (NCB); ARSLA "Therapeutic approach for Amyotrophic Lateral Sclerosis" (NCB); ERC-2012-StG #310659 "RNA DISEASES" (NCB); ANR-17-CE16-0001-01

© 2020 The Authors

"C9ORF72 ALS" (NCB); ANR-18-CE16-0019 "NewMutALS" (NCB); ANR-10-LABX-0030-INRT (IGBMC) and ANR-10-IDEX-0002-02 (IGBMC).

Author contributions

Experiments were performed by MB, VP, AG, FR, and CS. Data were collected and analyzed by LN, NC-B, and CS. The study was designed, coordinated, and written by CS and NC-B.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Almeida S, Gascon E, Tran H, Chou HJ, Gendron TF, Degroot S, Tapper AR, Sellier C, Charlet-Berguerand N, Karydas A *et al* (2013) Modeling key pathological features of frontotemporal dementia with C9ORF72 repeat expansion in iPSC-derived human neurons. *Acta Neuropathol* 126: 385 – 399

Almeida S, Krishnan G, Rushe M, Gu Y, Kankel MW, Gao FB (2019) Production of poly(GA) in C9ORF72 patient motor neurons derived from induced pluripotent stem cells. *Acta Neuropathol* 138: 1099 – 1101

Amick J, Roczniak-Ferguson A, Ferguson SM (2016) C9orf72 binds SMCR8, localizes to lysosomes, and regulates mTORC1 signaling. *Mol Biol Cell* 27: 3040 – 3051

Ash PE, Bieniek KF, Gendron TF, Caulfield T, Lin WL, van Blitterswijk MM, Jansen-West K, Paul JW III, Rademakers R, Boylan KB *et al* (2013) Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron* 77: 639 – 646

Atanasio A, Decman V, White D, Ramos M, Ikiz B, Lee HC, Siao CJ, Brydges S, LaRosa E, Bai Y *et al* (2016) C9orf72 ablation causes immune dysregulation characterized by leukocyte expansion, autoantibody production, and glomerulonephropathy in mice. *Sci Rep* 16: 23204

Barmada SJ, Serio A, Arjun A, Bilican B, Daub A, Ando DM, Tsvetkov A, Pleiss M, Li X, Peisach D *et al* (2014) Autophagy induction enhances TDP43 turnover and survival in neuronal ALS models. *Nat Chem Biol* 10: 677 – 685

van Blitterswijk M, Gendron TF, Baker MC, DeJesus-Hernandez M, Finch NA, Brown PH, Daughrity LM, Murray ME, Heckman MG, Jiang J *et al* (2015) Novel clinical associations with specific C9ORF72 transcripts in patients with repeat expansions in C9ORF72. *Acta Neuropathol* 130: 863 – 876

Boeynaems S, Bogaert E, Michiels E, Gijselinck I, Sieben A, Jovičić A, De Baets G, Scheveneels W, Steyaert J, Cuijt I *et al* (2016) *Drosophila* screen connects nuclear transport genes to DPR pathology in c9ALS/FTD. *Sci Rep* 12: 20877

Burberry A, Suzuki N, Wang JY, Moccia R, Mordes DA, Stewart MH, Suzuki-Uematsu S, Ghosh S, Singh A, Merkle FT *et al* (2016) Loss-of-function mutations in the C9ORF72 mouse ortholog cause fatal autoimmune disease. *Sci Transl Med* 8: 347ra93

Castillo K, Nassif M, Valenzuela V, Rojas F, Matus S, Mercado G, Court FA, van Zundert B, Hetz C (2013) Trehalose delays the progression of amyotrophic lateral sclerosis by enhancing autophagy in motoneurons. *Autophagy* 9: 1308 – 1320

Cheng W, Wang S, Mestre AA, Fu C, Makarem A, Xian F, Hayes LR, Lopez-Gonzalez R, Drenner K, Jiang J *et al* (2018) C9ORF72 GGGGCC repeat-associated non-AUG translation is upregulated by stress through eIF2α phosphorylation. *Nat Commun* 9: 51

Chew J, Gendron TF, Prudencio M, Sasaguri H, Zhang YJ, Castanedes-Casey M, Lee CW, Jansen-West K, Kurti A, Murray ME *et al* (2015) C9ORF72 repeat expansions in mice cause TDP-43 pathology, neuronal loss, and behavioral deficits. *Science* 348: 1151 – 1154

Chitiprolu M, Jagow C, Tremblay V, Bondy-Chorney E, Paris G, Savard A, Palidwor G, Barry FA, Zinman L, Keith J *et al* (2018) A complex of C9ORF72 and p62 uses arginine methylation to eliminate stress granules by autophagy. *Nat Commun* 9: 2794

Choi SY, Lopez-Gonzalez R, Krishnan G, Phillips HL, Li AN, Seeley WW, Yao WD, Almeida S, Gao FB (2019) C9ORF72-ALS/FTD-associated poly(GR) binds Atp5a1 and compromises mitochondrial function *in vivo*. *Nat Neurosci* 22: 851 – 862

Ciura S, Lattante S, Le Ber I, Latouche M, Tostivint H, Brice A, Kabashi E (2013) Loss of function of C9orf72 causes motor deficits in a zebrafish model of amyotrophic lateral sclerosis. *Ann Neurol* 74: 180 – 187

Conlon EG, Lu L, Sharma A, Yamazaki T, Tang T, Shneider NA, Manley JL (2016) The C9ORF72 GGGGCC expansion forms RNA G-quadruplex inclusions and sequesters hnRNP H to disrupt splicing in ALS brains. *Elife* 13: 5

Conlon EG, Fagegaltier D, Agius P, Davis-Porada J, Gregory J, Hubbard I, Kang K, Kim D, New York Genome Center ALS Consortium, Phatnani H *et al* (2018) Unexpected similarities between C9ORF72 and sporadic forms of ALS/FTD suggest a common disease mechanism. *Elife* 7: e37754

Cooper-Knock J, Walsh MJ, Higginbottom A, Robin Highley J, Dickman MJ, Edbauer D, Ince PG, Wharton SB, Wilson SA, Kirby J *et al* (2014) Sequestration of multiple RNA recognition motif-containing proteins by C9orf72 repeat expansions. *Brain* 137(Pt 7): 2040 – 2051

Corrionero A, Horvitz HR (2018) A C9orf72 ALS/FTD ortholog acts in endolysosomal degradation and lysosomal homeostasis. *Curr Biol* 28: 1522 – 1535.e5

Crippa V, Sau D, Rusmini P, Boncoraglio A, Onesto E, Bolzoni E, Galbiati M, Fontana E, Marino M, Carra S *et al* (2010) The small heat shock protein B8 (HspB8) promotes autophagic removal of misfolded proteins involved in amyotrophic lateral sclerosis (ALS). *Hum Mol Genet* 19: 3440 – 3456

Cristofani R, Crippa V, Vezzoli G, Rusmini P, Galbiati M, Cicardi ME, Meroni M, Ferrari V, Tedesco B, Piccolella M *et al* (2018) The small heat shock protein B8 (HSPB8) efficiently removes aggregating species of dipeptides produced in C9ORF72-related neurodegenerative diseases. *Cell Stress Chaperones* 23: 1 – 12

Dafinca R, Scaber J, Ababneh N, Lalic T, Weir G, Christian H, Vowles J, Douglas AG, Fletcher-Jones A, Browne C *et al* (2016) C9orf72 hexanucleotide expansions are associated with altered endoplasmic reticulum calcium homeostasis and stress granule formation in induced pluripotent stem cell-derived neurons from patients with amyotrophic lateral sclerosis and frontotemporal dementia. *Stem Cells* 34: 2063 – 2078

Davidson Y, Robinson AC, Liu X, Wu D, Troakes C, Rollinson S, Masuda-Suzukake M, Suzuki G, Nonaka T, Shi J *et al* (2016) Neurodegeneration in frontotemporal lobar degeneration and motor neurone disease associated with expansions in C9orf72 is linked to TDP-43 pathology and not associated with aggregated forms of dipeptide repeat proteins. *Neuropathol Appl Neurobiol* 42: 242 – 254

DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford NJ, Nicholson AM, Finch NA, Flynn H, Adamson J *et al* (2011) Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* 72: 245 – 256

Donnelly CJ, Zhang PW, Pham JT, Heusler AR, Mistry NA, Vidensky S, Daley EL, Poth EM, Hoover B, Fines DM *et al* (2013) RNA toxicity from the ALS/FTD C9ORF72 expansion is mitigated by antisense intervention. *Neuron* 80: 415 – 428

 © 2020 The Authors

Freibaum BD, Lu Y, Lopez-Gonzalez R, Kim NC, Almeida S, Lee KH, Badders N, Valentine M, Miller BL, Wong PC *et al* (2015) GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. *Nature* 525: 129 – 133

Frick P, Sellier C, Mackenzie IRA, Cheng CY, Tahraoui-Bories J, Martinat C, Pasterkamp RJ, Prudlo J, Edbauer D, Oulad-Abdelghani M *et al* (2018) Novel antibodies reveal presynaptic localization of C9orf72 protein and reduced protein levels in C9orf72 mutation carriers. *Acta Neuropathol Commun* 6: 72

Fritsch C, Herrmann A, Nothnagel M, Szafranski K, Huse K, Schumann F, Schreiber S, Platzer M, Krawczak M, Hampe J *et al* (2012) Genome-wide search for novel human uORFs and N-terminal protein extensions using ribosomal footprinting. *Genome Res* 22: 2208 – 2218

Galluzzi L, Bravo-San Pedro JM, Levine B, Green DR, Kroemer G (2017) Pharmacological modulation of autophagy: therapeutic potential and persisting obstacles. *Nat Rev Drug Discov* 16: 487 – 511

Gami P, Murray C, Schottlaender L, Bettencourt C, De Pablo Fernandez E, Mudanohwo E, Mizielinska S, Polke JM, Holton JL, Isaacs AM *et al* (2015) A 30-unit hexanucleotide repeat expansion in C9orf72 induces pathological lesions with dipeptide-repeat proteins and RNA foci, but not TDP-43 inclusions and clinical disease. *Acta Neuropathol* 130: 599 – 601

Gao FB, Richter JD, Cleveland DW (2017) Rethinking Unconventional Translation in Neurodegeneration. *Cell* 171: 994 – 1000

Gendron TF, Bieniek KF, Zhang YJ, Jansen-West K, Ash PE, Caulfield T, Daughrity L, Dunmore JH, Castanedes-Casey M, Chew J *et al* (2013) Antisense transcripts of the expanded C9ORF72 hexanucleotide repeat form nuclear RNA foci and undergo repeat-associated non-ATG translation in c9FTD/ALS. *Acta Neuropathol* 126: 829 – 844

Gijselinck I, Van Langenhove T, van der Zee J, Sleegers K, Philtjens S, Kleinberger G, Janssens J, Bettens K, Van Cauwenberghe C, Pereson S *et al* (2012) A C9orf72 promoter repeat expansion in a Flanders-Belgian cohort with disorders of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum: a gene identification study. *Lancet Neurol* 11: 54 – 65

Green KM, Glineburg MR, Kearse MG, Flores BN, Linsalata AE, Fedak SJ, Goldstrohm AC, Barmada SJ, Todd PK (2017) RAN translation at C9orf72-associated repeat expansions is selectively enhanced by the integrated stress response. *Nat Commun* 8: 2005

Haeusler AR, Donnelly CJ, Periz G, Simko EA, Shaw PG, Kim MS, Maragakis NJ, Troncoso JC, Pandey A, Sattler R *et al* (2014) C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* 507: 195 – 200

Hetz C, Thielen P, Matus S, Nassif M, Court F, Kiffin R, Martinez G, Cuervo AM, Brown RH, Glimcher LH (2009) XBP-1 deficiency in the nervous system protects against amyotrophic lateral sclerosis by increasing autophagy. *Genes Dev* 23: 2294 – 2306

Ho WY, Tai YK, Chang JC, Liang J, Tyan SH, Chen S, Guan JL, Zhou H, Shen HM, Koo E *et al* (2019) The ALS-FTD-linked gene product, C9orf72, regulates neuronal morphogenesis via autophagy. *Autophagy* 15: 827 – 842

Ingolia NT, Lareau LF, Weissman JS (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147: 789 – 802

Jiang J, Zhu Q, Gendron TF, Saberi S, McAlonis-Downes M, Seelman A, Stauffer JE, Jafar-Nejad P, Drenner K, Schulte D *et al* (2016) Gain of toxicity from ALS/FTD-linked repeat expansions in C9ORF72 Is alleviated by antisense oligonucleotides targeting GGGGCC-containing RNAs. *Neuron* 90: 535 – 550

Jovičić A, Mertens J, Boeynaems S, Bogaert E, Chai N, Yamada SB, Paul JW III, Sun S, Herdy JR, Bieri G *et al* (2015) Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. *Nat Neurosci* 18: 1226 – 1229

Jung J, Nayak A, Schaeffer V, Starzetz T, Kirsch AK, Müller S, Dikic I, Mittelbronn M, Behrends C (2017) Multiplex image-based autophagy RNAi screening identifies SMCR8 as ULK1 kinase activity and gene expression regulator. *Elife* 6: e23063

Kearse MG, Green KM, Krans A, Rodriguez CM, Linsalata AE, Goldstrohm AC, Todd PK (2016) CGG repeat-associated Non-AUG translation utilizes a cap-dependent scanning mechanism of initiation to produce toxic proteins. *Mol Cell* 62: 314 – 322

Khosravi B, Hartmann H, May S, Möhl C, Ederle H, Michaelsen M, Schludi MH, Dormann D, Edbauer D (2017) Cytoplasmic poly-GA aggregates impair nuclear import of TDP-43 in C9orf72 ALS/FTLD. *Hum Mol Genet* 26: 790 – 800

Koppers M, Blokhuis AM, Westeneng HJ, Terpstra ML, Zundel CA, Vieira de Sá R, Schellevis RD, Waite AJ, Blake DJ, Veldink JH *et al* (2015) C9orf72 ablation in mice does not cause motor neuron degeneration or motor deficits. *Ann Neurol* 78: 426 – 438

Kozak M (1989) Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems. *Mol Cell Biol* 9: 5073 – 5080

Kwon I, Xiang S, Kato M, Wu L, Theodoropoulos P, Wang T, Kim J, Yun J, Xie Y, McKnight SL (2014) Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science* 345: 1139 – 1145

Lagier-Tourenne C, Baughn M, Rigo F, Sun S, Liu P, Li HR, Jiang J, Watt AT, Chun S, Katz M *et al* (2013) Targeted degradation of sense and antisense C9orf72 RNA foci as therapy for ALS and frontotemporal degeneration. *Proc Natl Acad Sci USA* 110: E4530 – E4539

Lan Y, Sullivan PM, Hu F (2019) SMCR8 negatively regulates AKT and MTORC1 signaling to modulate lysosome biogenesis and tissue homeostasis. *Autophagy* 15: 871 – 885

Lee YB, Chen HJ, Peres JN, Gomez-Deza J, Attig J, Stalekar M, Troakes C, Nishimura AL, Scotter EL, Vance C *et al* (2013) Hexanucleotide repeats in ALS/FTD form length-dependent RNA foci, sequester RNA binding proteins, and are neurotoxic. *Cell Rep* 5: 1178 – 1186

Lee YB, Baskaran P, Gomez-Deza J, Chen HJ, Nishimura AL, Smith BN, Troakes C, Adachi Y, Stepto A, Petrucelli L *et al* (2017) C9orf72 poly GA RAN-translated protein plays a key role in amyotrophic lateral sclerosis via aggregation and toxicity. *Hum Mol Genet* 26: 4765 – 4777

Liu Y, Pattamatta A, Zu T, Reid T, Bardhi O, Borchelt DR, Yachnis AT, Ranum LP (2016) C9orf72 BAC mouse model with motor deficits and neurodegenerative features of ALS/FTD. *Neuron* 90: 521 – 534

Lomen-Hoerth C, Anderson T, Miller B (2002) The overlap of amyotrophic lateral sclerosis and frontotemporal dementia. *Neurology* 59: 1077 – 1079

Lopez-Gonzalez R, Lu Y, Gendron TF, Karydas A, Tran H, Yang D, Petrucelli L, Miller BL, Almeida S, Gao FB (2016) Poly(GR) in C9ORF72-Related ALS/FTD compromises mitochondrial function and increases oxidative stress and DNA damage in iPSC-derived motor neurons. *Neuron* 92: 383 – 391

Lopez-Gonzalez R, Yang D, Pribadi M, Kim TS, Krishnan G, Choi SY, Lee S, Coppola G, Gao FB (2019) Partial inhibition of the overactivated Ku80-dependent DNA repair pathway rescues neurodegeneration in C9ORF72-ALS/FTD. *Proc Natl Acad Sci USA* 116: 9628 – 9633

Majounie E, Renton AE, Mok K, Dopper EG, Waite A, Rollinson S, Chiò A, Restagno G, Nicolaou N, Simon-Sanchez J *et al* (2012) Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study. *Lancet Neurol* 11: 323 – 330

Marrone L, Poser I, Casci I, Japtok J, Reinhardt P, Janosch A, Andree C, Lee HO, Moebius C, Koerner E *et al* (2018) Isogenic FUS-eGFP iPSC reporter lines

© 2020 The Authors

enable quantification of FUS stress granule pathology that is rescued by drugs inducing autophagy. *Stem Cell Reports* 10: 375 – 389

Marrone L, Drexler HCA, Wang J, Tripathi P, Distler T, Heisterkamp P, Anderson EN, Kour S, Moraiti A, Maharana S *et al* (2019) FUS pathology in ALS is linked to alterations in multiple ALS-associated proteins and rescued by drugs stimulating autophagy. *Acta Neuropathol* 138: 67 – 84

May S, Hornburg D, Schludi MH, Arzberger T, Rentzsch K, Schwenk BM, Grässer FA, Mori K, Kremmer E, Banzhaf-Strathmann J *et al* (2014) C9orf72 FTLD/ALS-associated Gly-Ala dipeptide repeat proteins cause neuronal toxicity and Unc119 sequestration. *Acta Neuropathol* 128: 485 – 503

Mackenzie IR, Frick P, Grässer FA, Gendron TF, Petrucelli L, Cashman NR, Edbauer D, Kremmer E, Prudlo J, Troost D *et al* (2015) Quantitative analysis and clinico-pathological correlations of different dipeptide repeat protein pathologies in C9ORF72 mutation carriers. *Acta Neuropathol* 130: 845 – 861

McGoldrick P, Zhang M, van Blitterswijk M, Sato C, Moreno D, Xiao S, Zhang AB, McKeever PM, Weichert A, Schneider R *et al* (2018) Unaffected mosaic C9orf72 case: RNA foci, dipeptide proteins, but upregulated C9orf72 expression. *Neurology* 90: e323 – e331

Mizielinska S, Lashley T, Norona FE, Clayton EL, Ridler CE, Fratta P, Isaacs AM (2013) C9orf72 frontotemporal lobar degeneration is characterised by frequent neuronal sense and antisense RNA foci. *Acta Neuropathol* 126: 845 – 857

Mizielinska S, Grönke S, Niccoli T, Ridler CE, Clayton EL, Devoy A, Moens T, Norona FE, Woollacott IO, Pietrzyk J *et al* (2014) C9orf72 repeat expansions cause neurodegeneration in *Drosophila* through arginine-rich proteins. *Science* 345: 1192 – 1194

Moens TG, Mizielinska S, Niccoli T, Mitchell JS, Thoeng A, Ridler CE, Grönke S, Esser J, Heslegrave A, Zetterberg H *et al* (2018) Sense and antisense RNA are not toxic in *Drosophila* models of C9orf72-associated ALS/FTD. *Acta Neuropathol* 135: 445 – 457

Mori K, Weng SM, Arzberger T, May S, Rentzsch K, Kremmer E, Schmid B, Kretzschmar HA, Cruts M, Van Broeckhoven C *et al* (2013) The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. *Science* 339: 1335 – 1338

Neumann M, Sampathu DM, Kwong LK, Truax AC, Micsenyi MC, Chou TT, Bruce J, Schuck T, Grossman M, Clark CM *et al* (2006) Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* 314: 130 – 133

O'Rourke JG, Bogdanik L, Yáñez A, Lall D, Wolf AJ, Muhammad AK, Ho R, Carmona S, Vit JP, Zarrow J *et al* (2016) C9orf72 is required for proper macrophage and microglial function in mice. *Science* 351: 1324 – 1329

Patten SA, Aggad D, Martinez J, Tremblay E, Petrillo J, Armstrong GA, La Fontaine A, Maios C, Liao M, Ciura S *et al* (2017) Neuroleptics as therapeutic compounds stabilizing neuromuscular transmission in amyotrophic lateral sclerosis. *JCI Insight* 2: 97152

Peabody DS (1989) Translation initiation at non-AUG triplets in mammalian cells. *J Biol Chem* 264: 5031 – 5035

Perera ND, Sheean RK, Lau CL, Shin YS, Beart PM, Horne MK, Turner BJ (2017) Rilmenidine promotes MTOR-independent autophagy in the mutant SOD1 mouse model of amyotrophic lateral sclerosis without slowing disease progression. *Autophagy* 5: 1 – 18

Peters OM, Cabrera GT, Tran H, Gendron TF, McKeon JE, Metterville J, Weiss A, Wightman N, Salameh J, Kim J *et al* (2015) Human C9ORF72 hexanucleotide expansion reproduces RNA foci and dipeptide repeat proteins but not neurodegeneration in BAC transgenic mice. *Neuron* 88: 902 – 909

Renton AE, Majounie E, Waite A, Simón-Sánchez J, Rollinson S, Gibbs JR, Schymick JC, Laaksovirta H, van Swieten JC, Myllykangas L *et al* (2011) A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS/FTD. *Neuron* 72: 257 – 268

Ringholz GM, Appel SH, Bradshaw M, Cooke NA, Mosnik DM, Schulz PE (2005) Prevalence and patterns of cognitive impairment in sporadic ALS. *Neurology* 65: 586 – 590

Rubinsztein DC, Codogno P, Levine B (2012) Autophagy modulation as a potential therapeutic target for diverse diseases. *Nat Rev Drug Discov* 11: 709 – 730

Saberi S, Stauffer JE, Jiang J, Garcia SD, Taylor AE, Schulte D, Ohkubo T, Schloffman CL, Maldonado M, Baughn M *et al* (2017) Sense-encoded poly-GR dipeptide repeat proteins correlate to neurodegeneration and uniquely co-localize with TDP-43 in dendrites of repeat-expanded C9orf72 amyotrophic lateral sclerosis. *Acta Neuropathol* 135: 459 – 474

Schludi MH, Becker L, Garrett L, Gendron TF, Zhou Q, Schreiber F, Popper B, Dimou L, Strom TM, Winkelmann J *et al* (2017) Spinal poly-GA inclusions in a C9orf72 mouse model trigger motor deficits and inflammation without neuron loss. *Acta Neuropathol* 134: 241 – 254

Sellier C, Campanari ML, Corbier C, Gaucherot A, Kolb-Cheynel I, Oulad-Abdelghani M, Ruffenach F, Page A, Ciura S, Kabashi E *et al* (2016) Loss of C9ORF72 partly impairs autophagy and synergizes the toxicity of Ataxin-2 with intermediate size of polyglutamine expansion. *EMBO J* 35: 1276 – 1297

Sellier C, Buijsen RA, He F, Natla S, Jung L, Tropel P, Gaucherot A, Jacobs H, Meziane H, Vincent A *et al* (2017) Translation of expanded CGG repeats into FMRpolyG Is pathogenic and may contribute to fragile X tremor ataxia syndrome. *Neuron* 93: 331 – 347

Shao Q, Liang C, Chang Q, Zhang W, Yang M, Chen JF (2019) C9orf72 deficiency promotes motor deficits of a C9ALS/FTD mouse model in a dose-dependent manner. *Acta Neuropathol Commun* 7: 32

Shi Y, Lin S, Staats KA, Li Y, Chang WH, Hung ST, Hendricks E, Linares GR, Wang Y, Son EY *et al* (2018) Haploinsufficiency leads to neurodegeneration in C9ORF72 ALS/FTD human induced motor neurons. *Nat Med* 24: 313 – 325

Sonobe Y, Ghadge G, Masaki K, Sendoel A, Fuchs E, Roos RP (2018) Translation of dipeptide repeat proteins from the C9ORF72 expanded repeat is associated with cellular stress. *Neurobiol Dis* 116: 155 – 165

Sudria-Lopez E, Koppers M, de Wit M, van der Meer C, Westeneng HJ, Zundel CA, Youssef SA, Harkema L, de Bruin A, Veldink JH *et al* (2016) Full ablation of C9orf72 in mice causes immune system-related pathology and neoplastic events but no motor neuron defects. *Acta Neuropathol* 132: 145 – 147

Sullivan PM, Zhou X, Robins AM, Paushter DH, Kim D, Smolka MB, Hu F (2016) The ALS/FTLD associated protein C9orf72 associates with SMCR8 and WDR41 to regulate the autophagy-lysosome pathway. *Acta Neuropathol Commun* 4: 51

Tabet R, Schaeffer L, Freyermuth F, Jambeau M, Workman M, Lee CZ, Lin CC, Jiang J, Jansen-West K, Abou-Hamdan H *et al* (2018) CUG initiation and frameshifting enable production of dipeptide repeat proteins from ALS/FTD C9ORF72 transcripts. *Nat Commun* 9: 152

Tao Z, Wang H, Xia Q, Li K, Li K, Jiang X, Xu G, Wang G, Ying Z (2015) Nucleolar stress and impaired stress granule formation contribute to C9orf72 RAN translation-induced cytotoxicity. *Hum Mol Genet* 24: 2426 – 2441

Todd PK, Oh SY, Krans A, He F, Sellier C, Frazer M, Renoux AJ, Chen KC, Scaglione KM, Basrur V *et al* (2013) CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. *Neuron* 78: 440 – 455

 © 2020 The Authors

Tran H, Almeida S, Moore J, Gendron TF, Chalasani U, Lu Y, Du X, Nickerson JA, Petrucelli L, Weng Z *et al* (2015) Differential toxicity of nuclear RNA foci versus dipeptide repeat proteins in a *Drosophila* model of C9ORF72 FTD/ALS. *Neuron* 87: 1207 – 1214

Tsvetkov AS, Miller J, Arrasate M, Wong JS, Pleiss MA, Finkbeiner S (2010) A small-molecule scaffold induces autophagy in primary neurons and protects against toxicity in a Huntington disease model. *Proc Natl Acad Sci USA* 107: 16982 – 16987

Ugolino J, Ji YJ, Conchina K, Chu J, Nirujogi RS, Pandey A, Brady NR, Hamacher-Brady A, Wang J (2016) Loss of C9orf72 enhances autophagic activity via deregulated mTOR and TFEB signaling. *PLoS Genet* 12: e1006443

Viodé A, Fournier C, Camuzat A, Fenaille F, NeuroCEB Brain Bank, Latouche M, Elahi F, Le Ber I, Junot C, Lamari F *et al* (2018) New antibody-free mass spectrometry-based quantification reveals that C9ORF72 long protein isoform is reduced in the frontal cortex of hexanucleotide-repeat expansion carriers. *Front Neurosci* 12: 589.

Waite AJ, Bäumer D, East S, Neal J, Morris HR, Ansorge O, Blake DJ (2014) Reduced C9orf72 protein levels in frontal cortex of amyotrophic lateral sclerosis and frontotemporal degeneration brain with the C9ORF72 hexanucleotide repeat expansion. *Neurobiol Aging* 35: 1779.e5 – 1779.e13

Walker C, Herranz-Martin S, Karyka E, Liao C, Lewis K, Elsayed W, Lukashchuk V, Chiang SC, Ray S, Mulcahy PJ *et al* (2017) C9orf72 expansion disrupts ATM-mediated chromosomal break repair. *Nat Neurosci* 20: 1225 – 1235

Wang IF, Guo BS, Liu YC, Wu CC, Yang CH, Tsai KJ, Shen CK (2012) Autophagy activators rescue and alleviate pathogenesis of a mouse model with proteinopathies of the TAR DNA-binding protein 43. *Proc Natl Acad Sci USA* 109: 15024 – 15029

Webster CP, Smith EF, Bauer CS, Moller A, Hautbergue GM, Ferraiuolo L, Myszczynska MA, Higginbottom A, Walsh MJ, Whitworth AJ *et al* (2016) The C9orf72 protein interacts with Rab1a and the ULK1 complex to regulate initiation of autophagy. *EMBO J* 35: 1656 – 1676

Wen X, Tan W, Westergard T, Krishnamurthy K, Markandaiah SS, Shi Y, Lin S, Shneider NA, Monaghan J, Pandey UB *et al* (2014) Antisense proline-arginine RAN dipeptides linked to C9ORF72-ALS/FTD form toxic nuclear aggregates that initiate *in vitro* and *in vivo* neuronal death. *Neuron* 84: 1213 – 1225

Westergard T, McAvoy K, Russell K, Wen X, Pang Y, Morris B, Pasinelli P, Trotti D, Haeusler A (2019) Repeat-associated non-AUG translation in C9orf72-ALS/FTD is driven by neuronal excitation and stress. *EMBO Mol Med* 11: e9423

White MR, Mitrea DM, Zhang P, Stanley CB, Cassidy DE, Nourse A, Phillips AH, Tolbert M, Taylor JP, Kriwacki RW (2019) C9orf72 Poly(PR) dipeptide repeats disturb biomolecular phase separation and disrupt nucleolar function. *Mol Cell* 74: 713 – 728.e6

Xiao S, MacNair L, McGoldrick P, McKeever PM, McLean JR, Zhang M, Keith J, Zinman L, Rogaeva E, Robertson J (2015) Isoform-specific antibodies reveal distinct subcellular localizations of C9orf72 in amyotrophic lateral sclerosis. *Ann Neurol* 78: 568 – 583

Xiao S, MacNair L, McLean J, McGoldrick P, McKeever P, Soleimani S, Keith J, Zinman L, Rogaeva E, Robertson J (2016) C9orf72 isoforms in amyotrophic lateral sclerosis and frontotemporal lobar degeneration. *Brain Res* 15: 43 – 49

Yamakawa M, Ito D, Honda T, Kubo K, Noda M, Nakajima K, Suzuki N (2015) Characterization of the dipeptide repeat protein in the molecular pathogenesis of c9FTD/ALS. *Hum Mol Genet* 24: 1630 – 1645

Yang M, Liang C, Swaminathan K, Herrlinger S, Lai F, Shiekhattar R, Chen JF (2016) A C9ORF72/SMCR8-containing complex regulates ULK1 and plays a dual role in autophagy. *Sci Adv* 2: e1601167

Zhang X, Li L, Chen S, Yang D, Wang Y, Zhang X, Wang Z, Le W (2011) Rapamycin treatment augments motor neuron degeneration in SOD1 (G93A) mouse model of amyotrophic lateral sclerosis. *Autophagy* 7: 412 – 425

Zhang K, Donnelly CJ, Haeusler AR, Grima JC, Machamer JB, Steinwald P, Daley EL, Miller SJ, Cunningham KM, Vidensky S *et al* (2015) The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. *Nature* 525: 56 – 61

Zhang YJ, Gendron TF, Grima JC, Sasaguri H, Jansen-West K, Xu YF, Katzman RB, Gass J, Murray ME, Shinohara M *et al* (2016) C9ORF72 poly(GA) aggregates sequester and impair HR23 and nucleocytoplasmic transport proteins. *Nat Neurosci* 19: 668 – 677

Zhang YJ, Gendron TF, Ebbert MTW, O'Raw AD, Yue M, Jansen-West K, Zhang X, Prudencio M, Chew J, Cook CN *et al* (2018) Poly(GR) impairs protein translation and stress granule dynamics in C9orf72-associated frontotemporal dementia and amyotrophic lateral sclerosis. *Nat Med* 24: 1136 – 1142

Zhang YJ, Guo L, Gonzales PK, Gendron TF, Wu Y, Jansen-West K, O'Raw AD, Pickles SR, Prudencio M, Carlomagno Y *et al* (2019) Heterochromatin anomalies and double-stranded RNA accumulation underlie C9orf72 poly (PR) toxicity. *Science* 363: eaav2606

Zu T, Liu Y, Bañez-Coronel M, Reid T, Pletnikova O, Lewis J, Miller TM, Harms MB, Falchook AE, Subramony SH *et al* (2013) RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. *Proc Natl Acad Sci USA* 110: E4968 – E4977

License: This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Expanded View Figures

Figure EV1. Expanded G4C2 repeats are translated into polyGA.

A Scheme and sequence of the human *C9ORF72* sense transcript with 80 G4C2 repeats fused to the eGFP in the three possible frames and cloned into the pcDNA3.1 plasmid.

B, C Fluorescence (B) and RT–qPCR (C) GFP analyses of HEK293 cells transfected for 24 h with 80 G4C2 repeats embedded in the human sense *C9ORF72* sequence and fused in all three possible frames with the GFP deleted of its ATG.

D, E GFP expression (D) and immunoblotting (E) analysis of HEK293 cells transfected for 24 h with constructs expressing either polyGA, polyGP, or polyGR expressed under an artificial ATG start codon and fused to the GFP.

F Immunoblotting against the GFP or the GAPDH of proteins extracted from HEK293 cells transfected for 24 h with either a wild-type or a mutant (CTG into CTT) construct containing 80 G4C2 repeats embedded in sense *C9ORF72* fused to the GFP in the GA frame.

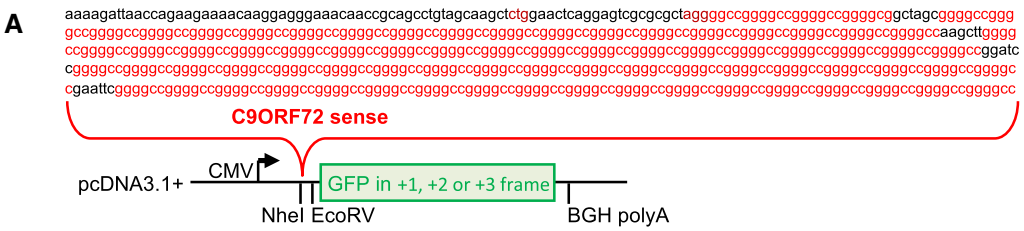
G Immunoblotting against the GFP or the GAPDH of proteins extracted from HEK293 cells transfected for 24 h with either a wild-type or a mutant (CTG into ATG) construct containing 80 G4C2 repeats embedded in sense *C9ORF72* fused to the GFP in the GA frame.

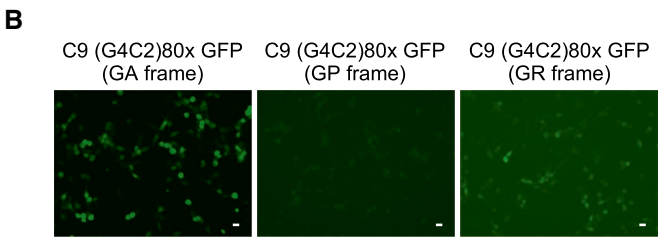
H RT–qPCR analysis of GFP expression of HEK293 cells transfected for 24 h with either wild-type or mutant (CTG into CTT or ATG) constructs containing 80 G4C2 repeats embedded in sense *C9ORF72* fused to the GFP in the GA frame.

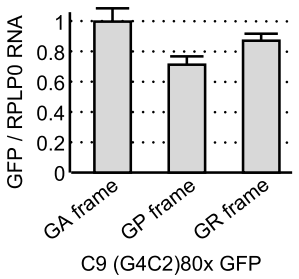
I Immunoblotting against the HA-tag or the GAPDH of proteins extracted from HEK293 cells transfected for 24 h with 20 G4C2 repeats embedded in the human sense *C9ORF72* sequence fused to a HA-tag in the GA frame and treated or not with MG132 and/or bafilomycin A1 for 15 h.

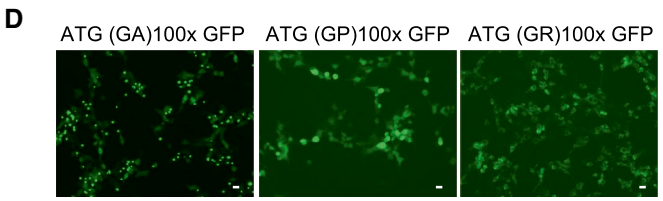
J RT–qPCR analysis of GFP expression of HEK293 cells transfected for 24 h with either wild-type or mutant (AGG into CGG) constructs containing 80 G4C2 repeats embedded in sense *C9ORF72* fused to the GFP in the GR frame.

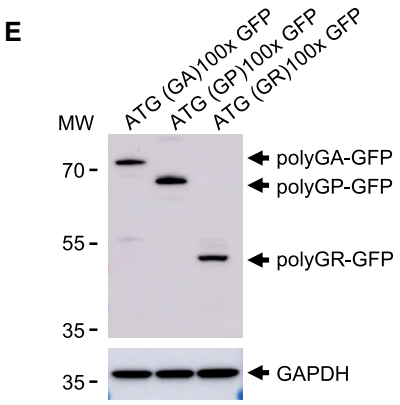
© 2020 The Authors

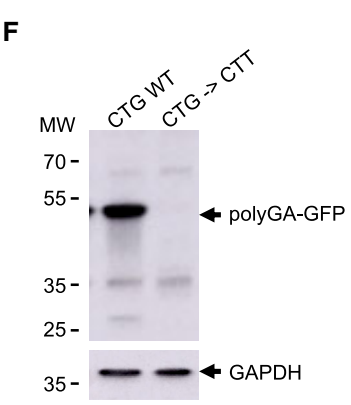


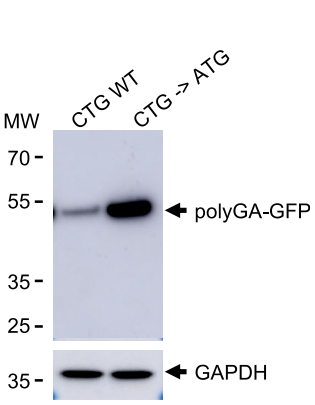


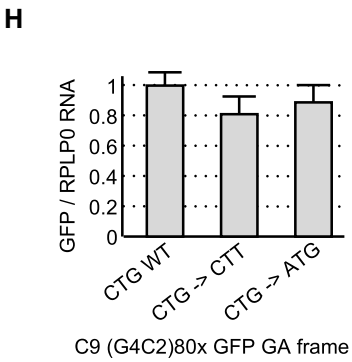


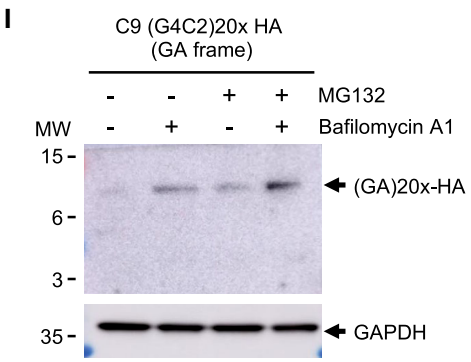












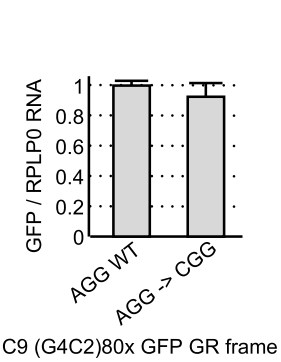


Figure EV1.

© 2020 The Authors

Figure EV2. Expanded C4G2 repeats are translated into polyPG.

A, B Fluorescence (A) and RT–qPCR (B) GFP analyses of HEK293 cells transfected for 24 h with 100 C4G2 repeats embedded in the human antisense *C9ORF72* sequence and fused in all three possible frames with the GFP deleted of its ATG.

C, D GFP expression (C) and immunoblotting (D) analysis of HEK293 cells transfected for 24 h with constructs expressing either polyPA or polyPR expressed under an artificial ATG start codon and fused to the GFP. Fluorescence of ATG-polyGP from Fig EV1D is shown as control.

E Immunoblotting against HA or the GAPDH of proteins extracted from HEK293 cells transfected for 24 h with either a wild-type or a mutant (ΔATG) construct containing 100 C4G2 repeats embedded in the antisense *C9ORF72* sequence fused to a HA-tag in the PG frame.

F RT–qPCR GFP expression analysis of HEK293 cells transfected for 24 h with either a wild-type or a mutant (ΔATG) construct containing 100 C4G2 repeats embedded in the antisense *C9ORF72* sequence fused to the GFP in the PG frame.

G Immunoblotting against the HA or the GAPDH of proteins extracted from HEK293 cells transfected for 24 h with either a wild-type or a Kozak consensus mutant construct containing 100 C4G2 repeats embedded in antisense *C9ORF72* fused to the HA-tag in the PG frame.

H Immunoblotting against the HA-tag or the GAPDH of proteins extracted from HEK293 cells transfected for 24 h with 10 C4G2 repeats embedded in the human antisense *C9ORF72* sequence fused to a HA-tag in the PG frame and treated or not with MG132 and/or bafilomycin A1 for 15 h.

© 2020 The Authors

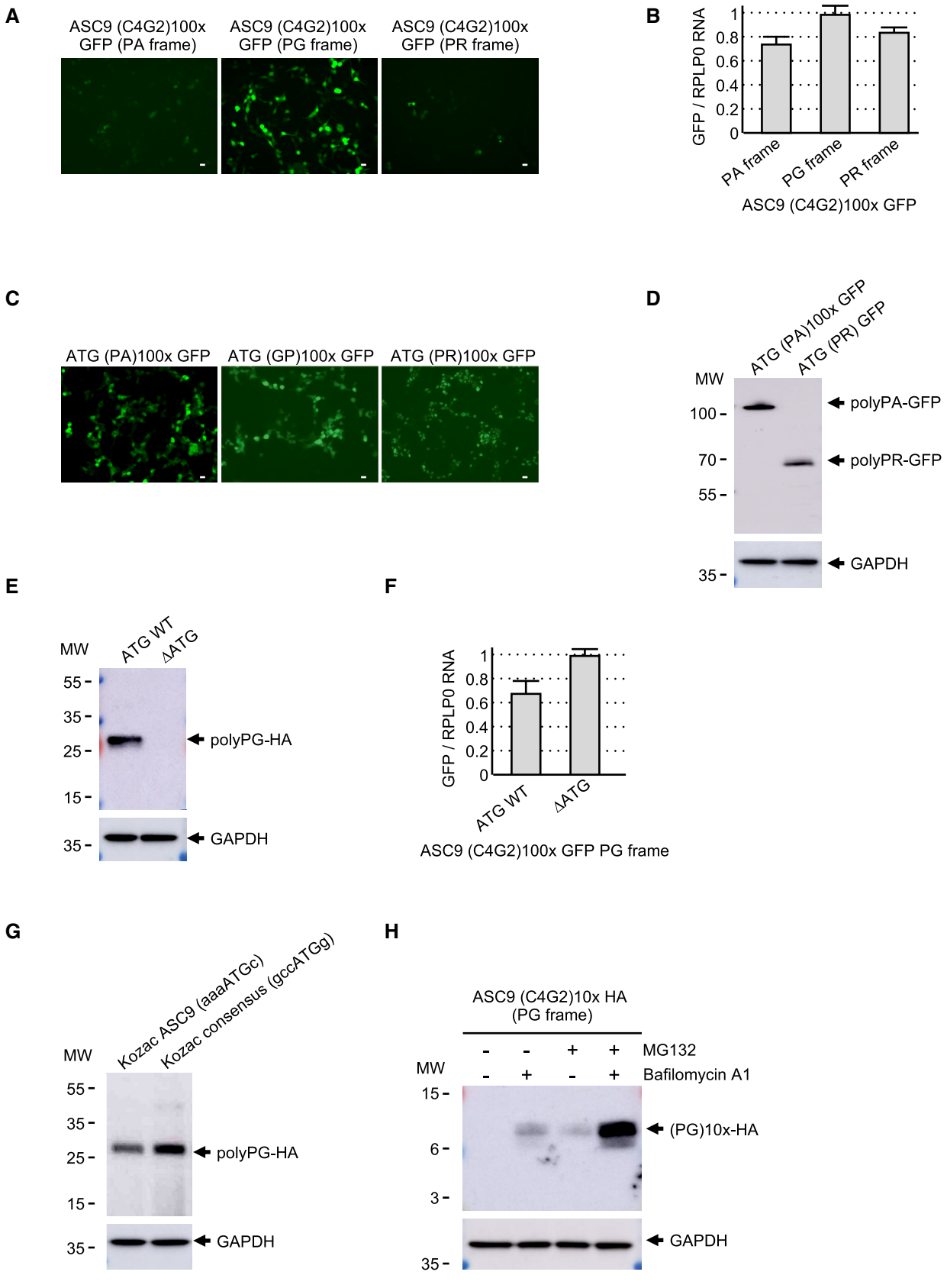

A
ASC9 (C4G2)100x GFP (PA frame)
ASC9 (C4G2)100x GFP (PG frame)
ASC9 (C4G2)100x GFP (PR frame)
B
GFP / RPLP0 RNA
1
0.8
0.6
0.4
0.2
0
PA frame
PG frame
PR frame
ASC9 (C4G2)100x GFP
C
ATG (PA)100x GFP
ATG (GP)100x GFP
ATG (PR)100x GFP
D
ATG (PA)100x GFP
ATG (PR) GFP
MW
100
70
55
35
polyPA-GFP
polyPR-GFP
GAPDH
E
MW
ATG WT
ΔATG
55
35
25
15
35
polyPG-HA
GAPDH
F
GFP / RPLP0 RNA
1
0.8
0.6
0.4
0.2
0
ATG WT
ΔATG
ASC9 (C4G2)100x GFP PG frame
G
MW
Kozac ASC9 (aaaATGc)
Kozac consensus (gccATGg)
55
35
25
15
35
polyPG-HA
GAPDH
H
ASC9 (C4G2)10x HA (PG frame)
MW
- - + + MG132
- + - + Bafilomycin A1
15
6
3
35
(PG)10x-HA
GAPDH

Figure EV2.

Figure EV3. Decreased expression of C9ORF72 synergizes DPR toxicity.

A Immunoblotting against the HA-tag or the GAPDH of proteins extracted from HEK293 cells transfected for 24 h with 80 G4C2 repeats embedded in the human sense *C9ORF72* sequence fused to a HA-tag in the GA frame and treated or not with MG132 and/or bafilomycin A1.

B As in (A) but with cells transfected with 100 C4G2 repeats embedded in the human antisense *C9ORF72* sequence fused to a HA-tag in the PG frame.

C, D Left panel, representative images of immunofluorescence labeling of endogenous P62/SQSTM1 (C) or ubiquitin (D) and the GFP in GT1-7 neuronal cells co-transfected for 24 h with either a control siRNA or a siRNA targeting *C9orf72* mRNA and a construct containing 80 G4C2 repeats embedded in sense *C9ORF72* fused to the GFP in the GA frame. Right panel, quantification of the percent of co-localization of P62 or ubiquitin with polyGA aggregates.

E Immunofluorescence labeling of M6PR in GT1-7 neuronal cells transfected for 24 h with either a control siRNA or a siRNA targeting *C9orf72* mRNA.

F Immunofluorescence labeling of EEA1 in GT1-7 neuronal cells transfected for 24 h with either a control siRNA or a siRNA targeting *C9orf72* mRNA.

Data information: Error bars indicate s.e.m. Student's *t*-test, ***$P < 0.001$. $n = 3$ independent transfection. Scale bars, 10 μm. Nuclei were counterstained with DAPI.

© 2020 The Authors

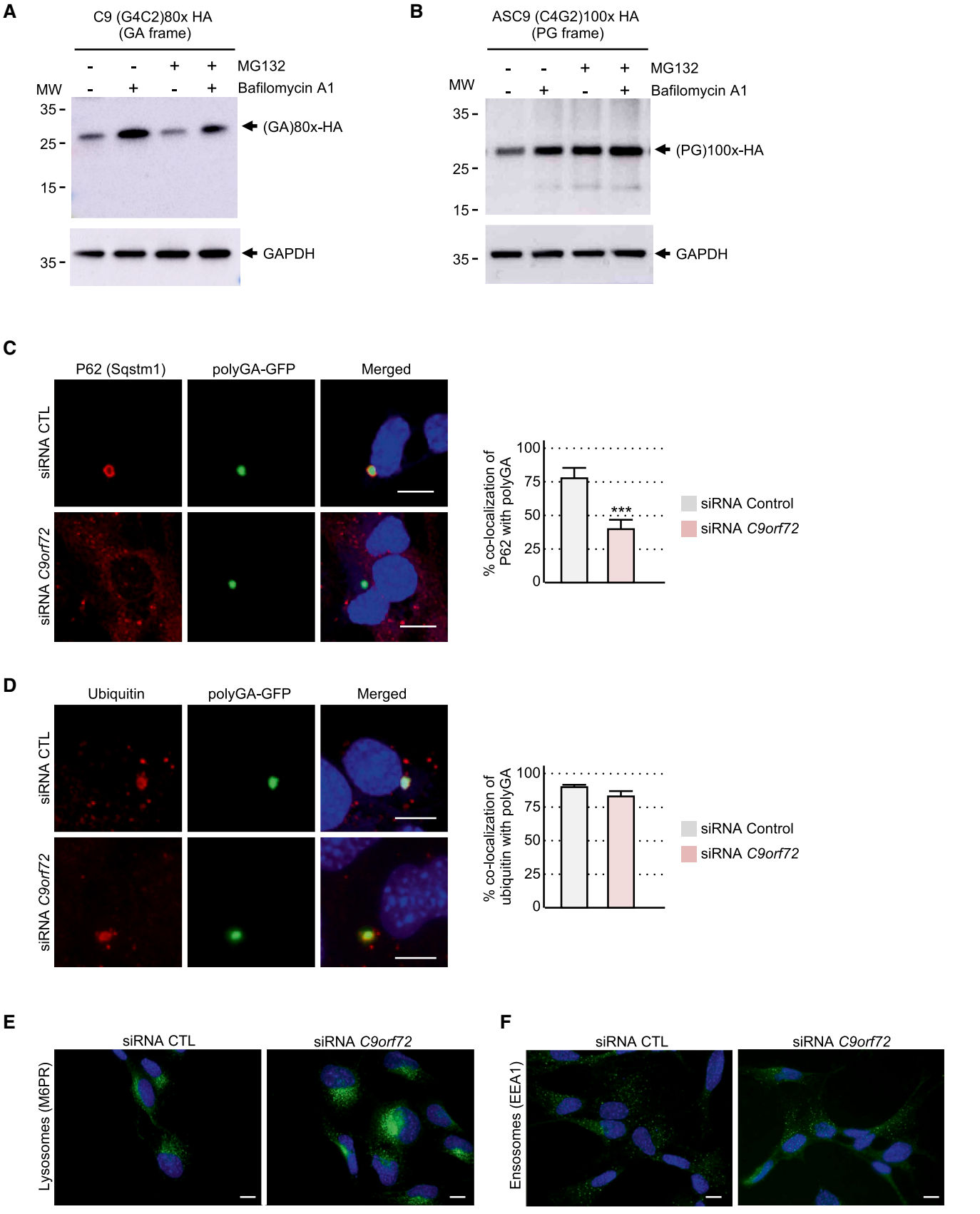

A
C9 (G4C2)80x HA
(GA frame)
MG132
Bafilomycin A1
MW
35
25
15
(GA)80x-HA
GAPDH
B
ASC9 (C4G2)100x HA
(PG frame)
MG132
Bafilomycin A1
MW
35
25
15
(PG)100x-HA
GAPDH
C
P62 (Sqstm1)
polyGA-GFP
Merged
siRNA CTL
siRNA C9orf72
% co-localization of P62 with polyGA
100
75
50
25
0
***
siRNA Control
siRNA C9orf72
D
Ubiquitin
polyGA-GFP
Merged
siRNA CTL
siRNA C9orf72
% co-localization of ubiquitin with polyGA
100
75
50
25
0
siRNA Control
siRNA C9orf72
E
siRNA CTL
siRNA C9orf72
Lysosomes (M6PR)
F
siRNA CTL
siRNA C9orf72
Ensosomes (EEA1)

Figure EV3.

Figure EV4. Promethazine reduces DPR protein accumulation and toxicity.

A Left panel, immunoblotting against the GFP or the GAPDH of proteins extracted from HEK293 cells transfected for 24 h with a construct expressing under an artificial ATG start codon 100 GA repeats fused to the GFP (ATG (GA)100× GFP) and treated 15 h with 10 μM of the indicated drug. Right panel, quantification of polyGA expression relative to the GAPDH.

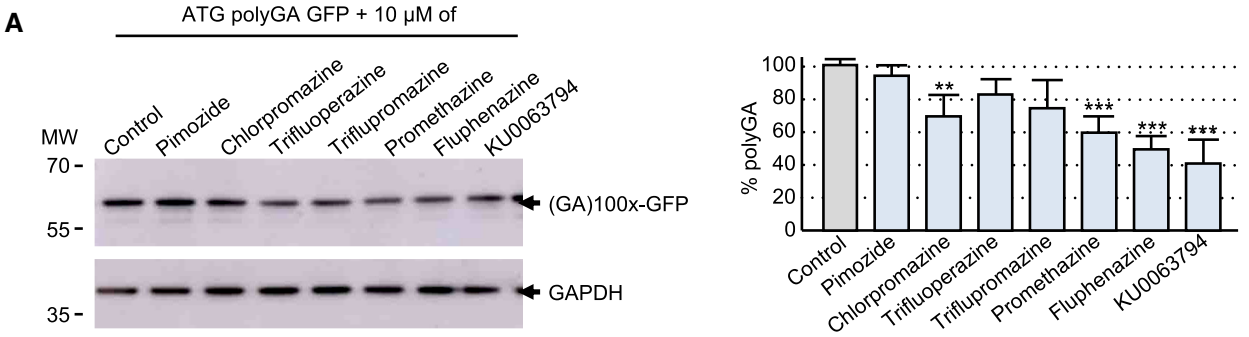
B Immunoblotting against the GFP or the GAPDH of proteins extracted from HEK293 cells transfected for 24 h with ATG (GA)100× GFP and treated 15 h with 1, 3, or 10 μM of the indicated drug.

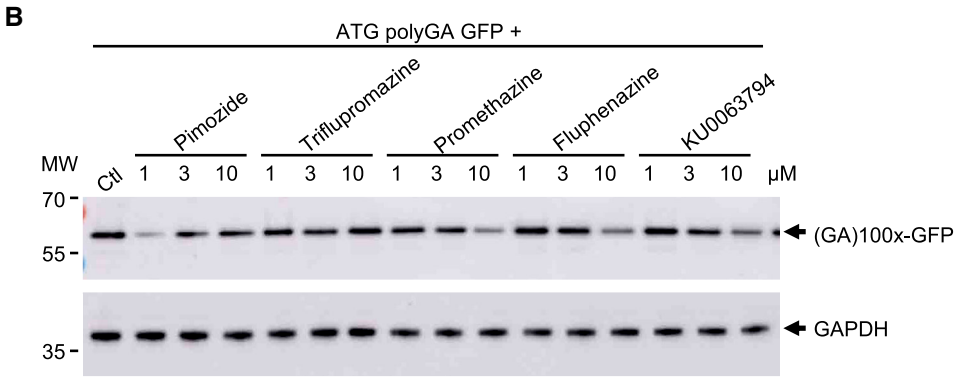
C Left panel, GFP fluorescence of HEK293 cells transfected for 24 h with ATG (GA)100× GFP and treated with 10 μM of the indicated compound. Scale bars, 10 μm. Nuclei were counterstained with DAPI. Right panel, quantification of polyGA aggregates.

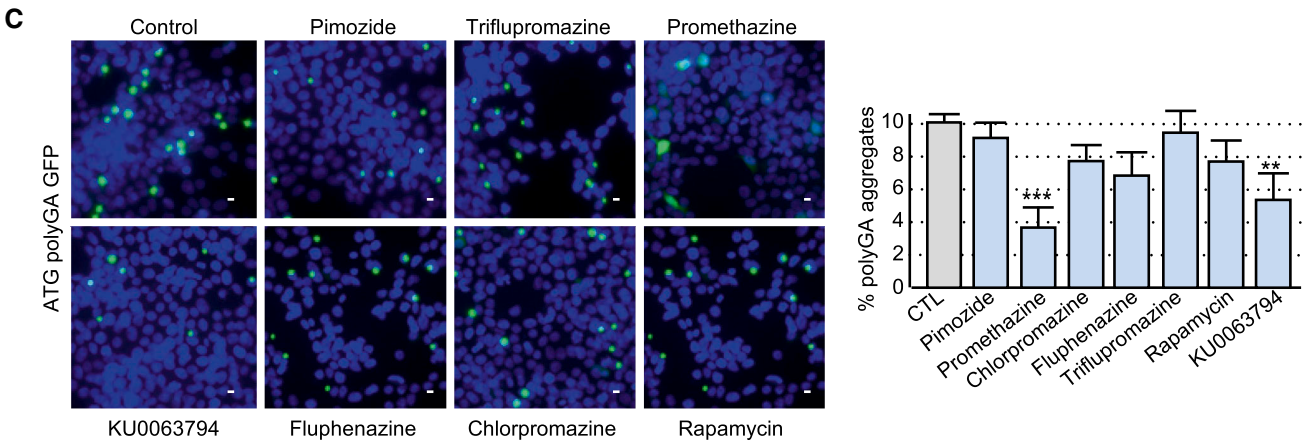
D Cell viability (TO-PRO-3 FACS staining) of GT1-7 neuronal cells treated with 1, 3, or 10 μM of promethazine and co-transfected for 24 h with ATG (GA)100× GFP and either a control siRNA or a siRNA targeting *C9orf72* mRNA.

Data information: Error bars indicate s.e.m. Student's *t*-test, $^{**}P < 0.01$, and $^{***}P < 0.001$. $n = 5$ independent transfection.

© 2020 The Authors


A
ATG polyGA GFP + 10 μM of
MW
Control
Pimozide
Chlorpromazine
Trifluoperazine
Triflupromazine
Promethazine
Fluphenazine
KU0063794
70
55
35
(GA)100x-GFP
GAPDH
% polyGA
100
80
60
40
20
0
**
***
***
***
Control
Pimozide
Chlorpromazine
Trifluoperazine
Triflupromazine
Promethazine
Fluphenazine
KU0063794


B
ATG polyGA GFP +
Pimozide
Triflupromazine
Promethazine
Fluphenazine
KU0063794
MW
Ctl
1
3
10
μM
70
55
35
(GA)100x-GFP
GAPDH


C
Control
Pimozide
Triflupromazine
Promethazine
ATG polyGA GFP
KU0063794
Fluphenazine
Chlorpromazine
Rapamycin
% polyGA aggregates
10
8
6
4
2
0
***
**
CTL
Pimozide
Promethazine
Chlorpromazine
Fluphenazine
Triflupromazine
Rapamycin
KU0063794

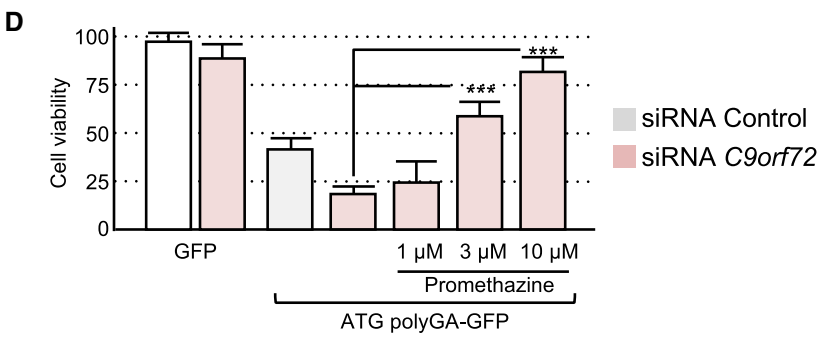

D
Cell viability
100
75
50
25
0
***
***
GFP
1 μM
3 μM
10 μM
Promethazine
ATG polyGA-GFP
siRNA Control
siRNA C9orf72

Figure EV4.

DISCUSSION

DISCUSSION

- CHAPITRE 1 -

Les maladies à polyGlycine

Nous allons maintenant aborder la partie discussion qui sera composée de 4 parties distinctes. La première partie portera sur les maladies à polyGlycine où je présenterai d'autres maladies causées par des expansions de répétitions CGG pouvant, tout comme dans NIID et FXTAS, être potentiellement traduites en protéines à polyG ; nous discuterons également du lien entre la taille des expansions CGG et la sévérité de ces pathologies puis, nous terminerons par le mécanisme de toxicité putatif des protéines à polyG. La seconde partie portera sur l'expansion de répétitions GGGGCC située dans le gène C9ORF72, responsable de la SLA/DFT où les points suivants seront abordés : la synergie de toxicité entre perte d'expression de la protéine C9ORF72 et expression des protéines ; les fonctions potentielles de la protéine C9ORF72 et ; la toxicité des protéines DPR en modèles murins. Dans la troisième partie, nous décrirons les possibles voies thérapeutiques concernant les maladies dues à l'expression de protéine composée de polyGlycine (FXTAS, NIID et OPDM) ainsi que pour la C9-SLA/DFT. Enfin, cette discussion terminera sur le mécanisme de traduction des expansions de répétitions (traduction canonique vs traduction « RAN »).

1. LES MALADIES A POLY-GLYCINE

Il y a maintenant 20 ans, la cause génétique du syndrome de tremblement et d'ataxie lié à l'X fragile (FXTAS) a été identifiée, il s'agit d'une expansion de 55 à 200 répétitions de tri-nucléotides CGG située dans la région 5'UTR du gène *FMR1* (Hagerman et al., 2001). Par la suite, il a été démontré que ces expansions étaient traduites en une nouvelle protéine comportant une séquence de résidus glycines appelée FMRpolyG (Todd et al., 2013 ; Buijsen et al., 2014 ; Sellier et al. 2017 ; Krans et al., 2019). Cette protéine est responsable de la formation des agrégats intranucléaires retrouvés chez les patients FXTAS et conduit à une ataxie chez les souris, faisant de FXTAS la première maladie à polyG (Sellier et al., 2017). Depuis, des expansions de séquence (tri-nucléotides CGG), de taille (55 à 200 répétitions) et de localisation (5'UTR) similaires ont été décrites dans les gènes *NOTCH2NLC*, *LRP12*, *GIPC1* et le locus *642361* responsables respectivement de NIID / OPDM3, OPDM1, OPDM2 et OPML. De manière intéressante, les patients NIID et OPDM présentent des agrégats intranucléaires comparables à ceux observés chez les patients FXTAS. Dans l'OPML, des analyses histopathologiques sur des biopsies de muscles de patients n'ont pas permis d'identifier des agrégats intranucléaires (Ishiura et al., 2019). Néanmoins, seuls 2 échantillons de muscles squelettiques ont été analysés et, au vu de l'expression du gène locus *642361* au niveau ARN

dans les différents organes chez l'Homme et des symptômes neurodégénératifs développés par les patients OPML, il serait intéressant de tester la présence d'inclusions intranucléaires dans le système nerveux central de ces patients.

Lors de ce travail de thèse, nous avons pu mettre en évidence que les expansions de nucléotides CGG situées dans les régions 5'UTR des gènes *NOTCH2NLC* et *GIPC1* étaient traduites en protéines composées de polyGlycine, que nous avons appelé respectivement uN2CpolyG et uGIPC1polyG. Ces pathologies, NIID et OPDM2/3, viennent donc agrandir la famille des maladies à polyG (Figure 55).

Il est à noter qu'en octobre 2021, Zhong et al. ont également montré que les expansions de nucléotides CGG dans le gène *NOTCH2NLC* étaient traduites en une protéine composée de polyGlycine, qu'ils ont appelé N2NLCpolyG. Cette protéine, N2NLCpolyG est identique à la protéine uN2CpolyG que nous avons décrite. Tout comme dans notre travail, ils ont démontré que l'initiation de la traduction a lieu dans une uORF au codon ATG situé en amont des répétitions CGG. L'expression de la protéine N2NLCpolyG conduit à une diminution du métabolisme cellulaire en culture primaire de neurones de souris, décelée par une hyperactivité de la lactate déshydrogénase. Enfin, ils ont mis en évidence que la protéine N2NLCpolyG altère, tout comme FMRpolyG, l'enveloppe nucléaire des cellules, conduisant à une altération du transport nucléocytoplasmique (Zhong et al., 2021).

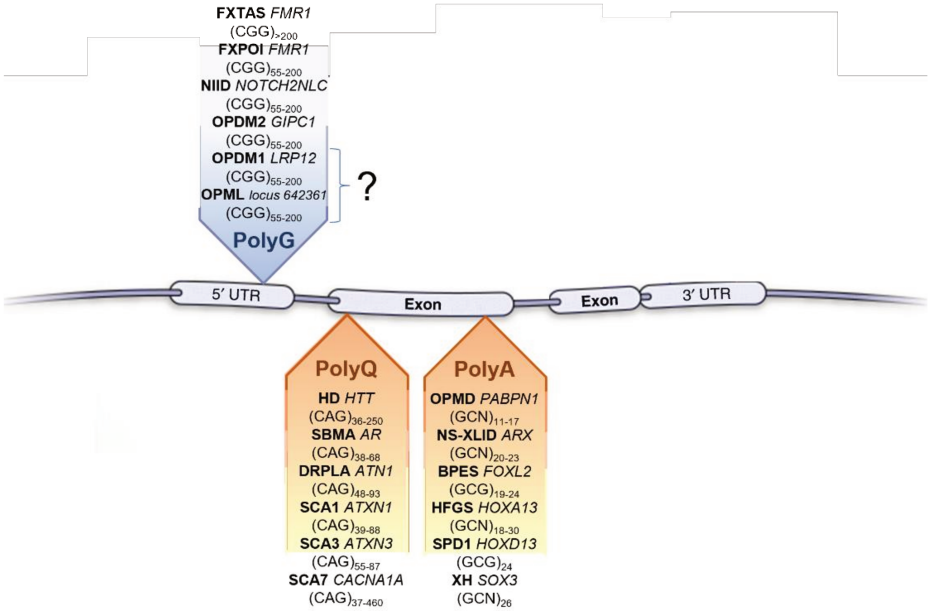


Figure 55. Expansions de répétitions traduites en protéines toxiques
Les expansions CGG localisées dans les régions 5'UTR de différents gènes sont traduites en protéines composées de polyG. Les expansions CAG et GCN localisées dans la phase codante des gènes sont traduites en protéines composées de polyG et de polyA respectivement.

1.1. Autres expansions CGG et maladies à polyG

Les progrès en termes de séquençage ont permis, ces dernières années, d'identifier de nombreuses expansions de répétitions dans différentes pathologies. Depuis la découverte en 1991 de l'expansion de répétitions CGG dans le gène *FMR1* causant FXS, plusieurs expansions pathologiques CGG ont été décrites dans les gènes *FMR1*, *XYLT1*, *NOTCH2NLC*, *LRP12*, *GIPC1* et dans le locus *642361* responsables respectivement de FXTAS, BSS, NIID / OPDM3, OPDM1, OPDM2 et OPML ; mettant en lumière une fréquence relativement élevée des expansions de répétitions CGG dans les maladies à expansions de répétitions (Hagerman et al., 2001 ; Ishiura et al., 2019 ; LaCroix et al., 2019 ; Sone et al., 2019 ; Tian et al., 2019 ; Deng et al., 2020). Parmi ces pathologies, nous avons pu montrer que ces répétitions CGG étaient traduites en protéines composées de polyGlycine dans FXTAS, NIID / OPDM3 et OPDM2. Il est à noter que les expansions dans le gène NOTCH2NLC peuvent conduire à 2 pathologies différentes : une maladie neurodégénérative et neuromusculaire, NIID, ou une maladie uniquement neuromusculaire, OPDM, suggérant un continuum entre ces pathologies. Il serait donc intéressant de tester si les répétitions CGG dans le gène *LRP12* et le locus *642361*, responsables respectivement de l'OPDM1 et l'OPML, sont également traduites en protéines composées de polyGlycine (Figure 55). De manière intéressante, des expériences préliminaires semblent indiquer que les expansions CGG dans le gène *LRP12* sont traduites dans les cadres de lecture glycine et arginine (cf Discussion 4.2). De plus, l'analyse de la séquence nucléotidique du locus *642361* révèle la présence d'un ATG en amont des répétitions CGG et dans le cadre de lecture glycine. Ces observations suggèrent que, tout comme dans FXTAS, NIID et OPDM2, les répétitions CGG dans l'OPDM1 et l'OPML pourraient être, elles aussi, traduites en protéines composées de polyGlycine.

Il est aussi intéressant de noter que les patients NIID porteurs d'une expansion de répétitions de nucléotides CGG dans le gène *NOTCH2NLC* sont presque exclusivement asiatiques. Au contraire, les patients NIID européens ne sont pas porteurs de cette mutation, suggérant que NIID est une maladie hétérogène au niveau génétique (Chen et al., 2020a). De plus, l'absence d'une expansion CGG dans le gène NOTCH2NLC chez les patients NIID européens soulève l'hypothèse de l'existence d'une expansion de nucléotides CGG localisée dans un autre gène. Ces observations suggèrent que, tout comme l'expansion de répétitions GGGGCC dans le gène *C9ORF72* dans la SLA/DFT, il existe un effet fondateur fort de ces mutations.

1.2. Taille de l'expansion CGG et sévérité

Dans la DM1 et les maladies à polyQ, il a été montré que plus les expansions de répétitions CUG ou CAG sont grandes, plus la maladie est sévère et plus les premiers symptômes de ces pathologies se développent tôt chez les patients. Cet allongement des répétitions peut se produire lors de la méiose et, donc, lors du passage d'une génération à la suivante. Ce phénomène conduit au mécanisme d'anticipation où les enfants d'un individu malade présentent généralement un nombre de répétitions plus élevé et, par conséquent, développent des manifestations cliniques plus précocement (pour revue, Lieberman et al., 2019). Toutefois, dans les maladies à polyG, un mécanisme d'anticipation et / ou une corrélation exacte entre la taille de l'expansion et la sévérité des symptômes ne fait pas actuellement consensus (Xi et al., 2021). En effet, dans NIID, une taille de répétitions comprise entre 40 et 80 tri-nucléotides augmente le risque de développer des syndromes parkinsoniens ou même la maladie de Parkinson (Ma et al., 2020), alors qu'une expansion supérieure à 100 nucléotides CGG est associée à une faiblesse musculaire et à une démence (Sone et al., 2019 ; Ishiura et al., 2019 ; Deng et al., 2019 ; Tian et al., 2019). Ces données suggèrent une sévérité différente selon la taille des répétitions. Toutefois, il a été observé que certains individus étaient porteurs d'expansions de répétitions CGG dans les gènes *NOTCH2NLC* ou *GIPC1* mais ne développaient pas la pathologie associée (Figure 56 ; Deng et al., 2020 ; Fukuda et al., 2021). En effet, des parents de patients NIID possédant un nombre de répétitions CGG supérieur à 200, conduisant à des modifications épigénétiques du promoteur et à une perte d'expression de la protéine NOTCH2NLC, ne développent pas de symptômes particuliers (Fukuda et al., 2021). De manière similaire, dans l'OPDM2, il a été récemment observé qu'un individu asymptomatique était porteur d'une expansion de plus de 500 répétitions CGG dans le gène *GIPC1* alors que ses 2 enfants sont porteurs d'expansions de répétitions CGG plus courtes (117 et 113 répétitions) et ont tous 2 développés la pathologie (Deng et al., 2020). Cette longue expansion de 500 répétitions CGG est méthylée, conduisant à une diminution d'expression de la protéine GIPC1 portée par cet allèle et donc à une absence d'expression de la protéine uGIPC1polyG toxique ; ce qui n'est pas le cas chez les enfants de cet individu qui sont porteurs d'expansions plus courtes (Deng et al., 2020). Ces données sont à mettre en perspective avec FXS et FXTAS. En effet, les expansions de répétitions CGG situées dans le gène *FMR1* peuvent conduire à la perte de fonction de la protéine FMRP dans FXS lorsque l'expansion CGG est longue et méthylée ou, à la génération d'une nouvelle protéine toxique FMRpolyG dans FXTAS

lorsque l'expansion est plus courte et donc transcrite ; faisant donc intervenir 2 mécanismes de toxicité distincts.

Ces informations engendrent plusieurs conséquences importantes. Premièrement, ces données montrent que l'haploinsuffisance des protéines NOTCH2NLC et GIPC1 due à de longues expansions CGG, ne conduirait pas aux maladies NIID et OPDM. Cette diminution d'expression de ces protéines peut probablement être compensée par le second allèle ; ce qui n'est pas le cas pour la protéine FMRP puisque le gène *FMR1* est porté par le chromosome X, l'absence de la protéine FMRP conduisant au syndrome de l'X fragile chez les hommes (Oberlé et al., 1991 ; Verkerk et al., 1991 ; Fu et al., 1991). Deuxièmement, de manière intéressante, une expansion supérieure à 200 répétitions CGG pourrait entrainer un effet protecteur puisque l'expansion serait alors méthylée, conduisant à une inhibition de la transcription du gène hôte et, par conséquent, une absence de la traduction des expansions CGG en protéine polyGlycine toxique.

Pour conclure, un phénomène d'anticipation n'est donc pas forcément observé dans les maladies à polyG, excepté dans FXS / FXTAS où il s'agit de maladies et de mécanismes de toxicité différents. De plus, la transmission de la pathologie n'est pas systématiquement dominante due à l'expansion et / ou la rétractation aléatoire de ces répétitions CGG, ce qui pourrait expliquer la complexité d'interprétation des arbres généalogiques de ces familles (Figure 56). En effet, les protéines à polyG sont toxiques à partir d'une cinquantaine de glycines et, au-dessus de 300 répétitions CGG, cette expansion est méthylée conduisant à une absence de transcription et, donc, de la traduction des protéines à polyG. Ces données suggèrent une fenêtre pathologique étroite des expansions de répétitions CGG.

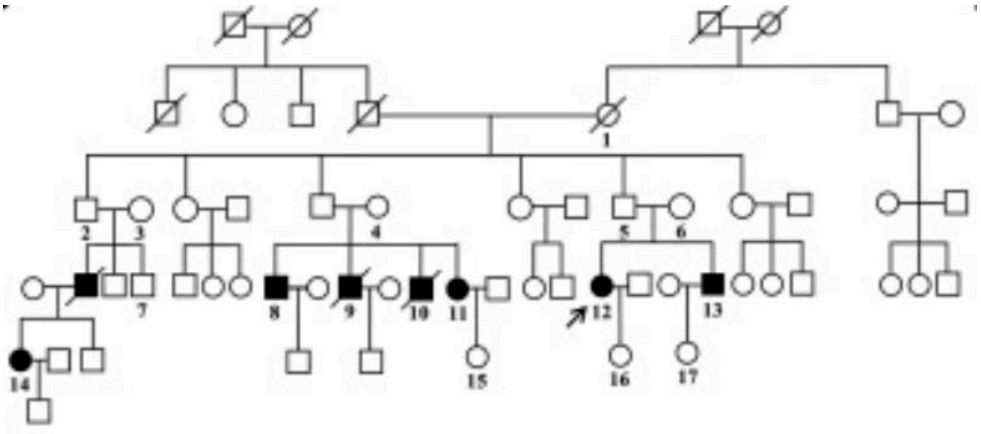


Figure 56. Arbre généalogique d'une famille OPDM
Les ronds et les carrés représentent les hommes et les femmes respectivement. Une diagonale indique que la personne est décédée. Les individus atteints d'OPDM sont représentés par une forme géométrique noire. Arbre généalogique de Deng et al., 2020.

1.3. Mécanisme de toxicité des protéines à polyG

Tout comme les protéines à polyQ responsables de plusieurs pathologies aux atteintes neurologiques et à l'histopathologie comparables suggérant un mécanisme de toxicité commun à ces maladies, les protéines à polyG sont maintenant retrouvées dans les maladies FXTAS, NIID et OPDM aux symptômes et à l'histopathologie similaires ; se pose alors la question d'un mécanisme de toxicité commun à ces maladies. Toutefois, ces maladies à polyG présentent également des particularités et, nous pouvons donc nous demander quels sont les mécanismes moléculaires sous-jacents à ces différences.

1.3.1. Niveau d'expression des protéines à polyG

A ce jour retrouvées dans 3 des maladies neurodégénératives et / ou neuromusculaires, les protéines à polyG forment des inclusions intranucléaires éosinophiliques identiques. Cependant, la sévérité et l'âge d'apparition sont différents entre ces pathologies. En effet, il existe des formes juvéniles de NIID, alors que l'OPDM apparaît généralement aux alentours de 30 ans, et que FXTAS touche des personnes âgées de 60 à 80 ans. Il est important de noter que les différentes protéines à polyG détectées ne sont pas exprimées de manière homogène. En effet, l'initiation de la traduction a lieu à un codon ATG dans NIID / OPDM3, CTG dans OPDM2 et ACG dans FXTAS ; conduisant à des niveaux d'expression de ces protéines différents. En effet, le codon ATG permet une initiation de la traduction plus forte que le codon CTG, lui-même initiant plus fortement que le codon ACG. Le niveau d'expression de ces protéines à polyG joue donc certainement un rôle dans l'âge d'apparition et la sévérité de la pathologie.

1.3.2. Toxicité tissu spécifique des protéines à polyG ?

Les maladies FXTAS, NIID, OPDM et OPML sont caractérisées par des altérations de systèmes et d'organes différents qui conduisent donc à des symptômes différents, mais avec des atteintes principalement neurologiques et musculaires (Figure 57). Nous pouvons nous demander d'où provient cette spécificité tissulaire. Une hypothèse probable est que, suivant l'expression tissulaire des gènes, ce ne sont pas les mêmes organes qui sont atteints. En effet, le gène *FMR1*

est fortement exprimé dans le cerveau et notamment dans le cervelet mais peu exprimé dans les muscles squelettiques, cette différence d'expression pourrait éventuellement expliquer le fait que FXTAS soit une maladie neurodégénérative sans atteinte musculaire.

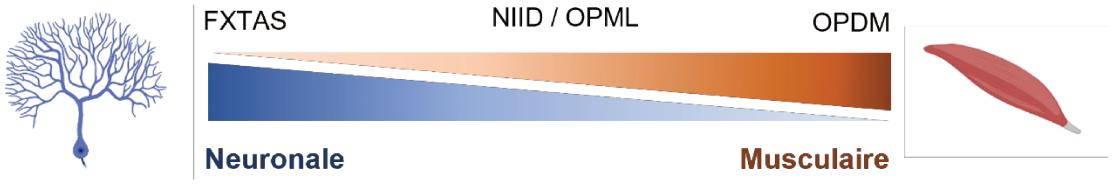


Figure 57. Les maladies à polyG : maladies neurodégénératives et neuromusculaires
Parmi les maladies à polyG, FXTAS est caractérisée par une neurodégénéresence, OPDM se traduit par des symptomes musculaires et, NIID et OPML présentent à la fois une altération des cellules neuronales et musculaires.

Le gène *NOTCH2NLC* est, lui, exprimé dans les neurones et les fibres musculaires, ce qui pourrait expliquer que des patients porteurs d'une expansion de répétitions CGG dans le gène *NOTCH2NLC* peuvent développer une maladie avec des atteintes neurodégénérative et neuromusculaire. Néanmoins, on observe une grande variabilité des symptômes d'un individu à l'autre. Peut-être que la protéine uN2CpolyG est différentiellement exprimée selon les tissus d'un patient à l'autre. En effet, les expansions de répétitions sont instables et la taille de l'expansion de répétitions peut varier d'un tissu à l'autre chez un même patient. Les répétitions CGG étant toxiques uniquement entre 60-70 et 200-300, des expansions trop courtes ou trop longues dans le cerveau ou dans les muscles permettraient d'expliquer l'absence de symptômes neurodégénératifs ou musculaires, respectivement. De plus, il est important de souligner que 2 études ayant réalisé une analyse histopathologique poussée des patients NIID révèlent la présence d'agrégats intranucléaires positifs à l'ubiquitine et à p62 dans tous les tissus analysés (poumon, cœur, foie, pancréas, etc.), suggérant que NIID n'est pas uniquement une maladie neurodégénérative et / ou neuromusculaire mais, une maladie multi-systémique avec de multiples atteintes sous cliniques (Yamaguchi et al., 2018 ; Chen et al., 2020b).

Enfin, de façon surprenante, bien que les gènes *LRP12* et *GIPC1* soient exprimés de manière ubiquitaire au niveau ARN, les patients OPDM développent des altérations musculaires sans atteinte neurodégénérative. Il serait donc intéressant de savoir si ces patients présentent des agrégats intranucléaires dans le système nerveux central. En effet, si la protéine uGIPC1polyG est retrouvée chez le cerveau des patients atteints d'OPDM, cela suggérerait que cette protéine n'est pas toxique pour les neurones mais uniquement pour les cellules musculaires. Ces

observations rappellent les maladies à polyQ où, bien que certaines protéines comme la HTT soient exprimées de manière ubiquitaire, seuls certains neurones dégénèrent, suggérant une sensibilité particulière du système nerveux central aux protéines à polyQ.

1.3.3. Les protéines à polyG sont-elles toxiques d'elles-mêmes ?

Durant ce travail de thèse, nous avons pu montrer que, comme dans FXTAS où les expansions de répétitions CGG situées dans une uORF du gène *FMR1* sont traduites en FMRpolyG ; les expansions de répétitions CGG localisées dans des uORF des gènes *NOTCH2NLC* et *GIPC1* étaient traduites en protéines composées de polyGlycine, appelées respectivement uN2CpolyG et uGIPC1polyG. Ces différentes protéines à polyG possèdent des séquences N-terminales et C-terminales qui leur sont propres. Nous pouvons nous demander si, comme dans les maladies à polyQ, la présence de ces séquences peut moduler la toxicité des protéines à polyG.

Lors de l'étude portant sur NIID, nous avons développé plusieurs virus : celui codant pour la protéine uN2CpolyG-GFP et celui codant pour la protéine uN2C-GFP (sans expansion de répétitions CGG). Seules les souris exprimant la protéine avec la suite de glycines développent des altérations locomotrices, montrant que la toxicité de cette protéine provient bien des résidus glycines et non des séquences bordant l'expansion. Toutefois, une contribution des séquences bordantes n'est pas à exclure. En effet, dans FXTAS, la délétion de la partie C-terminale de la protéine FMRpolyG conduit à une plus faible toxicité de cette protéine en cellules HEK293 et N2A (Sellier et al., 2017). De plus, des drosophiles exprimant la protéine FMRpolyG meurt plus précocement que des drosophiles exprimant une protéine composée uniquement de polyGlycines (Sellier et al., 2017). De même, nous avons récemment développé un virus codant pour une protéine composée uniquement de polyGlycine sans séquences bordantes et, mes résultats préliminaires montrent que ces souris ne développent pas les altérations phénotypiques observées chez les souris exprimant uN2CpolyG. Néanmoins, cette protéine polyG semble être moins exprimée, les séquences bordantes pourraient donc jouer un rôle dans la stabilité des différentes protéines à polyG (Figure 58).

Enfin, mes données suggèrent que la toxicité des protéines à polyG dépend fortement du modèle utilisé. En effet, lors de l'étude de la protéine uN2CpolyG, je n'ai pas observé de toxicité de cette protéine dans des cellules immortalisées de type HEK293, U2OS et GT1-7, mais une toxicité dans des cultures primaires de neurones d'embryons de souris. De plus, j'ai observé de

grandes variations d'expression de la protéine composée uniquement de polyGlycine où, cette protéine s'exprime fortement dans les fibres musculaires de souris conduisant à la présence de nombreux agrégats, alors que cette protéine s'exprime très faiblement et / ou ne s'agrège pas dans le cerveau des souris, suggérant que les protéines à polyG pourraient être plus ou moins stables selon le type cellulaire ou le tissu considéré.

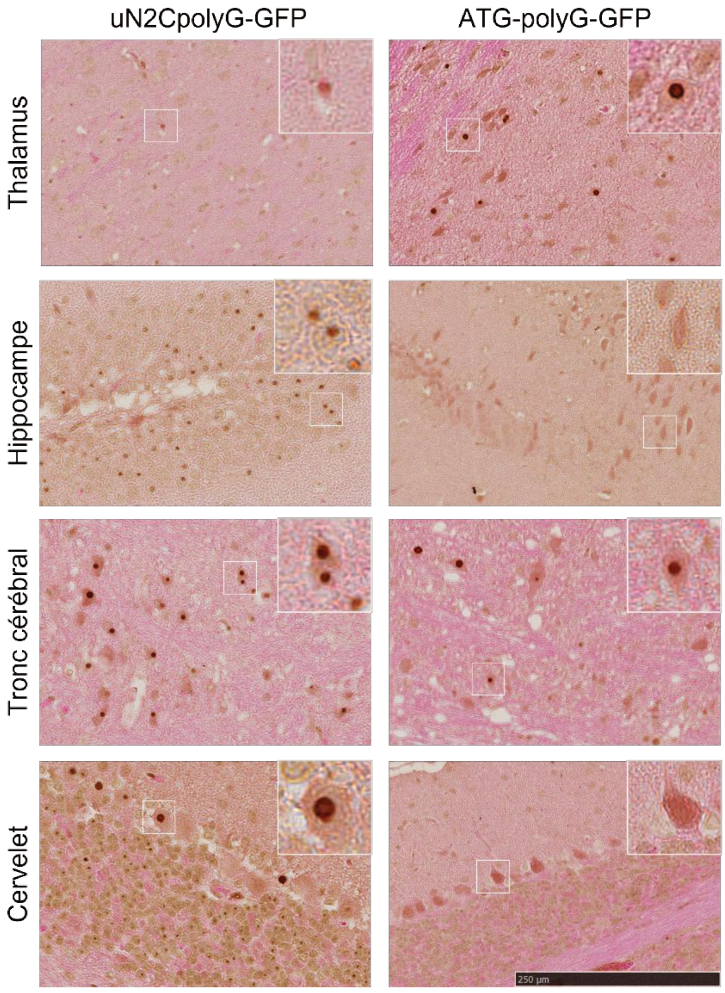


Figure 58. Expression de uN2CpolyG-GFP est ATG-polyG-GFP dans le cerveau de souris
IHC réalisées contre la protéine p62 et coloration à l'hématoxyline et éosine de cerveaux de souris sacrifiées à 3 mois après injection de $1{,}5.10^{13}$ vg/kg d'AAV/PHP.eB. La protéine uN2CpolyG-GFP forme des inclusions intranucléaires de manière homogène dans les différentes zones du cerveau alors que les souris exprimant la protéine ATG-polyG-GFP présentent des agrégats principalement dans le thalamus. Barre d'échelle = 250 µm.

Ces résultats suggèrent que les séquences entourant les polyGlycine participent à la stabilité et à la toxicité des protéines à polyG mais aussi, que chaque cellule / tissu pourrait répondre différemment à l'expression de ces protéines.

1.3.4. Localisation des protéines à polyG et toxicité

Les analyses histopathologiques sur des tissus de patients FXTAS, NIID et OPDM révèlent la présence d'inclusions éosinophiliques causées par l'agrégation des différentes protéines à polyG. De manière intéressante, les inclusions observées chez les patients, ainsi que dans les différents modèles murins exprimant les protéines à polyG, sont intranucléaires. Toutefois, l'expression des protéines à polyG dans différentes lignées de cellules immortalisées conduit à l'agrégation de ces protéines majoritairement dans le cytoplasme et, plus rarement dans le noyau.

Nous pouvons donc nous demander pourquoi, en modèle cellulaire, ces protéines forment principalement des agrégats cytoplasmiques alors qu'en modèle murin et chez les patients, ces protéines forment des agrégats nucléaires. Nous avons tout d'abord émis l'hypothèse que la GFP, une protéine de 25 kDa, pouvait empêcher de par sa taille la migration des protéines à polyG dans le noyau. Malheureusement, le remplacement de la GFP par de petites étiquettes de type FLAG ou HA ne conduit pas à une plus forte localisation nucléaire de ces protéines. L'ajout d'une séquence de localisation nucléaire (NLS) artificielle à ces protéines ne conduit pas non plus à leur localisation intranucléaire. Il est à noter que la transfection d'un plasmide en cellules conduit à une expression rapide et élevée des protéines à polyG, qui s'agrègent alors en moins de 12 h ; chez l'Homme ou en modèle murin, l'expression de ces protéines est certainement plus faible, conduisant peut-être à une agrégation plus lente, permettant alors aux protéines à polyG de migrer dans le noyau. Cependant, la raison de la localisation nucléaire des polyGlycines reste un mystère. En effet, nous n'avons identifié aucune séquence NLS ou commune entre les protéines FMRpolyG, uN2CpolyG et uGIPC1polyG pouvant expliquer la localisation intranucléaire de ces protéines, suggérant que les résidus glycines induiraient d'eux-mêmes la migration dans le noyau.

A l'inverse des modèles cellulaires, les protéines à polyG sont nucléaires en modèles murins et semblent très toxiques. Il serait intéressant donc de développer une protéine à polyG mutante qui ne migrerait pas dans le noyau, peut-être par l'ajout d'une séquence d'exportation nucléaire (NES), afin de comparer les 2 modèles murins en parallèle.

Enfin, il est intéressant de noter que les patients atteints d'OPDM, en plus des agrégats intranucléaires typiques des maladies à polyG, présentent également des vacuoles cytoplasmiques positives à l'ubiquitine et p62. Les muscles de souris injectées avec un virus

codant pour la protéine uGIPC1polyG développent également de nombreuses inclusions cytoplasmiques positives pour ces l'ubiquitine et p62. De manière intéressante, dans nos souris, ces inclusions sont également positives pour la protéine uGIPC1polyG. Ces résultats obtenus chez la souris questionnent sur la nature des agrégats cytoplasmiques retrouvés chez les patients. Le développement en cours d'anticorps dirigés contre cette protéine permettra de déterminer si, comme chez la souris, la protéine uGIPC1polyG forme des inclusions musculaires cytoplasmiques chez les patients OPDM2.

1.3.5. L'agrégation des protéines à polyG est-elle nécessaire pour leur toxicité ?

Au contraire des séquences entourant les expansions de polyQ qui peuvent moduler l'agrégation des protéines, comme les 17 premiers acides aminés de la protéine HTT-polyQ qui module sa capacité d'agrégation (Zheng et al., 2013), les glycines seules sont capables de s'agréger très rapidement et ceci dès 30 résidus. Cette forte capacité d'agrégation des polyGlycines suppose que les protéines à polyG existent majoritairement sous forme d'inclusions et que la part de forme soluble de ces protéines est probablement faible.

Néanmoins, afin de tester si l'agrégation des protéines à polyG est en corrélation avec la mortalité cellulaire, nous pourrions reproduire des expériences réalisées lors de l'étude des maladies à polyQ. En effet, des analyses entre la dynamique d'agrégation des protéines à polyQ et la mortalité neuronale ont été effectuées en culture primaire de neurones, ce qui a permis de mettre en évidence qu'il s'agissait de la forme soluble des protéines à polyQ qui sont responsables de la toxicité (Saudou et al., 1998 ; Arrasate et al., 2004)

1.3.6. Perte de fonction et / ou gain de fonction ?

Dans les maladies à polyG, ce sont les uORF situées amont des ORF principales des gènes qui sont traduites et qui génèrent la production de « nouvelles » protéines toxiques. Nous avons montré que dans FXTAS, NIID et OPDM2, la traduction de ces uORF avait lieu systématiquement et n'était pas donc dépendante de l'expansion de répétitions CGG. Ces protéines sans expansions de glycines sont petites, instables et donc sujettes à dégradation. Toutefois, nous pouvons nous demander si ces petites protéines codées par les uORF ne possèdent pas de fonction propre. En effet, il est à noter qu'il existe des micro-protéines

fonctionnelles comme la sarcolipine, phospholamban (pour revue, Shaikh et al., 2016), myoréguline (Anderson et al., 2015), Minion (Zhang et al., 2017a) et DWORF (Nelson et al., 2020) exprimées et avec des fonctions importantes pour le muscle squelettique ; ou encore MOTS-c (Lee et al., 2015 ; Kim et al., 2018), mitoréguline (Stein et al., 2018), MIEF1 (Rathore et al., 2018) et PIGBOS (Chu et al., 2019) localisées aux mitochondries et impliquées dans la traduction et le métabolisme. Ces protéines ne sont pas codées par des uORF mais par de petites ORF (smORF) situées majoritairement dans de longs ARN précédemment décrits comme « non codants » (ARNlnc).

Nous pouvons donc nous demander si ces petites protéines générées à partir des uORF ont une fonction. Par exemple, dans FXTAS, ce sont les séquences bordantes de la protéine FMRpolyG qui permettent l'interaction avec la protéine Lap2β (Sellier et al., 2017). De manière similaire, dans NIID, la protéine uN2C (sans expansion de polyGlycine) interagit avec les protéines KU70 et KU80, deux protéines impliquées dans la réparation de l'ADN. Nous avons alors montré que la surexpression de uN2C en cellules U2OS stimulait la réparation de l'ADN après irradiation des cellules, suggérant donc une fonction de cette micro-protéine uN2C. Cependant, nous n'avons pas détecté la présence de la protéine uN2C ni en cellules iPS, ni en organoïdes, questionnant sa stabilité et son niveau d'expression. Néanmoins, l'expression de certaines micro-protéines est régulée par des stress cellulaires spécifiques comme lors d'un stress du réticulum endoplasmique où certaines smORF sont spécifiquement sur-régulées (Martinez et al., 2020). Il est alors possible que la protéine uN2C ne soit stable et détectable qu'en réponse à des dommages à l'ADN.

De manière intéressante, le cortex chez les hominidés est plus complexe et volumineux que celui des autres espèces animales. En effet, durant le développement cortical humain, les protéines NOTCH2NLA et NOTCH2NLB régulent la voie NOTCH permettant la prolifération de progéniteurs neuronaux. Or, qui dit prolifération cellulaire dit dommage à l'ADN. Nous pouvons donc imaginer que le gène *NOTCH2NLC*, situé juste à côté des gènes *NOTCH2NLA* et *NOTCH2NLB*, est un gène en cours d'évolution afin de coder pour une nouvelle protéine permettant de stimuler la réparation de l'ADN lors du développement cérébral.

Enfin, nous n'avons pas observé d'altération au niveau de la réparation de l'ADN chez 8 patients NIID, suggérant qu'un gain de fonction toxique d'une protéine impliquée dans la réparation de l'ADN est peu probable. De manière similaire, les patients FXTAS développent de faibles altérations de l'enveloppe nucléaire. Ces résultats suggèrent donc que même si ces

uORF codent pour des micro-protéines, les expansions de polyGlycines ne provoquent pas un gain de fonction toxique de celles-ci.

1.3.7. Mécanisme de toxicité des protéines à polyG

La similarité des symptômes neurodégénératifs, dont l'ataxie, entre FXTAS et NIID et neuromusculaires, comme la faiblesse et l'atrophie musculaires, entre NIID et l'OPDM laissent penser que les polyGlycines possèdent un mécanisme de toxicité commun. En culture de cellules, les protéines FMRpolyG et uN2CpolyG ont été décrites comme pouvant altérer le transport nucléocytoplasmique via la désorganisation de la lamine nucléaire (Sellier et al., 2017 ; Zhong et al., 2021) ; la protéine FMRpolyG pourrait également altérer le système ubiquitine-protéasome (Oh et al., 2015) et / ou l'intégrité mitochondriale (Gohel et al., 2019). Concernant la protéine uN2CpolyG, le marquage de nombreuses organelles dans des cellules U2OS comme la mitochondrie, le réticulum endoplasmique ou encore l'enveloppe nucléaire, etc. n'a pas révélé d'altération morphologique évidente, à part l'observation de quelques altérations du marquage de la lamine nucléaire. La divergence de ces résultats provient peut-être du type cellulaire utilisé, la plupart des études précédentes étant réalisées en cellules HEK293 ou HeLa alors que nous avons utilisé des cellules ou U2OS. De plus, nous avons développé des lignées U2OS exprimant stablement la protéine uN2CpolyG-GFP contrairement aux autres études qui ont utilisent des stratégies de transfection de plasmides transitoire, pouvant engendrer des altérations de certaines organelles. Enfin, il serait intéressant d'étudier des types cellulaires relevant pour ces maladies, tels que des cellules neuronales et / ou musculaires, notamment différenciées à partir de cellules iPS de patients.

1.3.8. Conclusion

En conclusion, nous pouvons faire un parallèle entre les maladies à polyQ et à polyG. En effet, ces 2 classes de pathologies sont causées par l'expression anormale et toxique de protéines contenant une suite d'acides aminés glutamines ou glycines. Ces protéines forment alors des agrégats détectés dans les zones atteintes des patients touchés par ces différentes maladies.

Tout comme les polyQ, la toxicité des polyG semble être variable et spécifique d'un tissu particulier et d'une maladie à l'autre. En effet, les maladies à polyG sont, à ce jour, caractérisées

par des symptômes neurodégénératifs et / ou neuromusculaires uniquement, bien que des inclusions des la protéine uN2CpolyG soit observée dans de nombreux tissus de patients NIID, suggérant une toxicité particulière des polyGlycines pour les neurones, les astrocytes et les fibres musculaires (Yamaguchi et al., 2018 ; Chen et al., 2020b). De manière intéressante, le niveau d'expression de ces protéines à polyG semble en corrélation avec l'âge d'apparition de ces différentes pathologies. De plus, il serait intéressant de tester si la toxicité des protéines à polyG est dépendante de leur agrégation et de leur localisation, nucléaire ou cytoplasmique, ainsi que déterminer l'importance des séquences bordant les régions répétées de glycines. Enfin, une perte de fonction ou un gain de fonction toxiques des micro-protéines encodées par les uORF n'est, à ce jour, pas à exclure (Figure 59).

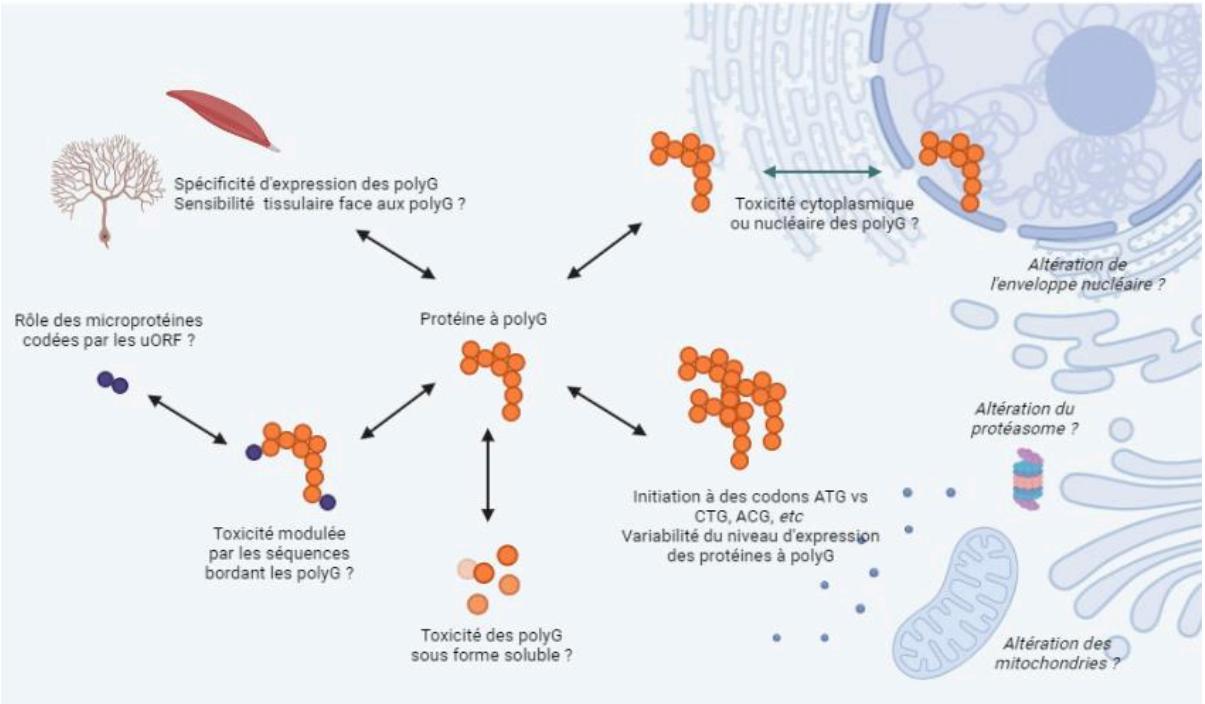


Figure 59. Mécanisme de toxicité des protéines à polyG
La toxicité des protéines à polyG semble toxique pour les muscles et les neurones et pourrait être modulée par leur niveau d'expression, leur localisation, leur agrégation ou encore par la composition des séquences entourant les glycines. Un possible rôle des protéines codées par les uORF n'est pas à exclure. Les protéines à polyG pourraient altérer l'enveloppe nucléaire, le protéasome et / ou les mitochondries. Schéma réalisé sur Biorender.

DISCUSSION

- CHAPITRE 2 -

Expansion de répétitions GGGGCC dans le gène *C9ORF72* dans la SLA/DFT

2. EXPANSION DE REPETITIONS GGGGCC DANS LE GENE *C9ORF72* DANS LA SLA/DFT

La sclérose latérale amyotrophique (SLA) est souvent associée à la démence fronto-temporale (DFT) et la cause génétique la plus fréquente de SLA, de DFT et de SLA/DFT est une expansion de répétitions des nucléotides GGGGCC dans le gène *C9ORF72*. Plusieurs mécanismes non-exclusifs sont proposés pour expliquer la pathogénicité de cette expansion dont (1) des modifications épigénétiques conduisant à une haploinsuffisance de la protéine C9ORF72, impliquée dans l'autophagie ; (2) leur transcription en ARN formant des foci séquestrant potentiellement des protéines liant l'ARN et (3) leur traduction en protéines composées de dipeptides répétés (DPR) potentiellement toxiques et s'agrégeant dans les cellules.

2.1. Synergie de toxicité entre perte de C9ORF72 et protéines DPR

Pendant ma thèse, nous avons pu montrer que la diminution d'expression de la protéine C9ORF72, provoquant une autophagie sous-optimale, conduit à une accumulation des protéines DPR, normalement dégradées par autophagie, augmentant ainsi la toxicité de ces protéines. Ces résultats suggèrent donc un mécanisme de toxicité à double cause des expansions de répétitions GGGGCC (Boivin et al., 2020). Depuis, deux autres études portant sur des motoneurones différenciés à partir de cellules iPS et sur un modèle murin ont confirmé que la déplétion de C9ORF72 exacerbe la toxicité des protéines DPR (Zhu et al., 2020 ; Pal et al., 2021). En effet, le croisement de souris transgéniques porteuses des expansions de répétitions GGGGCC avec des souris KO pour le gène *C9orf72* a permis de mettre en évidence des altérations cognitives et locomotrices associées à une dégénérescence des neurones de l'hippocampe uniquement chez les souris KO ou haploinsuffisantes pour C9orf72, le phénotype s'aggravant avec la perte d'expression de C9orf72 (Zhu et al., 2020). De manière similaire, des motoneurones différenciés à partir de cellules iPS de patients C9-SLA/DFT, où le gène *C9ORF72* a été invalidé, présentent de plus fortes altérations du transport axonal que les motoneurones exprimant la protéine C9ORF72 (Pal et al., 2021). De plus, des modèles murins transgéniques porteurs de l'expansion GGGGCC, mais sans perte d'expression de leur protéine C9orf72 endogène, ne développent pas de neurodégénérescence et cela, bien que l'expression des protéines DPR soit détectée chez ces animaux (O'Rourke et al., 2015 ; Peters et al., 2015 ; Jiang et al., 2016). Enfin, deux études ont décrit des individus avec des expansions de répétitions

GGGGCC dans le gène *C9ORF72* relativement courtes (entre 30 et 90) qui ne développent ni la SLA, ni la DFT. Pourtant, ces patients présentent des inclusions de protéines DPR, mais l'expression de la protéine C9ORF72 n'est pas diminuée. L'ensemble de ces études suggère que la seule expression des protéines DPR est insuffisante pour développer la maladie et que la perte de fonction partielle de la protéine C9ORF72 joue un rôle dans cette pathologie (Gami et al., 2015 ; McGoldrick et al., 2018).

De manière intéressante, il a été montré que la diminution d'expression de C9ORF72 sensibilise les cellules à d'autres stress cellulaires que l'expression des protéines DPR. Par exemple, il a été montré que la déplétion de la protéine C9ORF72 engendrait également l'accumulation et une plus forte toxicité de la protéine ATXN2 avec une taille intermédiaire d'expansion de polyGlutamine (30x) (Sellier et al., 2016). En effet, une expansion de répétitions de nucléotides CAG supérieures à 35 éléments dans le gène *ATXN2* est responsable de l'ataxie spinocérébelleuse de type 2 (SCA2), générant une protéine ATXN2-polyQ toxique. Au contraire, une taille intermédiaire de 24 à 33 répétitions de tri-nucléotides CAG dans l'ATXN2 est un facteur de risque de la SLA (Elden et al., 2010 ; Daoud et al., 2011 ; Sproviero et al., 2017), et cette mutation a été observée chez des patients porteurs d'expansions de répétitions GGGGCC dans le gène *C9ORF72* (Zhang et al., 2017b), ou de mutations dans les gènes *TARDP*, *SOD1* ou *FUS* (Farg et al., 2013 ; Chiò et al., 2015) mais aussi chez des patients sporadiques (Daoud et al., 2011). De plus, les neurones différenciés à partir de cellules iPS de patients C9-SLA/DFT sont plus sensibles à des molécules inhibant l'autophagie (Almeida et al., 2013) ou au glutamate (Shi et al., 2018). Ces données suggèrent que l'haploinsuffisance de la protéine C9ORF72 sensibilise les différents modèles (cellulaires et murins) à des stress variés, soulevant la question des fonctions moléculaires / cellulaires de cette protéine.

2.2. Rôles de la protéine C9ORF72

2.2.1. Rôle de la protéine C9ORF72 au lysosome

Durant ma thèse, je me suis intéressée à la fonction du complexe C9ORF72-SMCR8-WDR41 (CSW) dans l'autophagie (Figure 60A). Toutefois, le développement de cellules U2OS KO pour le gène *C9ORF72* et l'analyse du flux autophagique ne m'a pas permis d'observer une accumulation de p62, un marqueur classique de l'autophagie. J'ai toutefois pu noter une légère diminution de la lipidation de LC3b, décrite dans de nombreuses autres études (Figure 60B, 60C ; Sellier et al., 2016 ; Sullivan et al., 2016 ; Ugolino et al., 2016 ; Webster et al., 2016 ; Jung et al., 2017). J'ai alors confirmé ces altérations mineures de l'autophagie dans des cellules HEK293 KO pour le gène *C9ORF72*, ainsi que dans des cellules neuronales immortalisées Neuro-2A et GT1-7 où la protéine C9orf72 a été déplétée transitoirement par siARN. La diminution d'expression du complexe CSW ne semble donc pas conduire à une dysfonction majeure de l'initiation de l'autophagie dans ces différents modèles cellulaires.

Le complexe CSW étant localisé aux lysosomes (Amick et al., 2016), je me suis alors intéressée à ces organelles. L'analyse du nombre, de la taille et de l'acidité des lysosomes dans les cellules U2OS KO pour le gène *C9ORF72*, mais également dans des cellules GT1-7 déplétées de la protéine C9orf72 par siARN, ne m'a pas permis de mettre en évidence de dysfonction évidente (Figure 60D, 60E, 60F, 60G). Néanmoins, ces expériences ont été réalisées en condition basale et il serait intéressant de tester si le nombre, la taille et l'acidité des lysosomes est modifiée pendant une autophagie active grâce à une privation en acides aminés ou à des inhibiteurs de mTOR tels que la torine ou le WYE-354.

En parallèle, je me suis aussi intéressée à la localisation du complexe CSW. Mes premiers résultats de transfection cellulaire ne m'ont pas permis de déterminer une localisation précise de C9ORF72 ou SMCR8. Toutefois, nous avons pu mettre récemment en évidence que la protéine WDR41 permet de recruter C9ORF72 et SMCR8 aux lysosomes après induction de l'autophagie par restriction en acides aminés, mais également lors de l'autophagie induite par des molécules inhibant mTOR comme le WYE-354 ou la torine (Figure 60H). Il est à noter que ces résultats sont en accord avec les résultats de l'équipe du Dr Ferguson (Amick et al., 2018).

L'ensemble de mes données suggèrent donc que le complexe CSW n'intervient probablement pas lors de l'induction de l'autophagie, mais pourrait agir dans des étapes tardives de l'autophagie, notamment au niveau des lysosomes et de leur reformation. En effet, lors d'une

autophagie prolongée comme une privation en acides aminés de plusieurs heures, la plupart des lysosomes ont fusionné avec des autophagosomes, formant alors des auto-lysosomes. La cellule crée alors de nouveaux lysosomes par un mécanisme appelé la reformation du lysosome après autophagie (ALR) où de nouveaux lysosomes se forment à partir des auto-lysosomes. Ce phénomène est dû à la réactivation de mTOR, qui était inhibé pendant l'autophagie. Ce mécanisme ALR est caractérisé par la formation de tubules positifs pour la protéine LAMP1, formés à partir des auto-lysosomes puis, qui sont clivés par la Dynamine-2 pour former de nouveaux lysosomes (Yu et al., 2010). Le complexe CSW étant localisé aux lysosomes mais ne semblant pas fortement altérer l'autophagie classique, ni le nombre, la taille et l'acidité des lysosomes, nous nous sommes demandé s'il n'était pas impliqué dans ce mécanisme de reformation des lysosomes. Pour répondre à cette question, nous avons développé des cellules U2OS WT ou KO pour le gène *C9ORF72* qui expriment stablement la protéine LAMP1-mCHERRY, permettant ainsi de suivre en direct la dynamique des lysosomes. Après 8 h de privation en acides aminés, les cellules KO pour le gène *C9ORF72* présentent moins de tubules de lysosomes que les cellules WT (Figure 60I, 60J). Ces résultats suggèrent donc que le complexe CSW est impliqué dans la reformation des lysosomes. Toutefois, le rôle exact de la protéine C9ORF72 dans ce mécanisme reste à éclaircir. Le complexe CSW jouant le rôle de GAP pour des petites GTPases (Su et al., 2020 ; Tang et al., 2020), il serait donc intéressant de tester s'il régule une petite GTPase localisée aux lysosomes, comme les protéines Rag et Rheb, qui sont impliquées dans la localisation et l'activation de mTOR aux lysosomes (Sancak et al., 2010) ; ou la protéine ARL8, qui joue un rôle dans la localisation et la mobilité des lysosomes ; ou même la protéine Rab7, qui est impliquée dans la migration et la fusion des autophagosomes aux lysosomes (Mrakovic et al., 2012).

Figure 60. Implication du complexe CSW dans l'autophagie
(**A**) Schéma de l'autophagosome en formation puis, fusionnant au lysosome. Schéma fait sur Biorender. (**B**) Western blot dirigé contre LC3b après induction de l'autophagie à la fluphénazine pendant 16h dans des cellules U2OS WT ou KO pour le gène *C9ORF72* et, (**C**) sa quantification. (**D**) Cellules GT1-7 déplétées ou non de C9orf72 par siARN où les lysotracker et lysensor ont été ajoutés afin de visualiser les lysosomes et quantification de leur (**E**) acidité, (**F**) nombre et (**G**) taille. (**H**) WDR41-GFP est localisé aux lysosomes après induction de l'autophagie par le WYE-354 pendant 8h. (**I**) Observation des tubules de lysosomes grâce au marquage LAMP1-mCHERRY après 8h de privation en acides aminés dans des cellules U2OS WT ou KO pour le gène *C9ORF72* et (**J**) quantification de leur nombre. n=3, t-test *p-valeur<1%, ***p-valeur<0,01%.

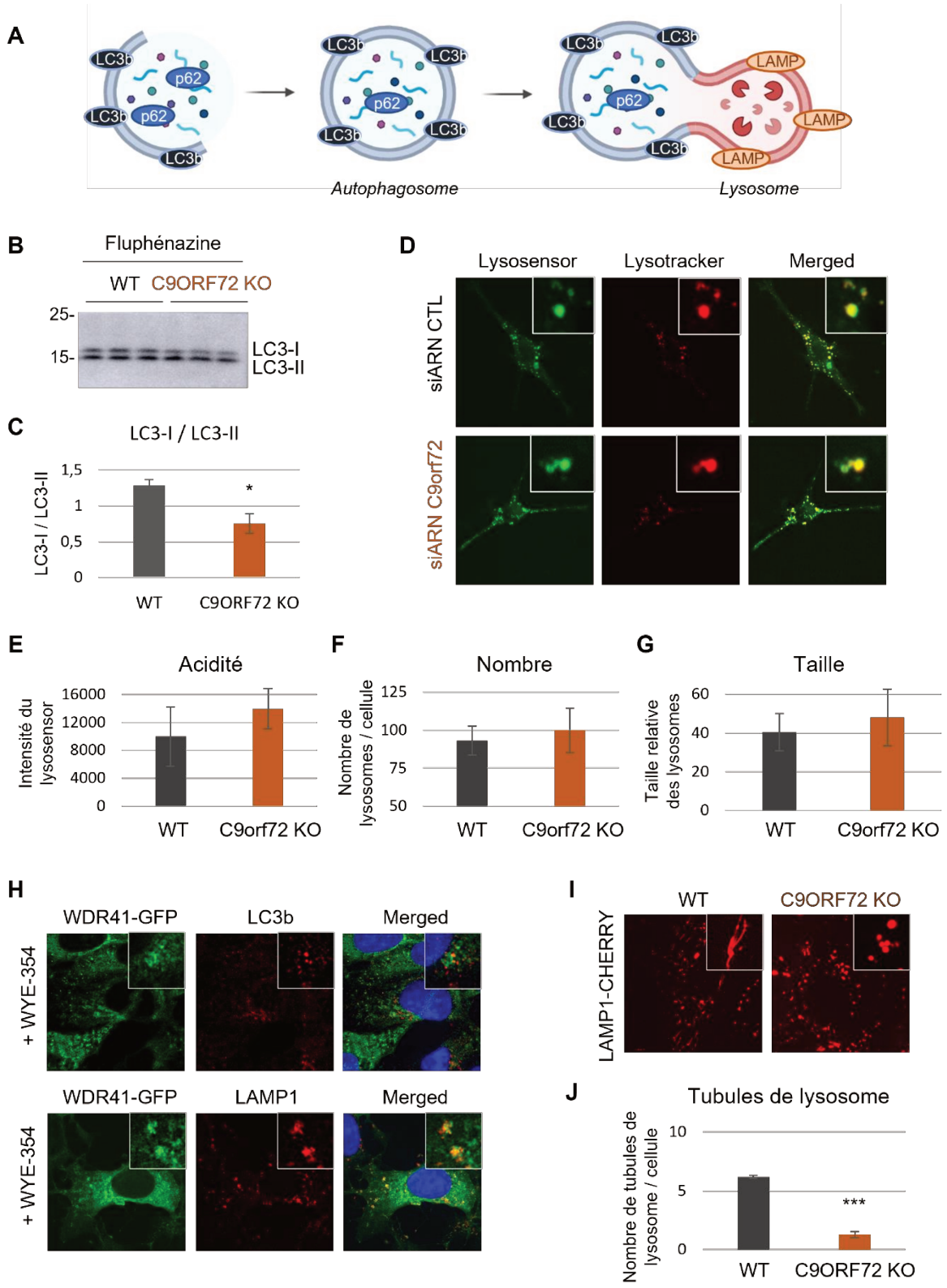


Figure 60. Implication du complexe CSW dans l'autophagie
Voir légende page précédente.

2.2.2. Rôle de la protéine C9ORF72 aux LDCV

Les résultats présentés précédemment ont été obtenus grâce à la surexpression des protéines du complexe CSW taguées. Afin de confirmer ces données sur les protéines endogènes, nous avons développé des anticorps dirigés contre les protéines C9ORF72 et SMCR8 qui ont permis de confirmer leur présence aux lysosomes (Frick et al., 2018). De plus, nous avons observé que les protéines SMCR8 et C9ORF72 sont localisées au niveau de vésicules, que nous avons alors identifié comme des LDCV pour « Large Dense Core Vesicle » (Figure 61A). Nous avons confirmé ces observations par immunofluorescence et immunoblot dans différents types cellulaires dont des cellules neuroendocrines de souris GT1-7, ainsi que dans des cultures primaires de motoneurones de souris et dans des motoneurones différenciés à partir de cellules iPS (Figure 61B, 61C, 61D). Ces vésicules permettent de stocker et de libérer des neuropeptides et des hormones des cellules chromaffines et les neurones. La formation des LDCV fait intervenir plusieurs étapes. Tout d'abord, des granules de sécrétion immatures (ISG) sont formées à partir de Golgi puis, elles fusionnent entres elles et sont enrichies en pompes à protons ATPase-V, ce qui permet leur acidification pendant leur transport jusqu'à la membrane plasmique (Figure 61A). Les LDCV sont alors matures et fonctionnelles (Wu et al., 2001).

Outre la localisation de C9ORF72 et SMCR8 aux LDCV, nous avons également montré que la diminution d'expression de la protéine C9ORF72 par siARN conduit à une diminution du nombre de LDCV matures, mais n'affecte pas leur fusion à la membrane plasmique dans des cellules endocrines PC12 (Figure 61E, 61F, 61G, 61H). Néanmoins, la souris KO pour le gène *C9orf72* ne présente pas de diminution du niveau de Chromogranine A (ChGA), un marqueur des LDCV, dans différents tissus. Ces résultant soulèvent la question d'une compensation fonctionnelle du complexe CSW *in vivo*. Chez les patients atteints de SLA, une diminution du nombre de vésicules de ChGA est observée dans les neurones (Schiffer et al., 1995). Enfin, la surexpression de la ChGA chez un modèle murin porteur d'une mutation dans le gène *SOD1* exacerbe le phénotype locomoteur, suggérant que la ChGA pourrait être un modulateur de la pathologie (Ezzi et al., 2010). De plus, une mutation P413L dans le gène *ChGB*, codant pour la chromogranine B, une protéine ayant une fonction similaire à la ChGA, est un facteur de risque de développer la SLA et un modificateur de l'âge d'apparition de la pathologie (Gros-Louis et al., 2009), bien que ces observations soient controversées (Blasco et al., 2011).

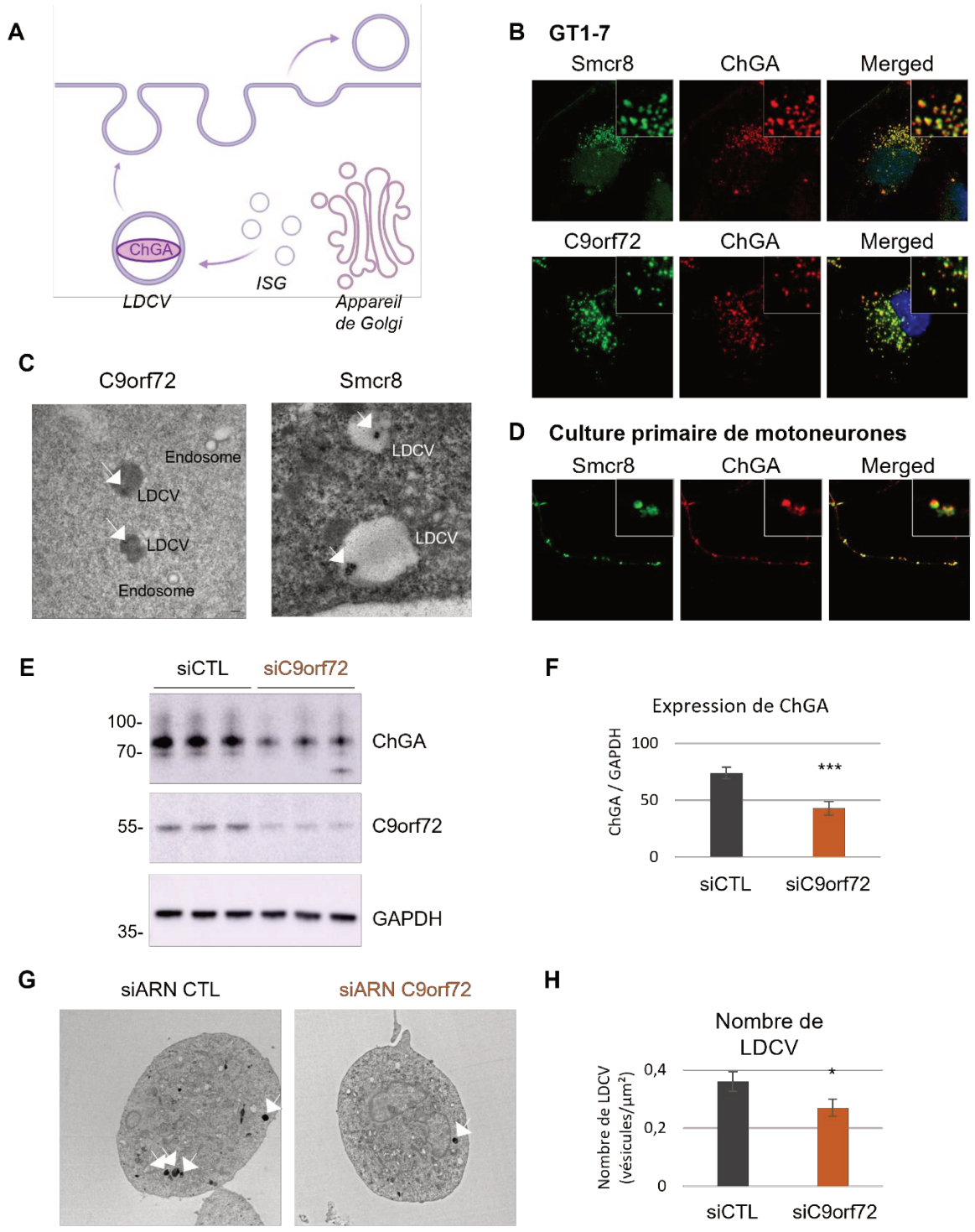


Figure 61. Rôle du complexe CSW pour les vésicules LDCV

(**A**) Schéma de la formation des LDCV. Schéma fait sur Biorender. (**B**, **C**, **D**) Localisation de C9orf72 et Smcr8 aux LDCV dans (**B**, **C**) des cellules GT1-7 et (**D**) des cultures primaires de motoneurones. (**C**, **D**) Immunofluorescence dirigée contre ChGA et C9orf72 ou Smcr8. (**C**) Localisation aux LDCV par microscopie électronique de C9orf72 et de Smcr8 détectées par un immunomarquage couplé à des billes d'or. (**E**) Western blot dirigé contre ChGA, C9orf72 et GAPDH après déplétion ou non de C9orf72 par siARN dans des cellules PC12 et (**F**) quantification de l'expression de ChGA. (**G**) LDCV observées dans des cellules PC12 par microscopie électronique après déplétion ou non de C9orf72 par siARN et (**H**) quantification du nombre de LDCV. n=3, t-test *p-valeur<1%, ***p-valeur<0,01%.

De manière intéressante, de grandes similitudes existent entre les LDCV et les lysosomes dont (1) l'acidité de ces organelles ; (2) la composition en protéines telles que les procathépsines B et L (Kuliawat et al., 1997), qui permettent la maturation des neuropeptides dans les LDCV et la dégradation des protéines dans le lysosome (pour revue, Funkelstein et al., 2010 ; Barrett et Kirschke, 1981) et ; (3) leur fusion possible à la membrane plasmique pour libérer le cargo des LDCV et réparer la membrane plasmique pour les lysosomes (Reddy et al., 2001). Nous pouvons nous demander si le complexe CSW ne joue pas un rôle au niveau de différents compartiments cellulaires acides, tels que le lysosome et les LDCV, et ayant une fonction plus large qu'uniquement l'autophagie. Cette hypothèse est appuyée par d'autres études reportant un rôle de la protéine C9ORF72 dans d'autres voies de signalisation cellulaire, notamment l'exocytose. En effet, il a été reporté que le complexe CSW interagit et colocalise avec la famille des Rab GTPases Rab3, impliquées dans la formation et la fusion des vésicules synaptiques (Frick et al., 2018). De plus, la déplétion de C9ORF72 dans le poisson zèbre conduit à une diminution du nombre de vésicules positives pour Rab3 (Butti et al., 2021).

Enfin, il reste à déterminer l'impact physiopathologique de la diminution d'expression de C9ORF72 sur ces différentes vésicules acides dans des modèles cellulaires et animaux relevant pour l'étude de la SLA/DFT tels que des motoneurones différenciés à partir de cellules iPS de patients C9-SLA/DFT et des poissons zèbres ou des souris déplétés de C9ORF72.

2.3. La toxicité des protéines DPR

Nous avons montré que l'expansion de répétitions GGGGCC est toxique via 2 mécanismes agissant en synergie : (1) une diminution d'expression de la protéine C9ORF72 conduit à une diminution de l'autophagie et ; (2) la traduction des répétitions GGGGCC en protéines DPR, notamment dans les cadres de lecture polyGlycine-Alanine (polyGA) et polyGlycine-Proline (polyGP), formant des agrégats potentiellement toxiques et normalement dégradés par autophagie. Mes expériences ayant été réalisées en modèle cellulaire, nous avons voulu confirmer cette synergie de toxicité entre perte de C9ORF72 et expression des protéines DPR polyGA et polyGP dans un modèle animal.

Il est à noter que les souris avec un KO complet du gène *C9orf72* développent un phénotype inflammatoire conduisant à une mort précoce de ces animaux (Atanasio et al., 2016 ; Burberry et al., 2016 ; Jiang et al., 2016 ; O'Rourke et al., 2016 ; Sudria-Lopez et al., 2016). Afin de parer ce problème, nous avons généré une lignée murine *C9orf72* LoxP x Nestin-Cre que j'appellerai par la suite *C9orf72* SNC-KO, où l'expression de C9orf72 est déplétée uniquement dans le système nerveux central. Afin d'exprimer les protéines DPR polyGA et polyGP, nous avons alors opté pour une approche virale via des AAV PHP.eB, ayant la propriété de passer la barrière hémato-encéphalique des souris après une simple injection intraveineuse, permettant ainsi l'expression de notre protéine DPR d'intérêt dans le système nerveux central de ces animaux. Nous avons alors injecté des souris WT ou *C9orf72* SNC-KO avec des virus codant soit pour la GFP en contrôle, soit pour la protéine polyGA, soit pour la protéine polyGP. Enfin, il est à noter que nous avons cloné les expansions GGGGCC et CCCCGG entourées de leurs séquences naturelles et non sous un codon ATG artificiel.

2.3.1. Toxicité de la protéine DPR polyGA

L'analyse du phénotype locomoteur des souris surexprimant la protéine polyGA (Figure 62A) a mis en évidence une altération des capacités motrices des souris *C9orf72* SNC-KO injectées en comparaison avec les souris WT, et ceci à partir de 9 mois post-injection, notamment grâce au test de l'open field et du rotarod (Figure 62B, 62C). L'analyse histopathologique montre que de nombreux agrégats de la protéine polyGA sont observés dans le cerveau des souris, que celles-ci soient contrôles ou *C9orf72* SNC-KO (Figure 62D). Ces résultats préliminaires sont importants car ils confirment notre hypothèse d'un mécanisme de synergie toxique entre protéines DPR et perte de la protéine C9ORF72, ce qui pourrait expliquer la mort des neurones dans la SLA. Dans le but de confirmer ces données, j'ai injecté récemment un plus grand nombre d'animaux WT et *C9orf72* SNC-KO, soit avec un virus codant pour la GFP, soit avec un virus codant pour la protéine polyGA et l'étude phénotypique de ces animaux est en cours.

D'un point de vue mécanistique, il a été montré que la diminution d'expression de la protéine C9ORF72 conduit à une autophagie sous-optimale (Sellier et al., 2016 ; Sullivan et al., 2016 ; Ugolino et al., 2016 ; Webster et al., 2016 ; Jung et al., 2017 ; Shi et al., 2018 ; Shao et al., 2020). De plus, l'expression de la protéine polyGA altère la fonction du protéasome (May et al., 2014 ; Schludi et al., 2015 ; Zhang et al., 2016 ; Schludi et al., 2017 ; Guo et al., 2018 ; Riemslagh et al., 2019 ; Zhang et al., 2021). Ces résultats suggèrent une altération des 2 mécanismes cataboliques principaux de la cellule pouvant ainsi, peut-être, expliquer la synergie de toxicité et la mort neuronale dans la SLA/DFT. En effet, les neurones sont particulièrement sensibles à l'accumulation de protéines toxiques ou d'organelles dysfonctionnels puisqu'ils ne peuvent pas les éliminer lors de la division cellulaire (pour revue, Damme et al., 2015).

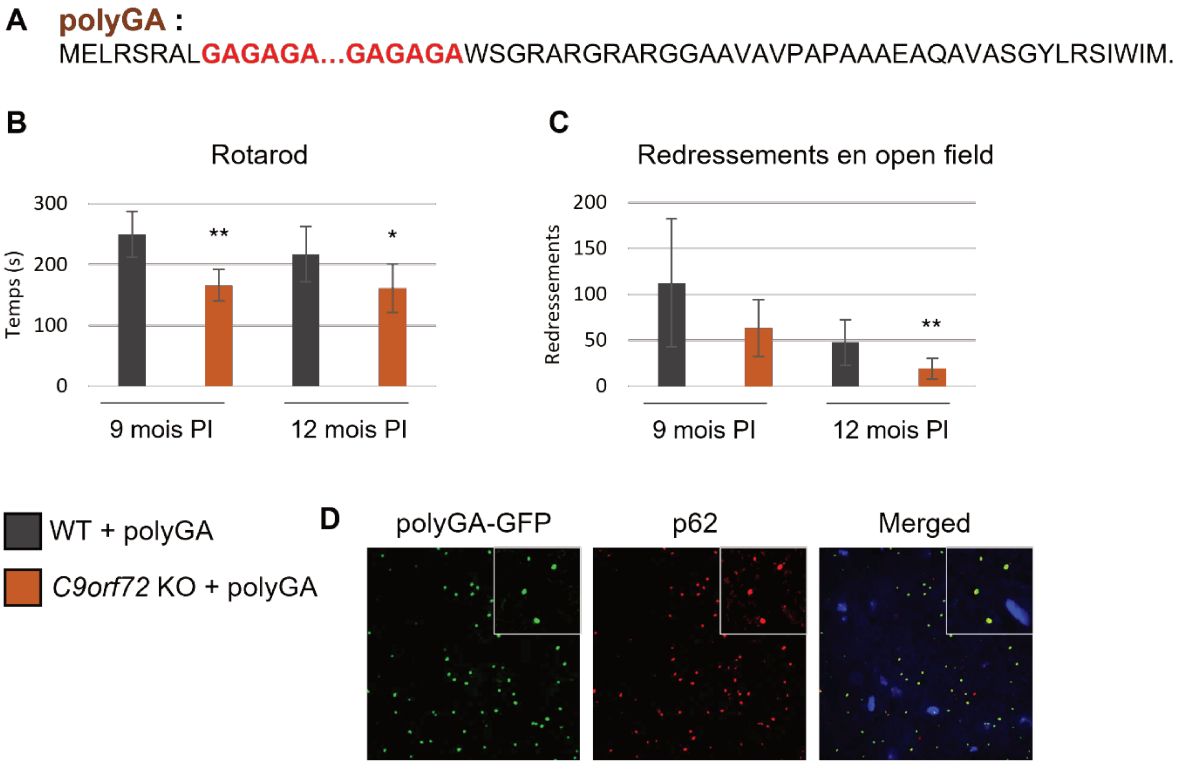


Figure 62. Phénotype et histopathologie des souris WT et *C9orf72* SNC-KO exprimant la protéine polyGA

(**A**) Séquence peptidique de la protéine polyGA. (**B**, **C**) Etude des capacités locomotrices des souris WT (noir) et *C9orf72* SNC-KO (brun) injectées avec le virus codant pour la protéine polyGA à 9 et 12 mois après l'injection. (**B**) Temps passé par les souris sur le rotarod avant la chute. (**C**) Nombre de redressements des souris pendant durant le test de l'open field. (**D**) Immunofluorescence réalisée sur le cerveau d'une souris WT injectée avec le virus codant pour la protéine polyGA sacrifiée 12 mois post-injection révélant la présence d'agrégats protéiques formés par la protéine polyGA positifs à p62. n=3 souris/groupe, t-test *p-valeur<10%, **p-valeur<5%.

2.3.2. Toxicité de la protéine DPR polyGP

De manière étonnante, les souris injectées avec le virus codant pour la protéine polyPG ont développé rapidement des altérations locomotrices, associées à une nette diminution de la survie de ces souris, et ceci qu'elles soient WT ou *C9orf72* SNC-KO (Figure 63A). Ces résultats mettent en lumière une forte toxicité de la protéine polyGP, qui est pourtant reportée comme peu toxique dans la littérature (Mizielinska et al., 2014, Wen et al., 2014 ; Freibaum et al., 2015 ; Lee et al., 2016). Néanmoins, il est à noter que les études précédentes ont été réalisées uniquement avec une protéine polyGP exprimée sous un ATG artificiel et donc tronquée. En effet, nous avons montré que l'initiation de la traduction de la protéine polyGP avait lieu en amont des répétitions et donc conduisant à l'expression d'une protéine polyGP avec une partie N-terminale non décrite précédemment. L'analyse de cette séquence N-terminale révèle la présence de nombreux acides aminés chargés positivement, notamment des résidus arginines, connus pour leur toxicité. En effet, les protéines DPR riches en arginine polyGR et polyPR sont décrites comme étant très toxiques dans différents modèles cellulaires et animaux (Mizielinska et al., 2014 ; Wen et al., 2014 ; Freibaum et al., 2015 ; Tao et al., 2015 ; Lee et al., 2016 ; Zhang et al., 2018 ; Hao et al., 2019 ; Cook et al., 2020 ; Li et al., 2020 ; Maor-Nof et al., 2021).

Nous nous sommes donc demandé si cette nouvelle séquence N-terminale pouvait expliquer la toxicité de la protéine polyGP. Pour répondre à cette question, nous avons développé un nouveau virus où la protéine polyGP est délétée de sa séquence N-terminale (ΔNter-polyGP), que nous avons alors injecté une nouvelle cohorte de souris WT (Figure 63B). Il est intéressant de noter que les souris exprimant la protéine ΔNter-polyGP ne développent pas d'altérations locomotrices évidentes et ne meurent pas prématurément, contrairement aux souris exprimant la protéine polyGP entière (Figure 63C, 63D, 63E, 63F, 63G). Ces résultats suggèrent que la suite de Glycine-Proline n'est pas toxique d'elle-même, ce qui est en accord avec la littérature, mais que la séquence N-terminale riche en arginines de cette protéine est responsable de sa toxicité. Afin de confirmer ces résultats, j'ai récemment injecté une nouvelle cohorte avec un nouveau virus codant uniquement pour la séquence N-terminale de la protéine polyGP, sans les répétitions de Glycine-Proline.

Ces résultats sur la protéine polyGP sont intéressants puisqu'ils suggèrent que cette protéine DPR pourrait faire partie des protéines DPR riches en arginine, comme les protéines polyGR et polyPR, expliquant ainsi sa toxicité. De plus, sa plus forte abondance chez les patients pourrait expliquer la mort des neurones dans la SLA. Toutefois, ces résultats préliminaires sont à confirmer.

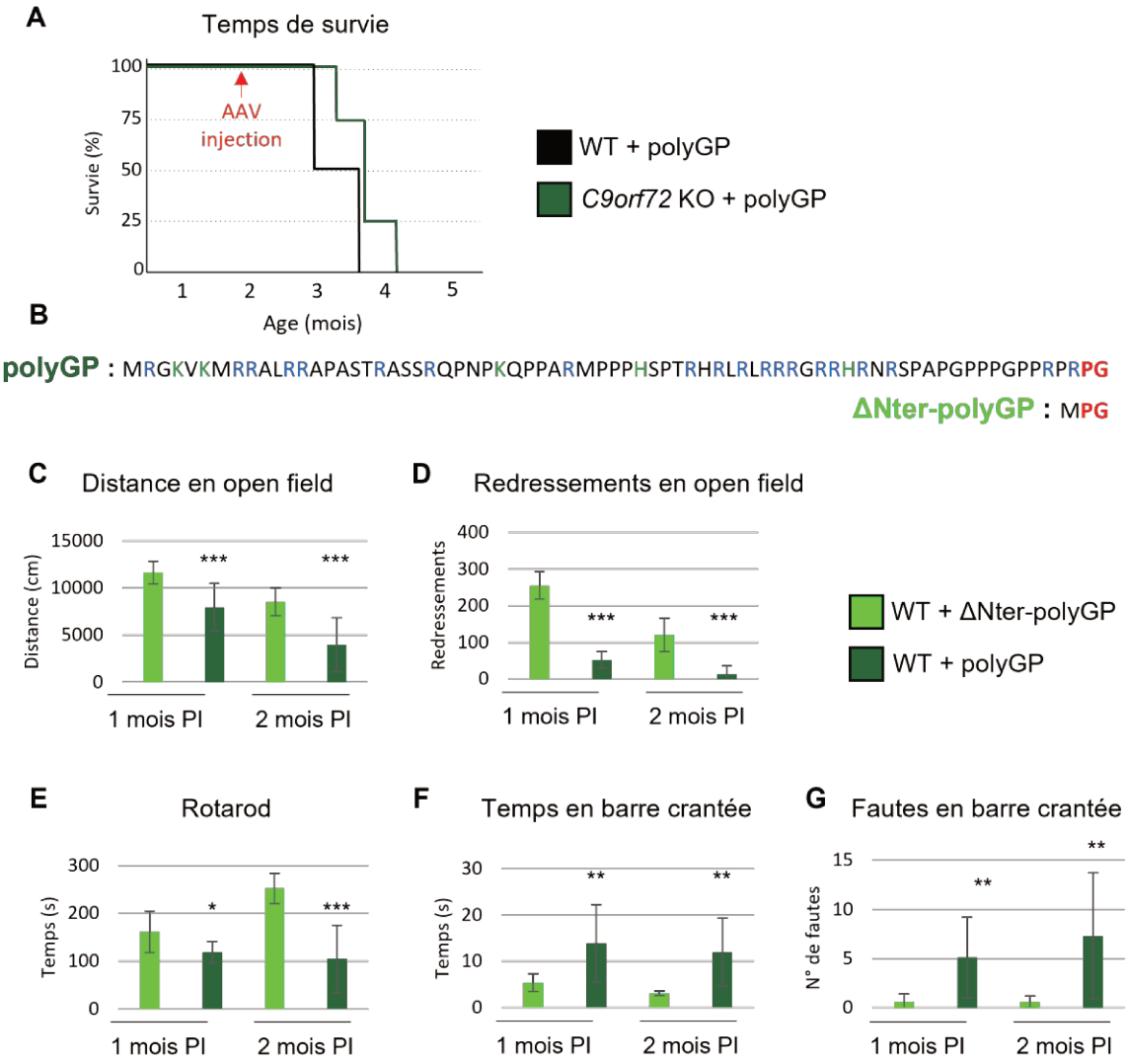


Figure 63. Phénotype locomoteur des souris exprimant la protéine polyGP
(**A**) Courbe de survie des souris WT (noir) et *C9orf72* SNC-KO (vert foncé) injectées avec le virus codant pour la protéine polyGP. (**B**) Composition N-terminale des protéines polyGP (**C-H**) Etude des capacités locomotrices des souris WT injectées avec 8.10^{12} vg/kg de virus AAV/PHP.eB codant pour la protéine polyGP entière (vert foncé) ou délétée de sa partie N-terminale (ΔNter-polyGP) (vert clair) à 1 et 2 mois post-injection. (**C**, **D**) Résultats de l'open field avec (**C**) la distance parcourue et (**D**) le nombre de redressements. (**E**) Temps passé par les souris sur le rotarod avant la chute. (**F**, **G**) Résultats de la barre crantée avec (**F**) le temps de traversée et (**G**) le nombre de fautes. n=6 souris/groupe, t-test *p-valeur<10%, **p-valeur<5%, ***p-valeur<1%..

2.3.3. Détection des protéines DPR chez les patients C9-SLA/DFT

Parallèlement à cette étude en modèle murin, nous voulions développer de nouveaux outils pour suivre l'expression des protéines DPR polyGA et polyGP, notamment par la production d'anticorps dirigés spécifiquement contre ces protéines. En effet, actuellement, les anticorps disponibles sont dirigés uniquement contre les stretchs d'acides aminés répétés (Glycine-Alanine ou Glycine-Proline). Ces anticorps sont donc susceptibles de reconnaître d'autres protéines contenant ces séquences répétées. Ayant identifié les parties N-terminales des protéines polyGA et polyGP, nous avons donc souhaité développer de nouveaux anticorps dirigés contre les acides aminés en amont des dipeptides répétés. Pour cela, nous avons produit et purifié des protéines recombinantes exprimant ces régions N-terminales, puis le service anticorps de l'IGBMC a injecté ces protéines chez des souris afin de produire des anticorps monoclonaux. Nous avons récemment obtenu des clones d'anticorps et plusieurs de ces clones semblent spécifiques des séquences N-terminales des protéines polyGA et polyGP en cellules transfectées (Figure 64A, 64B) ainsi qu'en tissu du modèle murin surexprimant la protéine polyGA-GFP (Figure 64C). Nous souhaitons désormais tester ces anticorps sur des lames de cerveaux de patients atteints de SLA/DFT afin d'attester d'une initiation de la traduction en amont des répétitions et d'étudier la localisation des protéines DPR. Enfin, la protéine polyGP étant retrouvée dans le liquide cérébrospinal des patients C9-SLA/DFT (Gendron et al., 2017 ; Lehmer et al., 2017 ; Meeter et al., 2018), des anticorps dirigés contre la partie N-terminale de cette protéine pourraient être utilisés pour le diagnostic précoce de ces patients.

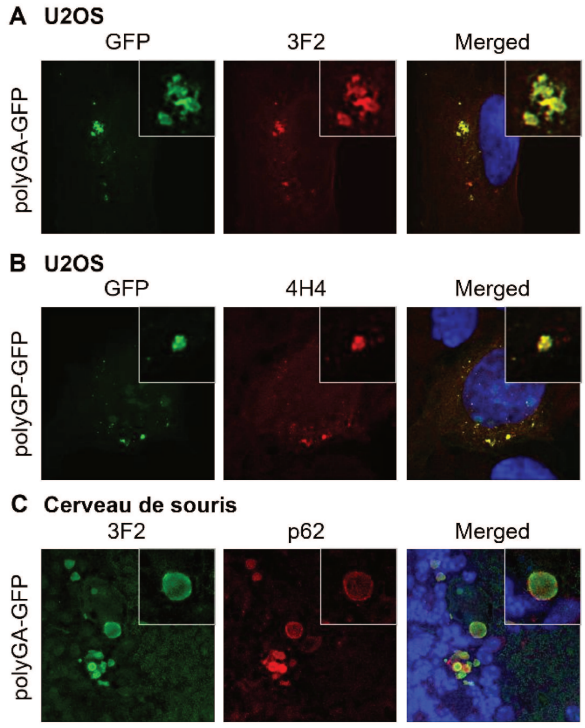


Figure 64. Anticorps dirigés contre les protéines polyGA et polyGP
Immunofluorescences dirigées contre les protéines (A, B) GFP et (A) polyGA ou (B) polyGP dans des cellules U2OS transfectées avec un plasmide codant pour la (A) protéine polyGA-GFP ou (B) protéine polyGP-GFP ou ; (C) contre les protéines polyGA et p62 sur une lame de cerveau de souris exprimant la protéine polyGA-GFP

DISCUSSION

- CHAPITRE 3 -

Traitements

3. TRAITEMENTS

3.1. Thérapie par oligonucléotides antisens (ASO)

A l'heure actuelle, aucun traitement efficace contre les maladies à polyG, FXTAS, NIID et OPDM, et contre la SLA/DFT n'existe. La cause des maladies à polyG étant l'expression de protéines similaires, toutes composées de polyGlycine, le ciblage de cette région polyGlycine commune pourrait s'avérer une piste thérapeutique prometteuse. Il a été montré récemment qu'un ASO dirigé contre les répétitions de nucléotides CGG de l'ARN codant pour la protéine FMRpolyG diminue le nombre d'agrégats formés par la protéine FMRpolyG et améliore le phénotype locomoteur des souris (Derbis et al., 2021), suggérant qu'un tel ASO pourrait également être utilisé dans NIID et OPDM. Concernant les maladies affectant le système nerveux central, FXTAS et NIID, l'injection de l'ASO pourrait être réalisé par voie intrathécale qui diffuserait alors via le liquide cérébrospinal, tout comme pour le traitement de la SMA. Pour les maladies affectant directement le système musculaire comme l'OPDM et, dans une moindre mesure NIID, l'injection de l'ASO pourrait se faire par voie intraveineuse, comme pour le traitement de la dystrophie de Duchenne avec l'Eteplirsen (NCT02255552). Néanmoins, l'utilisation d'un ASO bloquant la traduction de l'ARN conduit à une perte de fonction de la protéine. Or, la perte de la protéine FMRP, codée par le même ARN que FMRpolyG, conduit à FXS, laissant penser que cette stratégie pourrait se révéler risquée. Au contraire, l'haploinsuffisance des protéines NOTCH2NLC, LRP12 et GIPC1 n'est pas actuellement reportée comme étant responsable d'une pathologie. En effet, des membres de la famille de patients NIID et OPDM2 possédant un nombre de répétitions CGG supérieur à 200, conduisant respectivement à la perte d'expression des protéines NOTCH2NLC et GIPC1 ne développent pas de symptômes particuliers (Deng et al., 2020 ; Fukuda et al., 2021). Ces données suggèrent qu'un ASO ciblant les expansions de répétitions CGG pourrait être testé dans NIID et l'OPDM, mais pourrait être une stratégie risquée dans le cas de FXTAS.

Concernant la SLA/DFT due à une expansion de répétitions GGGGCC dans le gène *C9ORF72*, plusieurs essais cliniques utilisant un ASO dégradant l'ARN sens contenant les répétitions sont en cours (NCT03626012). Nous avons montré que la protéine polyGP était principalement codée par l'ARN antisens composé de répétitions CCCCGG, qui n'est pas ciblée dans ces essais. L'initiation de la traduction de la protéine polyGP ayant lieu en amont des répétitions, cette protéine possède une séquence N-terminale particulière qui pourrait alors être ciblée par

un ASO. Nous pourrions tester cette stratégie dans nos souris injectées avec un virus AAV/PHP.eB codant pour la protéine polyGP.

3.2. Thérapie par molécules activant l'autophagie

De nombreuses pathologies, telles que la maladie d'Alzheimer, la maladie de Parkinson, les maladies à polyQ, les maladies à polyG, la SLA, la DFT etc. sont des protéinopathies, c'est-à-dire des maladies caractérisées par l'accumulation de protéines sous forme d'agrégats. Ces inclusions protéiques sont sensibles à l'autophagie, suggérant que la stimulation de ce processus de dégradation pourrait être une piste thérapeutique intéressante. Nous avons montré que l'expression des protéines DPR polyGA et polyGP, et de la protéine uN2CpolyG, est diminuée après stimulation de l'autophagie avec la torine, un inhibiteur de mTOR, et avec la fluphénazine, la chlorpromazine et la prométhazine, des composés approuvés par la FDA mais avec une forte action neuroleptique. Il serait intéressant de tester si ces molécules améliorent le phénotype locomoteur des souris exprimant les protéines DPR polyGA et polyGP ou uN2CpolyG. Néanmoins, leur utilisation en clinique reste compliquée puisque ces composés présentent un fort effet sédatif.

DISCUSSION

- CHAPITRE 4 -

Mécanismes de traduction des répétitions

4. MECANISMES DE TRADUCTION DES REPETITIONS

Ma thèse a porté sur l'étude des mécanismes de toxicité dans différentes maladies à expansion de répétitions, la SLA/DFT, NIID et OPDM2, et notamment sur la traduction de ces répétitions. En effet, nous avons montré que les expansions de répétitions GGGGCC situées dans le premier intron du gène *C9ORF72* et CGG dans les régions 5'UTR des gènes *NOTCH2NLC* et *GIPC1*, bien que situées dans des régions décrites comme « non-codantes », sont traduites en protéines DPR polyGA et polyGP dans la SLA ; et en protéines composées de polyGlycine dans NIID et l'OPDM2.

4.1. Mécanisme de traduction des protéines DPR dans la SLA/DFT

L'expansion de répétitions GGGGCC dans le gène C9ORF72 peut être traduite en 5 protéines DPR différentes selon le cadre de lecture ouvert : les protéines polyGA et polyGR à partir de l'ARN sens ; les protéines polyPA et polyPR à partir de l'ARN antisens et ; la protéine polyGP pouvant être traduite à partir de 2 ARN. Néanmoins, ces protéines ne sont pas retrouvées en abondance égale chez les patients. En effet, la protéine polyGA est la protéine la plus largement retrouvée, suivie des protéines de polyGP, de polyGR, de polyPA puis de polyPR (Mackenzie et al., 2013).

Nous avons montré que l'initiation de la traduction de la protéine polyGA avait lieu à un codon CTG situé en amont des répétitions GGGGCC, un résultat confirmé par plusieurs autres études (Green et al., 2017 ; Sonobe et al., 2018 ; Tabet et al., 2018 ; Almeida et al., 2019 ; Sonobe et al., 2021). La protéine polyGP est la seule protéine DPR pouvant être générée à partir de l'ARN sens composé de répétitions GGGGCC et de l'ARN antisens CCCCGG. Il a tout d'abord été montré que cette protéine polyGP était produite via un changement du cadre de lecture ouvert du ribosome passant du cadre de lecture Glycine-Alanine au cadre de lecture Glycine-Proline à partir de l'ARN sens ; suggérant que la protéine polyGP était, elle aussi, traduite à partir du codon CTG (Tabet et al., 2018). De même, l'utilisation d'un ASO dirigé contre l'ARN sens en neurones différenciés à partir de cellules iPS de patients C9-SLA/DFT et en modèle murin transgénique exprimant les répétitions de nucléotides GGGGCC conduit à une diminution de l'expression de la protéine polyGP, suggérant donc que cette protéine est codée à partir de l'ARN sens (Gendron et al., 2017). Néanmoins, la délétion de ce codon CTG dans des cellules

iPS de patients C9-SLA/DFT ne conduit pas à une diminution d'expression de la protéine polyGP, suggérant donc que cette protéine n'est pas traduite à partir de ce codon CTG par changement du cadre de lecture (Almeida et al., 2019). De plus, le développement d'anticorps dirigés contre les régions C-terminales putatives des protéines polyGP traduites soit à partir de l'ARN sens, soit à partir de l'ARN antisens a permis de déterminer que cette protéine polyGP est majoritairement traduite à partir de l'ARN antisens (Zu et al., 2013), ce qui est en accord avec notre travail. En effet, nous avons montré que la protéine polyGP était traduite à partir de l'ARN antisens grâce à un codon ATG situé en amont des répétitions CCCCGG. Enfin, nous supposons que l'initiation de la traduction de la protéine DPR polyGR a lieu à un codon AGG situé, lui aussi, en amont des répétitions.

Des études ont reporté un mécanisme de traduction « RAN » des expansions de répétitions GGGGCC du gène *C9ORF72*. En effet, il a été montré que les répétitions sens GGGGCC et antisens CCCCGG étaient traduites en protéines composées de polyGA, polyGP, polyGR, polyPA et polyPR en culture cellulaire (Ash et al., 2013 ; Gendron et al., 2013 ; Mori et al., 2013a ; Zu et al., 2013). Néanmoins, une de ces études a montré que l'ARN sens était principalement traduit en protéine polyGA et suggérait une initiation de la traduction en amont des répétitions (Mori et al., 2013a). Le développement d'anticorps contre ces 5 protéines et l'analyse de tissus de patients C9-SLA/DFT a permis de confirmer leur présence dans le système nerveux central chez les patients, notamment dans l'hippocampe et rarement dans les cortex moteurs et temporaux ainsi que dans la moelle épinière (Gendron et al., 2013 ; Zu et al., 2013). Par la suite, des études histopathologiques complémentaires sur des patients C9-SLA/DFT ont permis d'affiner ces premiers résultats, notamment sur l'abondance relative de ces protéines DPR ainsi que leurs zones d'expression. En effet, les protéines polyGA et polyGP sont les 2 protéines DPR les plus fréquemment observées, principalement dans les cortex frontaux, temporaux et dans le cervelet et très rarement dans la moelle épinière (Mackenzie et al., 2015). Il est à noter que les 2 protéines DPR les plus fortement exprimées chez les patients sont celles où nous avons déterminé une initiation canonique de la traduction. L'ensemble de ces résultats suggèrent donc que 2 protéines DPR peuvent être traduites via un mécanisme de traduction classique à partir d'un codon ATG, ou ressemblant à un ATG, situé en amont des expansions de répétitions.

4.2. Mécanisme de traduction des protéines à polyG

Concernant les maladies à polyG, il est maintenant admis que, dans FXTAS, la traduction des expansions de répétitions CGG se fait principalement dans le cadre de lecture glycine, générant une protéine appelée FMRpolyG. L'initiation de la traduction se fait en amont des répétitions à un codon ACG (Kearse et al., 2017 ; Sellier et al., 2017) et ceci de manière dépendante du mécanisme de scan du ribosome (Kearse et al., 2017). D'autres protéines traduites à partir du mécanisme « RAN » et donc issues soit d'un autre cadre de lecture ouvert (polyA) soit à partir de l'ARN antisens (polyP, polyR et polyA) sont également décrites *in vitro* mais ne sont pas, ou très rarement observées chez les patients FXTAS (Todd et al., 2013 ; Kranz et al., 2016 ; Krans et al., 2019), soulevant la question de la relevance physiopathologique de ces protéines. De plus, l'équipe du Dr Charlet, par le développement d'un modèle murin exprimant 100 répétitions CCG mais dont la région en amont des répétitions est délétée du codon ACG (qui permet la traduction de la protéine FMRpolyG), a démontré que la présence de la protéine FMRpolyG est nécessaire au développement de la pathologie (Sellier et al., 2017). En effet, la délétion de cette région contenant l'ACG ne modifie pas le mécanisme de traduction « RAN », suggérant que les protéines issues de ce mécanisme ne sont pas suffisantes pour engendrer la maladie, contrairement à FMRpolyG.

Dans la cadre de l'étude de NIID, nous avons montré que, tout comme dans FXTAS, les répétitions CGG étaient traduites dans le cadre de lecture glycine, produisant une nouvelle protéine que nous avons nommé uN2CpolyG. L'initiation de la traduction de cette protéine uN2CpolyG a lieu en amont des répétitions de nucléotides CGG à un codon canonique de l'initiation de la traduction ATG. Ce résultat a été récemment confirmé par une seconde équipe (Zhong et al., 2021). Cette protéine forme des inclusions intranucléaires en cultures primaires de neurones de souris ainsi qu'en modèle murin, résultant à une neurodégénérescence. Ces données suggèrent que l'expression de uN2CpolyG est suffisante pour récapituler les caractéristiques histopathologiques de NIID et induire cette pathologie. Lors de cette étude, nous n'avons pas observé *in vitro* une traduction des répétitions CGG dans un autre cadre de lecture (alanine ou arginine), suggérant que le mécanisme de traduction « RAN » n'est pas prédominant dans cette pathologie.

Dans l'OPDM2, les expansions de répétitions CGG situées dans la région 5'UTR du gène *GIPC1* sont elles aussi traduites en une protéine composée de polyGlycine, appelé uGIPC1polyG. Nous avons montré que, là encore, l'initiation de la traduction a lieu en amont

des répétitions et, tout comme dans FXTAS, à un codon ressemblant à un ATG et, dans le cas de l'OPDM2, à un codon CTG. Il est à noter que les codons CTG permettent l'initiation de la traduction de plus de 40% des uORF ; ce résultat n'est donc pas inattendu. Les symptômes et les caractéristiques histopathologiques de l'OPDM ressemblent à ceux observés dans l'OPMD, une maladie neuromusculaire due à une expansion de 2 à 8 répétitions GCG dans le gène *PABPN1* qui est traduite en une protéine contenant une séquence polyA (cf Introduction 1.4.3). En effet, ces 2 pathologies sont caractérisées une faiblesse musculaire associée à un ptosis. De plus, les patients atteints d'OPDM ou d'OPMD présentent des vacuoles ainsi que des inclusions intranucléaires positives à l'ubiquitine dans leurs fibres musculaires. Afin de s'assurer que la protéine uGIPC1polyG est bien responsable des symptômes observés chez les patients OPDM, et non la protéine uGIPC1polyA qui pourrait être générée par le mécanisme de traduction « RAN », nous avons également développé des anticorps dirigés contre la protéine putative uGIPC1polyA afin de tester sa présence / son absence chez les patients OPDM.

L'analyse de la séquence nucléotidique du gène *LRP12*, dont une expansion de répétitions CGG est responsable de l'OPMD1, révèle la présence de codons CTG et ACG en amont des répétitions pouvant peut-être permettre une initiation de la traduction dans le cadre de lecture glycine (Figure 65A). De plus, un codon ATG se trouve dans le cadre de lecture arginine et, ce cadre de lecture ouvert permet également la traduction de la protéine LRP12, suggérant qu'une protéine LRP12 avec une séquence N-terminale plus longue et riche en arginines puisse exister (Figure 65A). Nous avons récemment pu montrer que ces protéines (à polyG et à polyR) étaient traduites à partir de l'uORF du gène *LRP12* (Figure 65B). Il serait maintenant intéressant d'étudier le ratio d'initiation *in vivo* entre ces codons CTG et ATG afin de conclure sur une expression d'une protéine composée de polyG et / ou d'une protéine à polyR. L'OPDM1 étant caractérisée, tout comme FXTAS, NIID et l'OPDM2, par la présence d'agrégats intranucléaires, nous pouvons penser qu'une protéine à polyG traduite à partir des répétitions CGG puisse être responsable de la présence de ces inclusions. Les protéines à polyR sont connues pour être extrêmement toxiques mais, elles sont également particulièrement instables. Le développement d'anticorps dirigés contre ces protéines à polyG et à polyR, ou contre la protéine LRP12, se révèlera donc nécessaire afin de tester la présence de ces protéines chez les patients OPDM1.

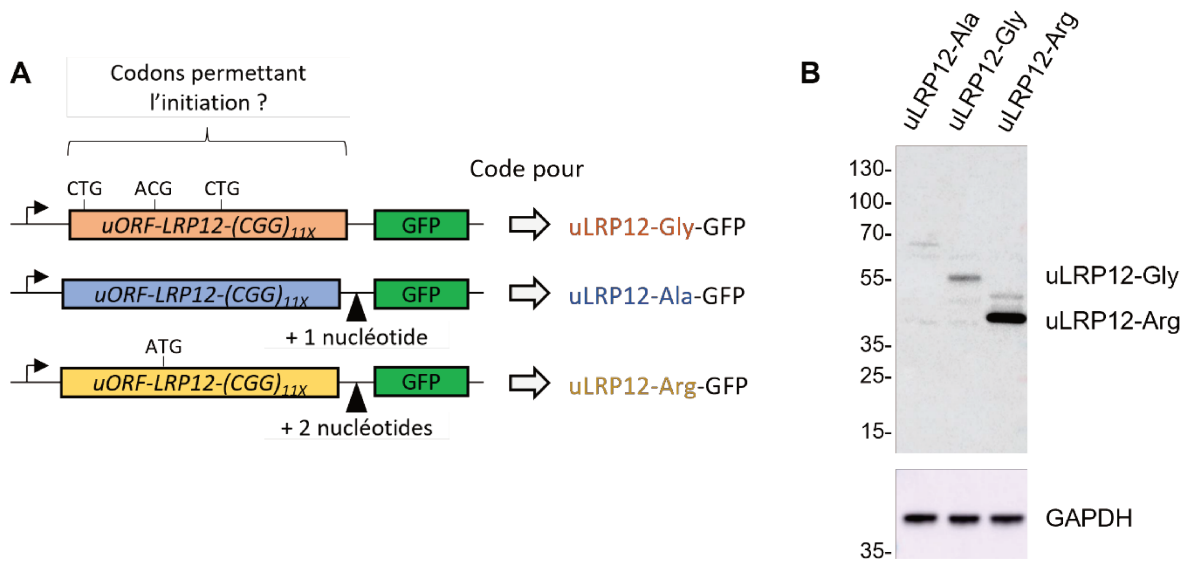


Figure 65. Traduction des répétitions CGG dans le gène LRP12
(**A**) Schéma représentant les 3 constructions plasmidiques dans les 3 cadres de lecture possible (Ala, Gly, Arg) avec les codons situés en amont des répétitions CGG pouvant, peut-être, permettre une initiation de la traduction de l'uORF du gène *LRP12* fusionnée à l'ADNc de la GFP. (**B**) Immunoblot dirigé contre la GFP et la GAPDH après transfection des différentes constructions dans des cellules HEK293.

Pour finir, l'OPML est causée par une expansion de répétitions CGG dans le locus *642361*. L'analyse de la séquence en nucléotides révèle la présence d'un ATG en amont des répétitions CGG transcrites en un long ARN décrit comme non-codant actuellement, dans le cadre de lecture glycine. Comme pour NIID, il serait intéressant de tester la présence de cette possible protéine à polyG chez les patients grâce à des anticorps spécifiques. Néanmoins, le faible nombre de personnes affecté par cette pathologie complexifie l'accès aux tissus et donc une éventuelle étude histopathologique.

4.3. La traduction de régions dites « non-codantes »

Dans l'OPML, les répétitions CGG pourraient être traduites en une protéine composée de polyGlycine à partir d'un ARNlnc. La traduction d'une protéine toxique à partir d'un ARNlnc n'est pas sans rappeler 2 autres maladies, SCA8 et maladie de Huntington 2 (HDL2). En effet, SCA8 est due à une expansion de 80 à 160 répétitions CAG dans le gène *ATXN8*, décrit comme étant transcrit en un long ARN non-codant (Ikeda et al., 2000). Le développement de modèles murins de SCA8 ainsi que l'utilisation d'anticorps dirigés contre les polyQ a permis de mettre en évidence un codon ATG situé en amont des répétitions CAG, permettant la traduction d'une protéine ATXN8-polyQ responsable de la pathologie (Moseley et al., 2006). De même, une expansion comprise entre 40 et 59 répétitions CAG dans le l'ARNlnc JPH3AS a été identifié chez les patients HDL2 (Holmes et al., 2001) et le développement d'un modèle murin de la pathologie révèle la présence la présence d'une protéine à polyQ codée à partir de cet ARN grâce à un codon AUG situé en amont des répétitions qui était, jusque-là, décrit comme non-codant (Wilburn et al., 2011).

La traduction d'expansions de répétitions localisées dans des régions dites non-codantes pourrait se révéler plus courante qu'initialement pensée dans les maladies à expansion de répétitions. En effet, dans SCA12, une expansion d'une centaine de répétitions CAG a été décrite dans le gène *PPP2R2B* (Holmes et al., 1999). Les patients SCA12 ne présentent pas d'agrégats de protéines composées de polyGlutamine, suggérant que SCA12 n'est pas une maladie à polyQ (O'Hearn et al., 2015). Bien que la cause de toxicité de cette expansion soit encore discutée, la présence d'agrégats intranucléaires dans les cellules de Purkinje ainsi que la présence d'un codon ATG dans le cadre de lecture sérine pourrait laisser penser à une accumulation toxique d'une protéine composée de polyS.

Récemment, il a été démontré que des protéines composées de polyGP étaient retrouvées dans une autre pathologie, l'ataxie spinocérébelleuse de type 36 (SCA36). Cette maladie est causée par une expansion de répétitions GGGCCT située dans le premier intron du gène *NOP56* (Kobayashi et al., 2011 ; García-Murias et al., 2012). Il est à noter que la séquence nucléotidique de cette expansion, ainsi que sa localisation, rappellent l'expansion GGGGCC intronique du gène *C9ORF72*. Cette expansion GGGCCT peut putativement encoder 5 protéines DPR : polyWA, polyGL, polyPR, polyAQ et polyGP. Le développement d'un modèle murin exprimant l'expansion de répétitions GGGCCT présente des foci d'ARN sens et antisens et les protéines DPR polyWA, polyGL, polyPR et polyGP sont détectées (Todd et al., 2020) ; ce qui

rappelle fortement ce qui est observé dans les modèles murins de SLA exprimant des répétitions GGGGCC (O'Rourke et al., 2015 ; Peters et al., 2015 ; Jiang et al., 2016 ; Liu et al., 2016b). Néanmoins, le développement d'anticorps dirigés contre ces protéines DPR et leur utilisation dans des fibroblastes de patients ainsi que sur des lames de cerveau de patients SCA36 révèle une présence majoritaire de la protéine polyGP sous forme soluble (McEachin et al., 2020). De manière intéressante, l'expansion de répétitions GGGCCT altère l'épissage de l'ARN de NOP56, conduisant à la rétention de l'intron porteur des répétitions, qui se retrouve alors dans la phase codante. En effet, le codon d'initiation de la traduction ATG de la protéine NOP56 permet la traduction des répétitions GGGCCT dans le cadre de lecture Glycine-Proline (McEachin et al., 2020). La présence de la protéine polyGP dans 2 pathologies, C9-SLA/DFT et SCA36, peut suggérer un mécanisme de toxicité similaire de cette protéine.

L'ensemble de ces études mettent en évidence que la majeure partie de ces expansions de répétitions sont traduites par un mécanisme canonique de l'initiation de la traduction. Il est important de noter que l'initiation de la traduction à ces codons ressemblant à des ATG, ACG et CTG, ou ATG est indépendante de la taille des répétitions (Kearse et al., 2017 ; Sellier et al., 2017 ; Zhong et al., 2021). Cependant, avec un faible nombre de répétitions, ces protéines sont petites, instables et donc difficiles à détecter. Ces données suggèrent que de nombreuses expansions de répétitions, bien que situées dans des régions dites non-codantes du génome, pouvant être traduites grâce à des codons ATG ou ressemblant à des ATG, et ceci d'autant plus que la majorité des gènes contiennent des uORF initiant à des codons ressemblant à des ATG (Ingolia et al., 2011).

Nous n'avons pas observé de traduction « RAN » lors de nos études sur ces différentes pathologies à expansion de répétitions. Ce mécanisme de traduction reste, à l'heure actuelle, mal décrit ; il s'agirait d'une initiation de la traduction se produisant dans les répétitions grâce à des structures secondaires d'ARN formées par les répétitions (Zu et al., 2011).

Il est à noter que dans de nombreuses études portant sur la traduction « RAN », les codons permettant l'initiation de la protéine majoritairement retrouvée chez les patients sont retirés de la séquence naturelle. Par exemple, dans l'étude originelle démontrant la traduction « RAN », le codon ATG permettant la traduction de la protéine ATXN8-polyQ a été supprimé, ce qui a permis la détection de protéines à polyS et à polyA, suggérant que la traduction a lieu dans tous les cadres de lecture possibles (Zu et al., 2011). De manière similaire, dans HD, toutes les protéines possiblement traduites à partir des expansions de répétitions sens CAG, mais

également antisens CTG, à savoir polyQ, polyA, polyS, polyL et polyC, sont détectées (Bañez-Coronel et al., 2015) mais, ces travaux sont réalisés à partir de constructions plasmidiques artificielles où les expansions de répétitions sont dénuées de leur séquence naturelle environnante. Nous pouvons donc nous demander si les vecteurs utilisés sont bien dépourvus de tous codon ATG et / ou codons ressemblant à des ATG comme les codons CTG, GTG, ACG, etc. permettant l'initiation de la traduction. A ce sujet, il est intéressant de noter que certaines des expansions de répétitions traduites par le mécanisme « RAN » sont composées de codons CTG : CCTG dans le gène *CNBP* dans la DM2 (Zu et al., 2017) et ; CTG dans la séquence antisens de la *HTT* dans HD (Bañez-Coronel et al., 2015). Or, il est connu que les codons CTG sont capables d'initier la traduction (Ingolia et al., 2011) ; comme observé pour la traduction de la protéine polyGA dans C9-SLA/DFT ou uGIPC1polyG dans l'OPDM2.

De plus, il est intéressant de noter que seules les protéines à polyA, qui sont détectées dans SCA8, FXTAS et HD, migrent sous forme de traînée, suggérant que l'initiation de la traduction n'a pas lieu à un codon unique (Zu et al., 2011 ; Todd et al., 2013 ; Bañez-Coronel et al., 2015). Néanmoins, nous avons montré que l'expression d'une protéine à polyA traduite grâce à un codon ATG situé en amont des répétitions CGG migre sous la forme d'une traînée, suggérant que les protéines à polyA ont un profil de migration anormal (Figure 66). Enfin, les autres protéines issues de la traduction « RAN » détectées dans différentes pathologies causées par des expansions de répétitions comme les protéines DPR issues des expansions GGGGCC du gène *C9ORF72* dans la SLA/DFT (Zu et al., 2013), ou encore pour les protéines polyLPAC et polyQAGR identifiées dans la DM2 (Zu et al., 2017) migrent à des tailles uniques. Ces résultats suggèrent donc que l'initiation de la traduction a lieu à un codon défini.

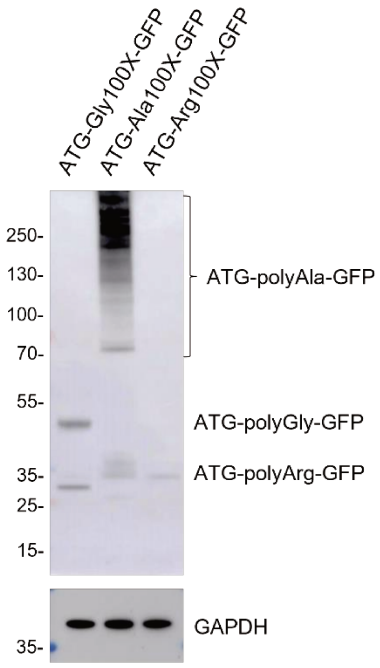


Figure 66. Profil de migration des protéines à polyA
Différentes constructions plasmides codant des protéines à polyGly, polyAla ou polyArg à partir d'un ATG situé en amont d'une centaine d'expansion CGG fusionnées à la GFP ont été transfectées dans des cellules HEK293 puis, un immunoblot contre la GFP et la GAPDH a été réalisé.

Enfin, la présence de ces différentes protéines « RAN » a été confirmée chez les patients grâce au développement d'anticorps. Toutefois, ceux-ci sont dirigés contre les séquences répétées de ces protéines, ces séquences pouvant être retrouvées dans d'autres protéines. De plus, ces protéines « RAN » sont majoritairement observées dans les zones non-atteintes chez les patients. En effet, dans la DM1, la protéine à polyQ est majoritairement retrouvée dans le sang (Zu et al., 2011) ; dans la DM2, la présence de la protéine polyLPAC est inversement corrélée à la présence de foci d'ARN CCUG, suggérant que cette protéine n'est pas dans les zones atteintes (Zu et al., 2017) et ; dans HD, la présence des protéines « RAN » est inversement corrélée à la protéine HTT-polyQ (Bañez-Coronel et al., 2015). Ces résultats soulèvent donc des questions sur la relevance clinique de ces protéines.

DISCUSSION

- CHAPITRE 5 -

Conclusion

5. CONCLUSION

Ce travail de thèse a permis de clarifier le mécanisme de toxicité des expansions CGG situées dans les régions 5'UTR des gènes *NOTCH2NLC* et *GIPC1*, responsables de NIID et de l'OPDM2 respectivement. Nous avons montré que ces répétitions étaient traduites en protéines composées de polyGlycine, responsables de la formation des inclusions intranucléaires caractéristiques de ces maladies et toxiques en modèles cellulaires et / ou murins. Ces résultats permettent ainsi de définir une nouvelle classe de maladie génétique humaine, les maladies à polyG, comportant actuellement 3 pathologies : FXTAS, NIID et OPDM2/3.

En parallèle, nous avons pu montrer que des expansions de répétitions GGGGCC situées dans la région intronique du gène *C9ORF72*, responsables de la SLA/DFT, sont, elles aussi traduites en protéines DPR de type polyGA à partir de l'ARN sens composé de répétitions GGGGCC et de type polyGP à partir de l'ARN antisens composé de répétitions CCCCGG. Ces protéines DPR sont faiblement exprimées et instables car elles sont constamment dégradées par autophagie. Or, l'expansion de répétitions GGGGCC conduit à haploinsuffisance de la protéine C9ORF72, qui est impliquée dans l'autophagie. La diminution d'expression de C9ORF72 conduit à une autophagie sous-optimale et, donc, à une accumulation et à une plus forte toxicité de ces protéines polyGA et polyGP en modèle cellulaire. L'expansion de répétitions GGGGCC serait donc toxique via 2 mécanismes : une sous-expression de la protéine C9ORF72 augmentant la toxicité des protéines DPR, pouvant peut-être expliquer la mort neuronale observée chez les patients atteints de SLA/DFT.

Enfin, lors de l'étude de la traduction de ces différentes protéines (uN2CpolyG, uGIPC1polyG, polyGA et polyGP), nous n'avons pas observé de traduction par le mécanisme « RAN ». En effet, dans nos études, l'initiation de la traduction de ces protéines a lieu systématiquement en amont des répétitions, soit à un codon ATG pour les protéines uN2CpolyG et polyGP, soit à un codon CTG qui permet l'initiation de la traduction par mésappariement avec l'ARNt initiateur, pour les protéines uGIPC1 et polyGA. En conclusion, ces données suggèrent que, bien que situées dans des régions dites « non-codantes », certaines expansions de répétitions peuvent tout de même être traduites en protéines toxiques, et ce, grâce à un mécanisme canonique de l'initiation de la traduction.

REFERENCES BIBLIOGRAPHIQUES

Références bibliographiques

A

Aartsma-Rus A. FDA Approval of Nusinersen for Spinal Muscular Atrophy Makes 2016 the Year of Splice Modulating Oligonucleotides. Nucleic Acid Ther. 2017 Apr;27(2):67-69. doi: 10.1089/nat.2017.0665. Epub 2017 Feb 21. PMID: 28346110.

Almeida S, Gascon E, Tran H, Chou HJ, Gendron TF, Degroot S, Tapper AR, Sellier C, Charlet-Berguerand N, Karydas A, Seeley WW, Boxer AL, Petrucelli L, Miller BL, Gao FB. Modeling key pathological features of frontotemporal dementia with C9ORF72 repeat expansion in iPSC-derived human neurons. Acta Neuropathol. 2013 Sep;126(3):385-99. doi: 10.1007/s00401-013-1149-y. Epub 2013 Jul 9. Erratum in: Acta Neuropathol. 2014 Jun;127(6):941. PMID: 23836290; PMCID: PMC3753484.

Almeida S, Krishnan G, Rushe M, Gu Y, Kankel MW, Gao FB. Production of poly(GA) in C9ORF72 patient motor neurons derived from induced pluripotent stem cells. Acta Neuropathol. 2019 Dec;138(6):1099-1101. doi: 10.1007/s00401-019-02083-z. Epub 2019 Oct 17. PMID: 31624870; PMCID: PMC6851345.

Amick J, Roczniak-Ferguson A, Ferguson SM. C9orf72 binds SMCR8, localizes to lysosomes, and regulates mTORC1 signaling. Mol Biol Cell. 2016 Oct 15;27(20):3040-3051. doi: 10.1091/mbc.E16-01-0003. Epub 2016 Aug 24. PMID: 27559131; PMCID: PMC5063613.

Amick J, Tharkeshwar AK, Amaya C, Ferguson SM. WDR41 supports lysosomal response to changes in amino acid availability. Mol Biol Cell. 2018 Sep 1;29(18):2213-2227. doi: 10.1091/mbc.E17-12-0703. Epub 2018 Jul 11. PMID: 29995611; PMCID: PMC6249801.

Amick J, Tharkeshwar AK, Talaia G, Ferguson SM. PQLC2 recruits the C9orf72 complex to lysosomes in response to cationic amino acid starvation. J Cell Biol. 2020 Jan 6;219(1):e201906076. doi: 10.1083/jcb.201906076. PMID: 31851326; PMCID: PMC7039192.

Anderson DM, Anderson KM, Chang CL, Makarewich CA, Nelson BR, McAnally JR, Kasaragod P, Shelton JM, Liou J, Bassel-Duby R, Olson EN. A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. Cell. 2015 Feb 12;160(4):595-606. doi: 10.1016/j.cell.2015.01.009. Epub 2015 Jan 29. PMID: 25640239; PMCID: PMC4356254.

Andrade NS, Ramic M, Esanov R, Liu W, Rybin MJ, Gaidosh G, Abdallah A, Del'Olio S, Huff TC, Chee NT, Anatha S, Gendron TF, Wahlestedt C, Zhang Y, Benatar M, Mueller C, Zeier Z. Dipeptide repeat proteins inhibit homology-directed DNA double strand break repair in C9ORF72 ALS/FTD. Mol Neurodegener. 2020 Feb 24;15(1):13. doi: 10.1186/s13024-020-00365-9. PMID: 32093728; PMCID: PMC7041170.

Arai T, Hasegawa M, Akiyama H, Ikeda K, Nonaka T, Mori H, Mann D, Tsuchiya K, Yoshida M, Hashizume Y, Oda T. TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Biochem Biophys Res Commun. 2006 Dec 22;351(3):602-11. doi: 10.1016/j.bbrc.2006.10.093. Epub 2006 Oct 30. PMID: 17084815.

Arnaud E, Touriol C, Boutonnet C, Gensac MC, Vagner S, Prats H, Prats AC. A new 34-kilodalton isoform of human fibroblast growth factor 2 is cap dependently synthesized by using a non-AUG start codon and behaves as a survival factor. Mol Cell Biol. 1999 Jan;19(1):505-14. doi: 10.1128/MCB.19.1.505. PMID: 9858574; PMCID: PMC83908.

Arnoux I, Willam M, Griesche N, Krummeich J, Watari H, Offermann N, Weber S, Narayan Dey P, Chen C, Monteiro O, Buettner S, Meyer K, Bano D, Radyushkin K, Langston R, Lambert JJ, Wanker E, Methner A, Krauss S, Schweiger S, Stroh A. Metformin reverses early cortical network dysfunction and behavior changes in Huntington's disease. Elife. 2018 Sep 4;7:e38744. doi: 10.7554/eLife.38744. PMID: 30179155; PMCID: PMC6156080.

Arrasate M, Mitra S, Schweitzer ES, Segal MR, Finkbeiner S. Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. Nature. 2004 Oct 14;431(7010):805-10. doi: 10.1038/nature02998. PMID: 15483602.

Artavanis-Tsakonas S, Rand MD, Lake RJ. Notch signaling: cell fate control and signal integration in development. Science. 1999 Apr 30;284(5415):770-6. doi: 10.1126/science.284.5415.770. PMID: 10221902.

Ash PE, Bieniek KF, Gendron TF, Caulfield T, Lin WL, Dejesus-Hernandez M, van Blitterswijk MM, Jansen-West K, Paul JW 3rd, Rademakers R, Boylan KB, Dickson DW, Petrucelli L. Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. Neuron. 2013 Feb 20;77(4):639-46. doi: 10.1016/j.neuron.2013.02.004. Epub 2013 Feb 12. PMID: 23415312; PMCID: PMC3593233.

Atanasio A, Decman V, White D, Ramos M, Ikiz B, Lee HC, Siao CJ, Brydges S, LaRosa E, Bai Y, Fury W, Burfeind P, Zamfirova R, Warshaw G, Orengo J, Oyejide A, Fralish M, Auerbach W, Poueymirou W, Freudenberg J, Gong G, Zambrowicz B, Valenzuela D, Yancopoulos G, Murphy A, Thurston G, Lai KM. C9orf72 ablation causes immune dysregulation characterized by leukocyte expansion, autoantibody production, and glomerulonephropathy in mice. Sci Rep. 2016 Mar 16;6:23204. doi: 10.1038/srep23204. PMID: 26979938; PMCID: PMC4793236.

B

Baborie A, Griffiths TD, Jaros E, Perry R, McKeith IG, Burn DJ, Masuda-Suzukake M, Hasegawa M, Rollinson S, Pickering-Brown S, Robinson AC, Davidson YS, Mann DM. Accumulation of dipeptide repeat proteins predates that of TDP-43 in frontotemporal lobar degeneration associated with hexanucleotide repeat expansions in C9ORF72 gene. Neuropathol Appl Neurobiol. 2015 Aug;41(5):601-12. Doi: 10.1111/nan.12178. Epub 2015 Apr 30. PMID: 25185840; PMCID: PMC4934135.

Bajc Česnik A, Darovic S, Prpar Mihevc S, Štalekar M, Malnar M, Motaln H, Lee YB, Mazej J, Pohleven J, Grosch M, Modic M, Fonovič M, Turk B, Drukker M, Shaw CE, Rogelj B. Nuclear RNA foci from *C9ORF72* expansion mutation form paraspeckle-like bodies. J Cell Sci. 2019 Mar 7;132(5):jcs224303. Doi: 10.1242/jcs.224303. PMID: 30745340.

Baker M, Mackenzie IR, Pickering-Brown SM, Gass J, Rademakers R, Lindholm C, Snowden J, Adamson J, Sadovnick AD, Rollinson S, Cannon A, Dwosh E, Neary D, Melquist S, Richardson A, Dickson D, Berger Z, Eriksen J, Robinson T, Zehr C, Dickey CA, Crook R, McGowan E, Mann D, Boeve B, Feldman H, Hutton M. Mutations in progranulin cause tau-negative frontotemporal dementia linked to chromosome 17. Nature. 2006 Aug 24;442(7105):916-9. Doi: 10.1038/nature05016. Epub 2006 Jul 16. PMID: 16862116.

Balendra R, Isaacs AM. C9orf72-mediated ALS and FTD: multiple pathways to disease. Nat Rev Neurol. 2018 Sep;14(9):544-558. Doi: 10.1038/s41582-018-0047-2. PMID: 30120348; PMCID: PMC6417666.

Bañez-Coronel M, Ayhan F, Tarabochia AD, Zu T, Perez BA, Tusi SK, Pletnikova O, Borchelt DR, Ross CA, Margolis RL, Yachnis AT, Troncoso JC, Ranum LP. RAN Translation in Huntington Disease. Neuron. 2015 Nov 18;88(4):667-77. Doi: 10.1016/j.neuron.2015.10.038. PMID: 26590344; PMCID: PMC4684947.

Banno H, Katsuno M, Suzuki K, Tanaka F, Sobue G. Neuropathology and therapeutic intervention in spinal and bulbar muscular atrophy. Int J Mol Sci. 2009 Mar;10(3):1000-12. Doi: 10.3390/ijms10031000. Epub 2009 Mar 10. PMID: 19399234; PMCID: PMC2672015.

Baratela WA, Bober MB, Tiller GE, Okenfuss E, Ditro C, Duker A, Krakow D, Stabley DL, Sol-Church K, Mackenzie W, Lachman R, Scott CI Jr. A newly recognized syndrome with characteristic facial features, skeletal dysplasia, and developmental delay. Am J Med Genet A. 2012 Aug;158A(8):1815-22. Doi: 10.1002/ajmg.a.35445. Epub 2012 Jun 18. PMID: 22711505; PMCID: PMC4164294.

Barker HV, Niblock M, Lee YB, Shaw CE, Gallo JM. RNA Misprocessing in *C9orf72*-Linked Neurodegeneration. Front Cell Neurosci. 2017 Jul 11;11:195. Doi: 10.3389/fncel.2017.00195. PMID: 28744202; PMCID: PMC5504096.

Barmada SJ, Serio A, Arjun A, Bilican B, Daub A, Ando DM, Tsvetkov A, Pleiss M, Li X, Peisach D, Shaw C, Chandran S, Finkbeiner S. Autophagy induction enhances TDP43 turnover and survival in neuronal ALS models. Nat Chem Biol. 2014 Aug;10(8):677-85. Doi: 10.1038/nchembio.1563. Epub 2014 Jun 29. PMID: 24974230; PMCID: PMC4106236.

Barnat M, Capizzi M, Aparicio E, Boluda S, Wennagel D, Kacher R, Kassem R, Lenoir S, Agasse F, Braz BY, Liu JP, Ighil J, Tessier A, Zeitlin SO, Duyckaerts C, Dommergues M, Durr A, Humbert S. Huntington's disease alters human neurodevelopment. Science. 2020 Aug 14;369(6505):787-793. Doi: 10.1126/science.aax3338. Epub 2020 Jul 16. PMID: 32675289; PMCID: PMC7859879.

Barth-Baus D, Bhasker CR, Zoll W, Merrick WC. Influence of translation factor activities on start site selection in six different mRNAs. Translation (Austin). 2013 Apr 1;1(1):e24419. Doi: 10.4161/trla.24419. PMID: 26824019; PMCID: PMC4718060.

Barrett AJ, Kirschke H. Cathepsin B, Cathepsin H, and cathepsin L. Methods Enzymol. 1981;80 Pt C:535-61. Doi: 10.1016/s0076-6879(81)80043-2. PMID: 7043200.

Battle MA, Maher VM, McCormick JJ. ST7 is a novel low-density lipoprotein receptor-related protein (LRP) with a cytoplasmic tail that interacts with proteins related to signal transduction pathways. Biochemistry. 2003 Jun 24;42(24):7270-82. Doi: 10.1021/bi034081y. PMID: 12809483.

Beck J, Poulter M, Hensman D, Rohrer JD, Mahoney CJ, Adamson G, Campbell T, Uphill J, Borg A, Fratta P, Orrell RW, Malaspina A, Rowe J, Brown J, Hodges J, Sidle K, Polke JM, Houlden H, Schott JM, Fox NC, Rossor MN, Tabrizi SJ, Isaacs AM, Hardy J, Warren JD, Collinge J, Mead S. Large C9orf72 hexanucleotide repeat expansions are seen in multiple neurodegenerative syndromes and are more frequent than expected in the UK population. Am J Hum Genet. 2013 Mar 7;92(3):345-53. Doi: 10.1016/j.ajhg.2013.01.011. Epub 2013 Feb 21. PMID: 23434116; PMCID: PMC3591848.

Behrends C, Sowa ME, Gygi SP, Harper JW. Network organization of the human autophagy system. Nature. 2010 Jul 1;466(7302):68-76. Doi: 10.1038/nature09204. Epub 2010 Jun 20. PMID: 20562859; PMCID: PMC2901998.

Belzil VV, Bauer PO, Prudencio M, Gendron TF, Stetler CT, Yan IK, Pregent L, Daughrity L, Baker MC, Rademakers R, Boylan K, Patel TC, Dickson DW, Petrucelli L. Reduced C9orf72 gene expression in c9FTD/ALS is caused by histone trimethylation, an epigenetic event detectable in blood. Acta Neuropathol. 2013 Dec;126(6):895-905. Doi: 10.1007/s00401-013-1199-1. Epub 2013 Oct 29. PMID: 24166615; PMCID: PMC3830740.

Benajiba L, Le Ber I, Camuzat A, Lacoste M, Thomas-Anterion C, Couratier P, Legallic S, Salachas F, Hannequin D, Decousus M, Lacomblez L, Guedj E, Golfier V, Camu W, Dubois B, Campion D, Meininger V, Brice A; French Clinical and Genetic Research Network on Frontotemporal Lobar Degeneration/Frontotemporal Lobar Degeneration with Motoneuron Disease. TARDBP mutations in motoneuron disease with frontotemporal lobar degeneration. Ann Neurol. 2009 Apr;65(4):470-3. Doi: 10.1002/ana.21612. PMID: 19350673.

Bennett CF, Krainer AR, Cleveland DW. Antisense Oligonucleotide Therapies for Neurodegenerative Diseases. Annu Rev Neurosci. 2019 Jul 8;42:385-406. Doi: 10.1146/annurev-neuro-070918-050501. PMID: 31283897; PMCID: PMC7427431.

Benussi L, Rossi G, Glionna M, Tonoli E, Piccoli E, Fostinelli S, Paterlini A, Flocco R, Albani D, Pantieri R, Cereda C, Forloni G, Tagliavini F, Binetti G, Ghidoni R. C9ORF72 hexanucleotide repeat number in frontotemporal lobar degeneration: a genotype-phenotype correlation study. J Alzheimers Dis. 2014;38(4):799-808. Doi: 10.3233/JAD-131028. Erratum in: J Alzheimers Dis. 2015;45(1):319. PMID: 24064469.

Bizzotto S, Walsh CA. Making a Notch in the Evolution of the Human Cortex. Dev Cell. 2018 Jun 4;45(5):548-550. Doi: 10.1016/j.devcel.2018.05.015. PMID: 29870717.

Blasco H, Corcia P, Veyrat-Durebex C, Coutadeur C, Fournier C, Camu W, Gordon P, Praline J, Andres CR, Vourc'h P; French ALS Study Group. The P413L chromogranin B variation in French patients with sporadic amyotrophic lateral sclerosis. Amyotroph Lateral Scler. 2011 May;12(3):210-4. Doi: 10.3109/17482968.2010.522587. Epub 2010 Oct 11. PMID: 20932227.

Blokhuis AM, Groen EJ, Koppers M, van den Berg LH, Pasterkamp RJ. Protein aggregation in amyotrophic lateral sclerosis. Acta Neuropathol. 2013 Jun;125(6):777-94. Doi: 10.1007/s00401-013-1125-6. Epub 2013 May 15. PMID: 23673820; PMCID: PMC3661910.

Boeynaems S, Bogaert E, Michiels E, Gijselinck I, Sieben A, Jovičić A, De Baets G, Scheveneels W, Steyaert J, Cuijt I, Verstrepen KJ, Callaerts P, Rousseau F, Schymkowitz J, Cruts M, Van Broeckhoven C, Van Damme P, Gitler AD, Robberecht W, Van Den Bosch L. Drosophila screen connects nuclear transport genes to DPR pathology in c9ALS/FTD. Sci Rep. 2016 Feb 12;6:20877. Doi: 10.1038/srep20877. PMID: 26869068; PMCID: PMC4751451.

Boeynaems S, Bogaert E, Kovacs D, Konijnenberg A, Timmerman E, Volkov A, Guharoy M, De Decker M, Jaspers T, Ryan VH, Janke AM, Baatsen P, Vercruysse T, Kolaitis RM, Daelemans D, Taylor JP, Kedersha N, Anderson P, Impens F, Sobott F, Schymkowitz J, Rousseau F, Fawzi NL, Robberecht W, Van Damme P, Tompa P, Van Den Bosch L. Phase Separation of C9orf72 Dipeptide Repeats Perturbs Stress Granule Dynamics. Mol Cell. 2017 Mar 16;65(6):1044-1055.e5. doi: 10.1016/j.molcel.2017.02.013. PMID: 28306503; PMCID: PMC5364369.

Božič J, Motaln H, Janež AP, Markič L, Tripathi P, Yamoah A, Aronica E, Lee YB, Heilig R, Fischer R, Thompson AJ, Goswami A, Rogelj B. Interactome screening of C9orf72 dipeptide repeats reveals VCP sequestration and functional impairment by polyGA. Brain. 2021 Sep 17:awab300. Doi: 10.1093/brain/awab300. Epub ahead of print. PMID: 34534264.

Bragg DC, Mangkalaphiban K, Vaine CA, Kulkarni NJ, Shin D, Yadav R, Dhakal J, Ton ML, Cheng A, Russo CT, Ang M, Acuña P, Go C, Franceour TN, Multhaupt-Buell T, Ito N, Müller U, Hendriks WT, Breakefield XO, Sharma N, Ozelius LJ. Disease onset in X-linked dystonia-parkinsonism correlates with expansion of a hexameric repeat within an SVA retrotransposon in *TAF1*. Proc Natl Acad Sci U S A. 2017 Dec 19;114(51):E11020-E11028. Doi: 10.1073/pnas.1712526114. Epub 2017 Dec 11. Erratum in: Proc Natl Acad Sci U S A. 2020 Mar 17;117(11):6277. PMID: 29229810; PMCID: PMC5754783.

Brais B, Bouchard JP, Xie YG, Rochefort DL, Chrétien N, Tomé FM, Lafrenière RG, Rommens JM, Uyama E, Nohira O, Blumen S, Korczyn AD, Heutink P, Mathieu J, Duranceau A, Codère F, Fardeau M, Rouleau GA. Short GCG expansions in the PABP2 gene cause oculopharyngeal muscular dystrophy. Nat Genet. 1998 Feb;18(2):164-7. Doi: 10.1038/ng0298-164. Erratum in: Nat Genet 1998 Aug;19(4):404. Korcyn AD [corrected to Korczyn AD]. PMID: 9462747.

Brasseur L, Coens A, Waeytens J, Melki R, Bousset L. Dipeptide repeat derived from C9orf72 hexanucleotide expansions forms amyloids or natively unfolded structures in vitro. Biochem Biophys Res Commun. 2020 May 28;526(2):410-416. Doi: 10.1016/j.bbrc.2020.03.108. Epub 2020 Mar 27. PMID: 32223927.

Brenner D, Yilmaz R, Müller K, Grehl T, Petri S, Meyer T, Grosskreutz J, Weydt P, Ruf W, Neuwirth C, Weber M, Pinto S, Claeys KG, Schrank B, Jordan B, Knehr A, Günther K, Hübers A, Zeller D, Kubisch C, Jablonka S, Sendtner M, Klopstock T, de Carvalho M, Sperfeld A, Borck G, Volk AE, Dorst J, Weis J, Otto M, Schuster J, Del Tredici K, Braak H, Danzer KM, Freischmidt A, Meitinger T, Strom TM, Ludolph AC, Andersen PM, Weishaupt JH; German ALS network MND-NET. Hot-spot KIF5A mutations cause familial ALS. Brain. 2018 Mar 1;141(3):688-697. Doi: 10.1093/brain/awx370. PMID: 29342275; PMCID: PMC5837483.

Bretherick KL, Fluker MR, Robinson WP. FMR1 repeat sizes in the gray zone and high end of the normal range are associated with premature ovarian failure. Hum Genet. 2005 Aug;117(4):376-82. Doi: 10.1007/s00439-005-1326-8. Epub 2005 Jun 2. PMID: 16078053.

Buhaescu I, Izzedine H, Covic A. Sirolimus—challenging current perspectives. Ther Drug Monit. 2006 Oct;28(5):577-84. Doi: 10.1097/01.ftd.0000245377.93401.39. PMID: 17038868.

Bui C, Huber C, Tuysuz B, Alanay Y, Bole-Feysot C, Leroy JG, Mortier G, Nitschke P, Munnich A, Cormier-Daire V. XYLT1 mutations in Desbuquois dysplasia type 2. Am J Hum Genet. 2014 Mar 6;94(3):405-14. Doi: 10.1016/j.ajhg.2014.01.020. Epub 2014 Feb 27. PMID: 24581741; PMCID: PMC3951945.

Buijsen RA, Sellier C, Severijnen LA, Oulad-Abdelghani M, Verhagen RF, Berman RF, Charlet-Berguerand N, Willemsen R, Hukema RK. FMRpolyG-positive inclusions in CNS and non-CNS organs of a fragile X premutation carrier with fragile X-associated tremor/ataxia syndrome. Acta Neuropathol Commun. 2014 Nov 26;2:162. Doi: 10.1186/s40478-014-0162-2. PMID: 25471011; PMCID: PMC4254384.

Buijsen RA, Visser JA, Kramer P, Severijnen EA, Gearing M, Charlet-Berguerand N, Sherman SL, Berman RF, Willemsen R, Hukema RK. Presence of inclusions positive for polyglycine containing protein, FMRpolyG, indicates that repeat-associated non-AUG translation plays a role in fragile X-associated primary ovarian insufficiency. Hum Reprod. 2016 Jan;31(1):158-68. Doi: 10.1093/humrep/dev280. Epub 2015 Nov 3. PMID: 26537920; PMCID: PMC4677964.

Burberry A, Suzuki N, Wang JY, Moccia R, Mordes DA, Stewart MH, Suzuki-Uematsu S, Ghosh S, Singh A, Merkle FT, Koszka K, Li QZ, Zon L, Rossi DJ, Trowbridge JJ, Notarangelo LD, Eggan K. Loss-of-function mutations in the C9ORF72 mouse ortholog cause fatal autoimmune disease.

Sci Transl Med. 2016 Jul 13;8(347):347ra93. Doi: 10.1126/scitranslmed.aaf6038. PMID: 27412785; PMCID: PMC5024536.

Bush JA, Aikawa H, Fuerst R, Li Y, Ursu A, Meyer SM, Benhamou RI, Chen JL, Khan T, Wagner-Griffin S, Van Meter MJ, Tong Y, Olafson H, McKee KK, Childs-Disney JL, Gendron TF, Zhang Y, Coyne AN, Wang ET, Yildirim I, Wang KW, Petrucelli L, Rothstein JD, Disney MD. Ribonuclease recruitment using a small molecule reduced c9ALS/FTD r(G_4C_2) repeat expansion in vitro and in vivo ALS models. Sci Transl Med. 2021 Oct 27;13(617):eabd5991. Doi: 10.1126/scitranslmed.abd5991. Epub 2021 Oct 27. PMID: 34705518.

Butti Z, Pan YE, Giacomotto J, Patten SA. Reduced C9orf72 function leads to defective synaptic vesicle release and neuromuscular dysfunction in zebrafish. Commun Biol. 2021 Jun 25;4(1):792. Doi: 10.1038/s42003-021-02302-y. PMID: 34172817; PMCID: PMC8233344.

C

Calado A, Tomé FM, Brais B, Rouleau GA, Kühn U, Wahle E, Carmo-Fonseca M. Nuclear inclusions in oculopharyngeal muscular dystrophy consist of poly(A) binding protein 2 aggregates which sequester poly(A) RNA. Hum Mol Genet. 2000 Sep 22;9(15):2321-8. doi: 10.1093/oxfordjournals.hmg.a018924. PMID: 11001936.

Campbell L, Potter A, Ignatius J, Dubowitz V, Davies K. Genomic variation and gene conversion in spinal muscular atrophy: implications for disease process and clinical phenotype. Am J Hum Genet. 1997 Jul;61(1):40-50. doi: 10.1086/513886. PMID: 9245983; PMCID: PMC1715870.

Campuzano V, Montermini L, Moltò MD, Pianese L, Cossée M, Cavalcanti F, Monros E, Rodius F, Duclos F, Monticelli A, Zara F, Cañizares J, Koutnikova H, Bidichandani SI, Gellera C, Brice A, Trouillas P, De Michele G, Filla A, De Frutos R, Palau F, Patel PI, Di Donato S, Mandel JL, Cocozza S, Koenig M, Pandolfo M. Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science. 1996 Mar 8;271(5254):1423-7. doi: 10.1126/science.271.5254.1423. PMID: 8596916.

Castro IH, Pignataro MF, Sewell KE, Espeche LD, Herrera MG, Noguera ME, Dain L, Nadra AD, Aran M, Smal C, Gallo M, Santos J. Frataxin Structure and Function. Subcell Biochem. 2019;93:393-438. doi: 10.1007/978-3-030-28151-9_13. PMID: 31939159.

Centenera MM, Harris JM, Tilley WD, Butler LM. The contribution of different androgen receptor domains to receptor dimerization and signaling. Mol Endocrinol. 2008 Nov;22(11):2373-82. doi: 10.1210/me.2008-0017. Epub 2008 Jul 10. PMID: 18617596.

Chang YJ, Jeng US, Chiang YL, Hwang IS, Chen YR. The Glycine-Alanine Dipeptide Repeat from C9orf72 Hexanucleotide Expansions Forms Toxic Amyloids Possessing Cell-to-Cell Transmission Properties. J Biol Chem. 2016 Mar 4;291(10):4903-11. doi: 10.1074/jbc.M115.694273. Epub 2016 Jan 14. PMID: 26769963; PMCID: PMC4777828.

Charlet-B N, Savkur RS, Singh G, Philips AV, Grice EA, Cooper TA. Loss of the muscle-specific chloride channel in type 1 myotonic dystrophy due to misregulated alternative splicing. Mol Cell. 2002 Jul;10(1):45-53. doi: 10.1016/s1097-2765(02)00572-5. PMID: 12150906.

Chen Z, Yan Yau W, Jaunmuktane Z, Tucci A, Sivakumar P, Gagliano Taliun SA, Turner C, Efthymiou S, Ibáñez K, Sullivan R, Bibi F, Athanasiou-Fragkouli A, Bourinaris T, Zhang D, Revesz T, Lashley T, DeTure M, Dickson DW, Josephs KA, Gelpi E, Kovacs GG, Halliday G, Rowe DB, Blair I, Tienari PJ, Suomalainen A, Fox NC, Wood NW, Lees AJ, Haltia MJ; Genomics England Research Consortium, Hardy J, Ryten M, Vandrovcova J, Houlden H. Neuronal intranuclear inclusion disease is genetically heterogeneous. Ann Clin Transl Neurol. 2020a Sep;7(9):1716-1725. doi: 10.1002/acn3.51151. Epub 2020 Aug 10. PMID: 32777174; PMCID: PMC7480908.

Chen H, Lu L, Wang B, Cui G, Wang X, Wang Y, Raza HK, Min Y, Li K, Cui Y, Miao Z, Wan B, Sun M, Xu X. Re-defining the clinicopathological spectrum of neuronal intranuclear inclusion disease. Ann Clin Transl Neurol. 2020b Oct;7(10):1930-1941. doi: 10.1002/acn3.51189. Epub 2020 Sep 15. PMID: 32931652; PMCID: PMC7545592.

Chew J, Gendron TF, Prudencio M, Sasaguri H, Zhang YJ, Castanedes-Casey M, Lee CW, Jansen-West K, Kurti A, Murray ME, Bieniek KF, Bauer PO, Whitelaw EC, Rousseau L, Stankowski JN, Stetler C, Daughrity LM, Perkerson EA, Desaro P, Johnston A, Overstreet K, Edbauer D,

Rademakers R, Boylan KB, Dickson DW, Fryer JD, Petrucelli L. Neurodegeneration. C9ORF72 repeat expansions in mice cause TDP-43 pathology, neuronal loss, and behavioral deficits. Science. 2015 Jun 5;348(6239):1151-4. doi: 10.1126/science.aaa9344. Epub 2015 May 14. PMID: 25977373; PMCID: PMC4692360.

Chiò A, Calvo A, Moglia C, Canosa A, Brunetti M, Barberis M, Restagno G, Conte A, Bisogni G, Marangi G, Moncada A, Lattante S, Zollino M, Sabatelli M, Bagarotti A, Corrado L, Mora G, Bersano E, Mazzini L, D'Alfonso S; PARALS. ATXN2 polyQ intermediate repeats are a modifier of ALS survival. Neurology. 2015 Jan 20;84(3):251-8. doi: 10.1212/WNL.0000000000001159. Epub 2014 Dec 19. PMID: 25527265.

Chiriboga CA, Swoboda KJ, Darras BT, Iannaccone ST, Montes J, De Vivo DC, Norris DA, Bennett CF, Bishop KM. Results from a phase 1 study of nusinersen (ISIS-SMN(Rx)) in children with spinal muscular atrophy. Neurology. 2016 Mar 8;86(10):890-7. doi: 10.1212/WNL.0000000000002445. Epub 2016 Feb 10. PMID: 26865511; PMCID: PMC4782111.

Choi SY, Lopez-Gonzalez R, Krishnan G, Phillips HL, Li AN, Seeley WW, Yao WD, Almeida S, Gao FB. C9ORF72-ALS/FTD-associated poly(GR) binds Atp5a1 and compromises mitochondrial function in vivo. Nat Neurosci. 2019 Jun;22(6):851-862. doi: 10.1038/s41593-019-0397-0. Epub 2019 May 13. PMID: 31086314; PMCID: PMC6800116.

Chowdhury S, Otomo C, Leitner A, Ohashi K, Aebersold R, Lander GC, Otomo T. Insights into autophagosome biogenesis from structural and biochemical analyses of the ATG2A-WIPI4 complex. Proc Natl Acad Sci U S A. 2018 Oct 16;115(42):E9792-E9801. doi: 10.1073/pnas.1811874115. Epub 2018 Sep 5. PMID: 30185561; PMCID: PMC6196511.

Chu Q, Martinez TF, Novak SW, Donaldson CJ, Tan D, Vaughan JM, Chang T, Diedrich JK, Andrade L, Kim A, Zhang T, Manor U, Saghatelian A. Regulation of the ER stress response by a mitochondrial microprotein. Nat Commun. 2019 Oct 25;10(1):4883. doi: 10.1038/s41467-019-12816-z. PMID: 31653868; PMCID: PMC6814811.

Cirulli ET, Lasseigne BN, Petrovski S, Sapp PC, Dion PA, Leblond CS, Couthouis J, Lu YF, Wang Q, Krueger BJ, Ren Z, Keebler J, Han Y, Levy SE, Boone BE, Wimbish JR, Waite LL, Jones AL, Carulli JP, Day-Williams AG, Staropoli JF, Xin WW, Chesi A, Raphael AR, McKenna-Yasek D, Cady J, Vianney de Jong JM, Kenna KP, Smith BN, Topp S, Miller J, Gkazi A; FALS Sequencing Consortium, Al-Chalabi A, van den Berg LH, Veldink J, Silani V, Ticozzi N, Shaw CE, Baloh RH, Appel S, Simpson E, Lagier-Tourenne C, Pulst SM, Gibson S, Trojanowski JQ, Elman L, McCluskey L, Grossman M, Shneider NA, Chung WK, Ravits JM, Glass JD, Sims KB, Van Deerlin VM, Maniatis T, Hayes SD, Ordureau A, Swarup S, Landers J, Baas F, Allen AS, Bedlack RS, Harper JW, Gitler AD, Rouleau GA, Brown R, Harms MB, Cooper GM, Harris T, Myers RM, Goldstein DB. Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. Science. 2015 Mar 27;347(6229):1436-41. doi: 10.1126/science.aaa3650. Epub 2015 Feb 19. PMID: 25700176; PMCID: PMC4437632.

Ciura S, Lattante S, Le Ber I, Latouche M, Tostivint H, Brice A, Kabashi E. Loss of function of C9orf72 causes motor deficits in a zebrafish model of amyotrophic lateral sclerosis. Ann Neurol. 2013 Aug;74(2):180-7. doi: 10.1002/ana.23946. PMID: 23720273.

Cleary JD, Pattamatta A, Ranum LPW. Repeat-associated non-ATG (RAN) translation. J Biol Chem. 2018 Oct 19;293(42):16127-16141. doi: 10.1074/jbc.R118.003237. Epub 2018 Sep 13. PMID: 30213863; PMCID: PMC6200949.

Clements JM, Laz TM, Sherman F. Efficiency of translation initiation by non-AUG codons in Saccharomyces cerevisiae. Mol Cell Biol. 1988 Oct;8(10):4533-6. doi: 10.1128/mcb.8.10.4533-4536.1988. PMID: 3141793; PMCID: PMC365530.

Coffee B, Zhang F, Warren ST, Reines D. Acetylated histones are associated with FMR1 in normal but not fragile X-syndrome cells. Nat Genet. 1999 May;22(1):98-101. doi: 10.1038/8807. Erratum in: Nat Genet 1999 Jun;22(2):209. PMID: 10319871.

Comery TA, Harris JB, Willems PJ, Oostra BA, Irwin SA, Weiler IJ, Greenough WT. Abnormal dendritic spines in fragile X knockout mice: maturation and pruning deficits. Proc Natl Acad Sci U S A. 1997 May 13;94(10):5401-4. doi: 10.1073/pnas.94.10.5401. PMID: 9144249; PMCID: PMC24690.

Congdon EE, Sigurdsson EM. Tau-targeting therapies for Alzheimer disease. Nat Rev Neurol. 2018 Jul;14(7):399-415. doi: 10.1038/s41582-018-0013-z. PMID: 29895964; PMCID: PMC6463489.

Conlon EG, Lu L, Sharma A, Yamazaki T, Tang T, Shneider NA, Manley JL. The C9ORF72 GGGGCC expansion forms RNA G-quadruplex inclusions and sequesters hnRNP H to disrupt splicing in ALS brains. Elife. 2016 Sep 13;5:e17820. doi: 10.7554/eLife.17820. PMID: 27623008; PMCID: PMC5050020.

Conlon EG, Fagegaltier D, Agius P, Davis-Porada J, Gregory J, Hubbard I, Kang K, Kim D; New York Genome Center ALS Consortium, Phatnani H, Shneider NA, Manley JL. Unexpected similarities between C9ORF72 and sporadic forms of ALS/FTD suggest a common disease mechanism. Elife. 2018 Jul 13;7:e37754. doi: 10.7554/eLife.37754. PMID: 30003873; PMCID: PMC6103746.

Cook CN, Wu Y, Odeh HM, Gendron TF, Jansen-West K, Del Rosso G, Yue M, Jiang P, Gomes E, Tong J, Daughrity LM, Avendano NM, Castanedes-Casey M, Shao W, Oskarsson B, Tomassy GS, McCampbell A, Rigo F, Dickson DW, Shorter J, Zhang YJ, Petrucelli L. *C9orf72* poly(GR) aggregation induces TDP-43 proteinopathy. Sci Transl Med. 2020 Sep 2;12(559):eabb3774. doi: 10.1126/scitranslmed.abb3774. PMID: 32878979; PMCID: PMC7989020.

Cooper-Knock J, Walsh MJ, Higginbottom A, Robin Highley J, Dickman MJ, Edbauer D, Ince PG, Wharton SB, Wilson SA, Kirby J, Hautbergue GM, Shaw PJ. Sequestration of multiple RNA recognition motif-containing proteins by C9orf72 repeat expansions. Brain. 2014 Jul;137(Pt 7):2040-51. doi: 10.1093/brain/awu120. Epub 2014 May 27. PMID: 24866055; PMCID: PMC4065024.

Cooper-Knock J, Moll T, Ramesh T, Castelli L, Beer A, Robins H, Fox I, Niedermoser I, Van Damme P, Moisse M, Robberecht W, Hardiman O, Panades MP, Assialioui A, Mora JS, Basak AN, Morrison KE, Shaw CE, Al-Chalabi A, Landers JE, Wyles M, Heath PR, Higginbottom A, Walsh T, Kazoka M, McDermott CJ, Hautbergue GM, Kirby J, Shaw PJ. Mutations in the Glycosyltransferase Domain of GLT8D1 Are Associated with Familial Amyotrophic Lateral Sclerosis. Cell Rep. 2019 Feb 26;26(9):2298-2306.e5. doi: 10.1016/j.celrep.2019.02.006. PMID: 30811981; PMCID: PMC7003067.

Corbett MA, Kroes T, Veneziano L, Bennett MF, Florian R, Schneider AL, Coppola A, Licchetta L, Franceschetti S, Suppa A, Wenger A, Mei D, Pendziwiat M, Kaya S, Delledonne M, Straussberg R, Xumerle L, Regan B, Crompton D, van Rootselaar AF, Correll A, Catford R, Bisulli F, Chakraborty S, Baldassari S, Tinuper P, Barton K, Carswell S, Smith M, Berardelli A, Carroll R, Gardner A, Friend KL, Blatt I, Iacomino M, Di Bonaventura C, Striano S, Buratti J, Keren B, Nava C, Forlani S, Rudolf G, Hirsch E, Leguern E, Labauge P, Balestrini S, Sander JW, Afawi Z, Helbig I, Ishiura H, Tsuji S, Sisodiya SM, Casari G, Sadleir LG, van Coller R, Tijssen MAJ, Klein KM, van den Maagdenberg AMJM, Zara F, Guerrini R, Berkovic SF, Pippucci T, Canafoglia L, Bahlo M, Striano P, Scheffer IE, Brancati F, Depienne C, Gecz J. Intronic ATTTC repeat expansions in STARD7 in familial adult myoclonic epilepsy linked to chromosome 2. Nat Commun. 2019 Oct 29;10(1):4920. doi: 10.1038/s41467-019-12671-y. PMID: 31664034; PMCID: PMC6820779.

Corman A, Jung B, Häggblad M, Bräutigam L, Lafarga V, Lidemalm L, Hühn D, Carreras-Puigvert J, Fernandez-Capetillo O. A Chemical Screen Identifies Compounds Limiting the Toxicity of C9ORF72 Dipeptide Repeats. Cell Chem Biol. 2019 Feb 21;26(2):235-243.e5. doi: 10.1016/j.chembiol.2018.10.020. Epub 2018 Dec 6. PMID: 30527999.

Cortese A, Simone R, Sullivan R, Vandrovcova J, Tariq H, Yau WY, Humphrey J, Jaunmuktane Z, Sivakumar P, Polke J, Ilyas M, Tribollet E, Tomaselli PJ, Devigili G, Callegari I, Versino M, Salpietro V, Efthymiou S, Kaski D, Wood NW, Andrade NS, Buglo E, Rebelo A, Rossor AM, Bronstein A, Fratta P, Marques WJ, Züchner S, Reilly MM, Houlden H. Biallelic expansion of an intronic repeat in RFC1 is a common cause of late-onset ataxia. Nat Genet. 2019 Apr;51(4):649-658. doi: 10.1038/s41588-019-0372-4. Epub 2019 Mar 29. Erratum in: Nat Genet. 2019 May;51(5):920. PMID: 30926972; PMCID: PMC6709527.

Coyle-Gilchrist IT, Dick KM, Patterson K, Vázquez Rodríquez P, Wehmann E, Wilcox A, Lansdall CJ, Dawson KE, Wiggins J, Mead S, Brayne C, Rowe JB. Prevalence, characteristics, and survival of frontotemporal lobar degeneration syndromes. Neurology. 2016 May 3;86(18):1736-43. doi: 10.1212/WNL.0000000000002638. Epub 2016 Apr 1. PMID: 27037234; PMCID: PMC4854589.

Crawford DC, Acuña JM, Sherman SL. FMR1 and the fragile X syndrome: human genome epidemiology review. Genet Med. 2001 Sep-Oct;3(5):359-71. doi: 10.1097/00125817-200109000-00006. PMID: 11545690; PMCID: PMC4493892.

Cummings CJ, Reinstein E, Sun Y, Antalffy B, Jiang Y, Ciechanover A, Orr HT, Beaudet AL, Zoghbi HY. Mutation of the E6-AP ubiquitin ligase reduces nuclear inclusion frequency while accelerating polyglutamine-induced pathology in SCA1 mice. Neuron. 1999 Dec;24(4):879-92. doi: 10.1016/s0896-6273(00)81035-1. PMID: 10624951.

D

Dafinca R, Scaber J, Ababneh N, Lalic T, Weir G, Christian H, Vowles J, Douglas AG, Fletcher-Jones A, Browne C, Nakanishi M, Turner MR, Wade-Martins R, Cowley SA, Talbot K. C9orf72 Hexanucleotide Expansions Are Associated with Altered Endoplasmic Reticulum Calcium Homeostasis and Stress Granule Formation in Induced Pluripotent Stem Cell-Derived Neurons from Patients with Amyotrophic Lateral Sclerosis and Frontotemporal Dementia. Stem Cells. 2016 Aug;34(8):2063-78. Doi: 10.1002/stem.2388. Epub 2016 May 4. PMID: 27097283; PMCID: PMC4979662.

Dafinca R, Barbagallo P, Farrimond L, Candalija A, Scaber J, Ababneh NA, Sathyaprakash C, Vowles J, Cowley SA, Talbot K. Impairment of Mitochondrial Calcium Buffering Links Mutations in C9ORF72 and TARDBP in iPS-Derived Motor Neurons from Patients with ALS/FTD. Stem Cell Reports. 2020 May 12;14(5):892-908. Doi: 10.1016/j.stemcr.2020.03.023. Epub 2020 Apr 23. PMID: 32330447; PMCID: PMC7220989.

Damme M, Suntio T, Saftig P, Eskelinen EL. Autophagy in neuronal cells: general principles and physiological and pathological functions. Acta Neuropathol. 2015 Mar;129(3):337-62. Doi: 10.1007/s00401-014-1361-4. Epub 2014 Nov 4. PMID: 25367385.

Dansithong W, Paul S, Comai L, Reddy S. MBNL1 is the primary determinant of focus formation and aberrant insulin receptor splicing in DM1. J Biol Chem. 2005;280:5773–5780. Doi: 10.1074/jbc.M410781200

Darnell JC, Jensen KB, Jin P, Brown V, Warren ST, Darnell RB. Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. Cell. 2001 Nov 16;107(4):489-99. Doi: 10.1016/s0092-8674(01)00566-9. PMID: 11719189.

Darnell JC, Fraser CE, Mostovetsky O, Stefani G, Jones TA, Eddy SR, Darnell RB. Kissing complex RNAs mediate interaction between the Fragile-X mental retardation protein KH2 domain and brain polyribosomes. Genes Dev. 2005 Apr 15;19(8):903-18. Doi: 10.1101/gad.1276805. Epub 2005 Apr 1. PMID: 15805463; PMCID: PMC1080130.

Darras BT, Chiriboga CA, Iannaccone ST, Swoboda KJ, Montes J, Mignon L, Xia S, Bennett CF, Bishop KM, Shefner JM, Green AM, Sun P, Bhan I, Gheuens S, Schneider E, Farwell W, De Vivo DC; ISIS-396443-CS2/ISIS-396443-CS12 Study Groups. Nusinersen in later-onset spinal muscular atrophy: Long-term results from the phase ½ studies. Neurology. 2019 May 21;92(21):e2492-e2506. Doi: 10.1212/WNL.0000000000007527. Epub 2019 Apr 24. PMID: 31019106; PMCID: PMC6541434.

Daoud H, Belzil V, Martins S, Sabbagh M, Provencher P, Lacomblez L, Meininger V, Camu W, Dupré N, Dion PA, Rouleau GA. Association of long ATXN2 CAG repeat sizes with increased risk of amyotrophic lateral sclerosis. Arch Neurol. 2011 Jun;68(6):739-42. Doi: 10.1001/archneurol.2011.111. PMID: 21670397.

Davidson Y, Robinson AC, Liu X, Wu D, Troakes C, Rollinson S, Masuda-Suzukake M, Suzuki G, Nonaka T, Shi J, Tian J, Hamdalla H, Ealing J, Richardson A, Jones M, Pickering-Brown S, Snowden JS, Hasegawa M, Mann DM. Neurodegeneration in frontotemporal lobar degeneration and motor neurone disease associated with expansions in C9orf72 is linked to TDP-43 pathology and not associated with aggregated forms of dipeptide repeat proteins. Neuropathol Appl Neurobiol. 2016 Apr;42(3):242-54. Doi: 10.1111/nan.12292. Epub 2015 Dec 7. PMID: 26538301; PMCID: PMC4832296.

Davidson YS, Flood L, Robinson AC, Nihei Y, Mori K, Rollinson S, Richardson A, Benson BC, Jones M, Snowden JS, Pickering-Brown S, Haass C, Lashley T, Mann DMA. Heterogeneous ribonuclear protein A3 (hnRNP A3) is present in dipeptide repeat protein containing inclusions in Frontotemporal Lobar Degeneration and Motor Neurone disease associated with expansions in C9orf72 gene. Acta Neuropathol Commun. 2017 Apr 21;5(1):31. Doi: 10.1186/s40478-017-0437-5. PMID: 28431575; PMCID: PMC5399321.

De Boulle K, Verkerk AJ, Reyniers E, Vits L, Hendrickx J, Van Roy B, Van den Bos F, de Graaff E, Oostra BA, Willems PJ. A point mutation in the FMR-1 gene associated with fragile X mental retardation. Nat Genet. 1993 Jan;3(1):31-5. Doi: 10.1038/ng0193-31. PMID: 8490650.

DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford NJ, Nicholson AM, Finch NA, Flynn H, Adamson J, Kouri N, Wojtas A, Sengdy P, Hsiung GY, Karydas A, Seeley WW, Josephs KA, Coppola G, Geschwind DH, Wszolek ZK, Feldman H, Knopman DS, Petersen RC, Miller BL, Dickson DW, Boylan KB, Graff-Radford NR, Rademakers R. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron. 2011 Oct 20;72(2):245-56. Doi: 10.1016/j.neuron.2011.09.011. Epub 2011 Sep 21. PMID: 21944778; PMCID: PMC3202986.

Deng HX, Zhai H, Bigio EH, Yan J, Fecto F, Ajroud K, Mishra M, Ajroud-Driss S, Heller S, Sufit R, Siddique N, Mugnaini E, Siddique T. FUS-immunoreactive inclusions are a common feature in sporadic and non-SOD1 familial amyotrophic lateral sclerosis. Ann Neurol. 2010 Jun;67(6):739-48. Doi: 10.1002/ana.22051. PMID: 20517935; PMCID: PMC4376270.

Deng HX, Chen W, Hong ST, Boycott KM, Gorrie GH, Siddique N, Yang Y, Fecto F, Shi Y, Zhai H, Jiang H, Hirano M, Rampersaud E, Jansen GH, Donkervoort S, Bigio EH, Brooks BR, Ajroud K, Sufit RL, Haines JL, Mugnaini E, Pericak-Vance MA, Siddique T. Mutations in UBQLN2 cause dominant X-linked juvenile and adult-onset ALS and ALS/dementia. Nature. 2011 Aug 21;477(7363):211-5. Doi: 10.1038/nature10353. PMID: 21857683; PMCID: PMC3169705.

Deng J, Gu M, Miao Y, Yao S, Zhu M, Fang P, Yu X, Li P, Su Y, Huang J, Zhang J, Yu J, Li F, Bai J, Sun W, Huang Y, Yuan Y, Hong D, Wang Z. Long-read sequencing identified repeat expansions in the 5'UTR of the *NOTCH2NLC* gene from Chinese patients with neuronal intranuclear inclusion disease. J Med Genet. 2019 Nov;56(11):758-764. Doi: 10.1136/jmedgenet-2019-106268. Epub 2019 Aug 14. PMID: 31413119.

Deng J, Yu J, Li P, Luan X, Cao L, Zhao J, Yu M, Zhang W, Lv H, Xie Z, Meng L, Zheng Y, Zhao Y, Gang Q, Wang Q, Liu J, Zhu M, Guo X, Su Y, Liang Y, Liang F, Hayashi T, Maeda MH, Sato T, Ura S, Oya Y, Ogasawara M, Iida A, Nishino I, Zhou C, Yan C, Yuan Y, Hong D, Wang Z. Expansion of GGC Repeat in GIPC1 Is Associated with Oculopharyngodistal Myopathy. Am J Hum Genet. 2020 Jun 4;106(6):793-804. Doi: 10.1016/j.ajhg.2020.04.011. Epub 2020 May 14. PMID: 32413282; PMCID: PMC7273532.

Depienne C, Mandel JL. 30 years of repeat expansion disorders: What have we learned and what are the remaining challenges? Am J Hum Genet. 2021 May 6;108(5):764-785. Doi: 10.1016/j.ajhg.2021.03.011. Epub 2021 Apr 2. PMID: 33811808; PMCID: PMC8205997.

Derbis M, Kul E, Niewiadomska D, Sekrecki M, Piasecka A, Taylor K, Hukema RK, Stork O, Sobczak K. Short antisense oligonucleotides alleviate the pleiotropic toxicity of RNA harboring expanded CGG repeats. Nat Commun. 2021 Feb 24;12(1):1265. Doi: 10.1038/s41467-021-21021-w. PMID: 33627639; PMCID: PMC7904788.

Didiot MC, Tian Z, Schaeffer C, Subramanian M, Mandel JL, Moine H. The G-quartet containing FMRP binding site in FMR1 mRNA is a potent exonic splicing enhancer. Nucleic Acids Res. 2008 Sep;36(15):4902-12. Doi: 10.1093/nar/gkn472. Epub 2008 Jul 24. PMID: 18653529; PMCID: PMC2528169.

DiFiglia M, Sapp E, Chase KO, Davies SW, Bates GP, Vonsattel JP, Aronin N. Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. Science. 1997 Sep 26;277(5334):1990-3. Doi: 10.1126/science.277.5334.1990. PMID: 9302293.

Disney MD, Liu B, Yang WY, Sellier C, Tran T, Charlet-Berguerand N, Childs-Disney JL. A small molecule that targets r(CGG)(exp) and improves defects in fragile X-associated tremor ataxia syndrome. ACS Chem Biol. 2012 Oct 19;7(10):1711-8. doi: 10.1021/cb300135h. Epub 2012 Sep 4. PMID: 22948243; PMCID: PMC3477254.

Donnelly CJ, Zhang PW, Pham JT, Haeusler AR, Mistry NA, Vidensky S, Daley EL, Poth EM, Hoover B, Fines DM, Maragakis N, Tienari PJ, Petrucelli L, Traynor BJ, Wang J, Rigo F, Bennett CF, Blackshaw S, Sattler R, Rothstein JD. RNA toxicity from the ALS/FTD C9ORF72 expansion is mitigated by antisense intervention. Neuron. 2013 Oct 16;80(2):415-28. doi: 10.1016/j.neuron.2013.10.015. Erratum in: Neuron. 2013 Nov 20;80(4):1102. Heusler, Aaron R [corrected to Haeusler, Aaron R]. PMID: 24139042; PMCID: PMC4098943.

Du J, Aleff RA, Soragni E, Kalari K, Nie J, Tang X, Davila J, Kocher JP, Patel SV, Gottesfeld JM, Baratz KH, Wieben ED. RNA toxicity and missplicing in the common eye disease fuchs endothelial corneal dystrophy. J Biol Chem. 2015 Mar 6;290(10):5979-90. doi: 10.1074/jbc.M114.621607. Epub 2015 Jan 15. PMID: 25593321; PMCID: PMC4358235.

Durmus H, Laval SH, Deymeer F, Parman Y, Kiyan E, Gokyigiti M, Ertekin C, Ercan I, Solakoglu S, Karcagi V, Straub V, Bushby K, Lochmüller H, Serdaroglu-Oflazer P. Oculopharyngodistal myopathy is a distinct entity: clinical and genetic features of 47 patients. Neurology. 2011 Jan 18;76(3):227-35. doi: 10.1212/WNL.0b013e318207b043. PMID: 21242490.

Dürr A, Cossee M, Agid Y, Campuzano V, Mignard C, Penet C, Mandel JL, Brice A, Koenig M. Clinical and genetic abnormalities in patients with Friedreich's ataxia. N Engl J Med. 1996 Oct 17;335(16):1169-75. doi: 10.1056/NEJM199610173351601. PMID: 8815938.

Duyao MP, Auerbach AB, Ryan A, Persichetti F, Barnes GT, McNeil SM, Ge P, Vonsattel JP, Gusella JF, Joyner AL, et al. Inactivation of the mouse Huntington's disease gene homolog Hdh. Science. 1995 Jul 21;269(5222):407-10. doi: 10.1126/science.7618107. PMID: 7618107.

E

Eberhart DE, Malter HE, Feng Y, Warren ST. The fragile X mental retardation protein is a ribonucleoprotein containing both nuclear localization and nuclear export signals. Hum Mol Genet. 1996 Aug;5(8):1083-91. doi: 10.1093/hmg/5.8.1083. PMID: 8842725.

Eichler L, Bellenberg B, Hahn HK, Köster O, Schöls L, Lukas C. Quantitative assessment of brain stem and cerebellar atrophy in spinocerebellar ataxia types 3 and 6: impact on clinical status. AJNR Am J Neuroradiol. 2011 May;32(5):890-7. doi: 10.3174/ajnr.A2387. Epub 2011 Mar 3. PMID: 21372168; PMCID: PMC7965570.

Elden AC, Kim HJ, Hart MP, Chen-Plotkin AS, Johnson BS, Fang X, Armakola M, Geser F, Greene R, Lu MM, Padmanabhan A, Clay-Falcone D, McCluskey L, Elman L, Juhr D, Gruber PJ, Rüb U, Auburger G, Trojanowski JQ, Lee VM, Van Deerlin VM, Bonini NM, Gitler AD. Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. Nature. 2010 Aug 26;466(7310):1069-75. doi: 10.1038/nature09320. PMID: 20740007; PMCID: PMC2965417.

Ellerby LM. Repeat Expansion Disorders: Mechanisms and Therapeutics. Neurotherapeutics. 2019 Oct;16(4):924-927. doi: 10.1007/s13311-019-00823-3. PMID: 31907874; PMCID: PMC6985307.

Emamian ES, Kaytor MD, Duvick LA, Zu T, Tousey SK, Zoghbi HY, Clark HB, Orr HT. Serine 776 of ataxin-1 is critical for polyglutamine-induced disease in SCA1 transgenic mice. Neuron. 2003 May 8;38(3):375-87. doi: 10.1016/s0896-6273(03)00258-7. PMID: 12741986.

Ezzi SA, Larivière R, Urushitani M, Julien JP. Neuronal over-expression of chromogranin A accelerates disease onset in a mouse model of ALS. J Neurochem. 2010 Dec;115(5):1102-11. doi: 10.1111/j.1471-4159.2010.06979.x. Epub 2010 Oct 26. PMID: 20807312.

F

Farg MA, Soo KY, Warraich ST, Sundaramoorthy V, Blair IP, Atkin JD. Ataxin-2 interacts with FUS and intermediate-length polyglutamine expansions enhance FUS-related pathology in amyotrophic lateral sclerosis. Hum Mol Genet. 2013 Feb 15;22(4):717-28. Doi: 10.1093/hmg/dds479. Epub 2012 Nov 19. Erratum in: Hum Mol Genet. 2020 Mar 13;29(4):703-704. PMID: 23172909.

Farg MA, Sundaramoorthy V, Sultana JM, Yang S, Atkinson RA, Levina V, Halloran MA, Gleeson PA, Blair IP, Soo KY, King AE, Atkin JD. C9ORF72, implicated in amytrophic lateral sclerosis and frontotemporal dementia, regulates endosomal trafficking. Hum Mol Genet. 2014 Jul 1;23(13):3579-95. Doi: 10.1093/hmg/ddu068. Epub 2014 Feb 18. Erratum in: Hum Mol Genet. 2017 Oct 15;26(20):4093-4094. PMID: 24549040; PMCID: PMC4049310.

Farg MA, Konopka A, Soo KY, Ito D, Atkin JD. The DNA damage response (DDR) is induced by the C9orf72 repeat expansion in amyotrophic lateral sclerosis. Hum Mol Genet. 2017 Aug 1;26(15):2882-2896. Doi: 10.1093/hmg/ddx170. PMID: 28481984.

Farhan SMK, Howrigan DP, Abbott LE, Klim JR, Topp SD, Byrnes AE, Churchhouse C, Phatnani H, Smith BN, Rampersaud E, Wu G, Wuu J, Shatunov A, Iacoangeli A, Al Khleifat A, Mordes DA,

Ghosh S; ALSGENS Consortium; FALS Consortium; Project MinE Consortium; CreATe Consortium, Eggan K, Rademakers R, McCauley JL, Schüle R, Züchner S, Benatar M, Taylor JP, Nalls M, Gotkine M, Shaw PJ, Morrison KE, Al-Chalabi A, Traynor B, Shaw CE, Goldstein DB, Harms MB, Daly MJ, Neale BM. Exome sequencing in amyotrophic lateral sclerosis implicates a novel gene, DNAJC7, encoding a heat-shock protein. Nat Neurosci. 2019 Dec;22(12):1966-1974. Doi: 10.1038/s41593-019-0530-0. Epub 2019 Nov 25. Erratum in: Nat Neurosci. 2019 Dec 19;: PMID: 31768050; PMCID: PMC6919277.

Fecto F, Yan J, Vemula SP, Liu E, Yang Y, Chen W, Zheng JG, Shi Y, Siddique N, Arrat H, Donkervoort S, Ajroud-Driss S, Sufit RL, Heller SL, Deng HX, Siddique T. SQSTM1 mutations in familial and sporadic amyotrophic lateral sclerosis. Arch Neurol. 2011 Nov;68(11):1440-6. Doi: 10.1001/archneurol.2011.250. PMID: 22084127.

Fiddes IT, Lodewijk GA, Mooring M, Bosworth CM, Ewing AD, Mantalas GL, Novak AM, van den Bout A, Bishara A, Rosenkrantz JL, Lorig-Roach R, Field AR, Haeussler M, Russo L, Bhaduri A, Nowakowski TJ, Pollen AA, Dougherty ML, Nuttle X, Addor MC, Zwolinski S, Katzman S, Kriegstein A, Eichler EE, Salama SR, Jacobs FMJ, Haussler D. Human-Specific NOTCH2NL Genes Affect Notch Signaling and Cortical Neurogenesis. Cell. 2018 May 31;173(6):1356-1369.e22. doi: 10.1016/j.cell.2018.03.051. Epub 2018 May 31. PMID: 29856954; PMCID: PMC5986104.

Florian RT, Kraft F, Leitão E, Kaya S, Klebe S, Magnin E, van Rootselaar AF, Buratti J, Kühnel T, Schröder C, Giesselmann S, Tschernoster N, Altmueller J, Lamiral A, Keren B, Nava C, Bouteiller D, Forlani S, Jornea L, Kubica R, Ye T, Plassard D, Jost B, Meyer V, Deleuze JF, Delpu Y, Avarello MDM, Vijfhuizen LS, Rudolf G, Hirsch E, Kroes T, Reif PS, Rosenow F, Ganos C, Vidailhet M, Thivard L, Mathieu A, Bourgeron T, Kurth I, Rafehi H, Steenpass L, Horsthemke B; FAME consortium, LeGuern E, Klein KM, Labauge P, Bennett MF, Bahlo M, Gecz J, Corbett MA, Tijssen MAJ, van den Maagdenberg AMJM, Depienne C. Unstable TTTTA/TTTCA expansions in MARCH6 are associated with Familial Adult Myoclonic Epilepsy type 3. Nat Commun. 2019 Oct 29;10(1):4919. Doi: 10.1038/s41467-019-12763-9. PMID: 31664039; PMCID: PMC6820781.

Fornai F, Longone P, Cafaro L, Kastsiuchenka O, Ferrucci M, Manca ML, Lazzeri G, Spalloni A, Bellio N, Lenzi P, Modugno N, Siciliano G, Isidoro C, Murri L, Ruggieri S, Paparelli A. Lithium delays progression of amyotrophic lateral sclerosis. Proc Natl Acad Sci U S A. 2008 Feb 12;105(6):2052-7. Doi: 10.1073/pnas.0708022105. Epub 2008 Feb 4. Erratum in: Proc Natl Acad Sci U S A. 2008 Oct 21;105(42):16404-7. PMID: 18250315; PMCID: PMC2538879.

Freibaum BD, Lu Y, Lopez-Gonzalez R, Kim NC, Almeida S, Lee KH, Badders N, Valentine M, Miller BL, Wong PC, Petrucelli L, Kim HJ, Gao FB, Taylor JP. GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. Nature. 2015 Sep 3;525(7567):129-33. Doi: 10.1038/nature14974. Epub 2015 Aug 26. PMID: 26308899; PMCID: PMC4631399.

Freischmidt A, Wieland T, Richter B, Ruf W, Schaeffer V, Müller K, Marroquin N, Nordin F, Hübers A, Weydt P, Pinto S, Press R, Millecamps S, Molko N, Bernard E, Desnuelle C, Soriani MH, Dorst J, Graf E, Nordström U, Feiler MS, Putz S, Boeckers TM, Meyer T, Winkler AS, Winkelman J, de Carvalho M, Thal DR, Otto M, Brännström T, Volk AE, Kursula P, Danzer KM, Lichtner P, Dikic I, Meitinger T, Ludolph AC, Strom TM, Andersen PM, Weishaupt JH. Haploinsufficiency of TBK1 causes familial ALS and fronto-temporal dementia. Nat Neurosci. 2015 May;18(5):631-6. Doi: 10.1038/nn.4000. Epub 2015 Mar 24. PMID: 25803835.

Freyermuth F, Rau F, Kokunai Y, Linke T, Sellier C, Nakamori M, Kino Y, Arandel L, Jollet A, Thibault C, Philipps M, Vicaire S, Jost B, Udd B, Day JW, Duboc D, Wahbi K, Matsumura T, Fujimura H, Mochizuki H, Deryckere F, Kimura T, Nukina N, Ishiura S, Lacroix V, Campan-Fournier A, Navratil V, Chautard E, Auboeuf D, Horie M, Imoto K, Lee KY, Swanson MS, de Munain AL, Inada S, Itoh H, Nakazawa K, Ashihara T, Wang E, Zimmer T, Furling D, Takahashi MP, Charlet-Berguerand N. Splicing misregulation of SCN5A contributes to cardiac-conduction delay and heart arrhythmia in myotonic dystrophy. Nat Commun. 2016 Apr 11;7:11067. Doi: 10.1038/ncomms11067. PMID: 27063795; PMCID: PMC4831019.

Frick P, Sellier C, Mackenzie IRA, Cheng CY, Tahraoui-Bories J, Martinat C, Pasterkamp RJ, Prudlo J, Edbauer D, Oulad-Abdelghani M, Feederle R, Charlet-Berguerand N, Neumann M. Novel antibodies reveal presynaptic localization of C9orf72 protein and reduced protein levels in C9orf72

mutation carriers. Acta Neuropathol Commun. 2018 Aug 3;6(1):72. Doi: 10.1186/s40478-018-0579-0. PMID: 30075745; PMCID: PMC6091050.

Friedman-Gohas M, Elizur SE, Dratviman-Storobinsky O, Aizer A, Haas J, Raanani H, Orvieto R, Cohen Y. FMRpolyG accumulates in FMR1 premutation granulosa cells. J Ovarian Res. 2020 Feb 26;13(1):22. Doi: 10.1186/s13048-020-00623-w. PMID: 32101156; PMCID: PMC7045455.

Fu YH, Kuhl DP, Pizzuti A, Pieretti M, Sutcliffe JS, Richards S, Verkerk AJ, Holden JJ, Fenwick RG Jr, Warren ST, et al. Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. Cell. 1991 Dec 20;67(6):1047-58. Doi: 10.1016/0092-8674(91)90283-5. PMID: 1760838.

Fugier C, Klein AF, Hammer C, Vassilopoulos S, Ivarsson Y, Toussaint A, Tosch V, Vignaud A, Ferry A, Messaddeq N, Kokunai Y, Tsuburaya R, de la Grange P, Dembele D, Francois V, Precigout G, Boulade-Ladame C, Hummel MC, Lopez de Munain A, Sergeant N, Laquerrière A, Thibault C, Deryckere F, Auboeuf D, Garcia L, Zimmermann P, Udd B, Schoser B, Takahashi MP, Nishino I, Bassez G, Laporte J, Furling D, Charlet-Berguerand N. Misregulated alternative splicing of BIN1 is associated with T tubule alterations and muscle weakness in myotonic dystrophy. Nat Med. 2011 Jun;17(6):720-5. Doi: 10.1038/nm.2374. Epub 2011 May 29. PMID: 21623381.

Fukuda H, Yamaguchi D, Nyquist K, Yabuki Y, Miyatake S, Uchiyama Y, Hamanaka K, Saida K, Koshimizu E, Tsuchida N, Fujita A, Mitsuhashi S, Ohbo K, Satake Y, Sone J, Doi H, Morihara K, Okamoto T, Takahashi Y, Wenger AM, Shioda N, Tanaka F, Matsumoto N, Mizuguchi T. Father-to-offspring transmission of extremely long NOTCH2NLC repeat expansions with contractions: genetic and epigenetic profiling with long-read sequencing. Clin Epigenetics. 2021 Nov 13;13(1):204. Doi: 10.1186/s13148-021-01192-5. PMID: 34774111; PMCID: PMC8590777.

Fumagalli L, Young FL, Boeynaems S, De Decker M, Mehta AR, Swijsen A, Fazal R, Guo W, Moisse M, Beckers J, Dedeene L, Selvaraj BT, Vandoorne T, Madan V, van Blitterswijk M, Raitcheva D, McCampbell A, Poesen K, Gitler AD, Koch P, Berghe PV, Thal DR, Verfaillie C, Chandran S, Van Den Bosch L, Bullock SL, Van Damme P. *C9orf72*-derived arginine-containing dipeptide repeats associate with axonal transport machinery and impede microtubule-based motility. Sci Adv. 2021 Apr 9;7(15):eabg3013. Doi: 10.1126/sciadv.abg3013. PMID: 33837088; PMCID: PMC8034861.

Funkelstein L, Beinfeld M, Minokadeh A, Zadina J, Hook V. Unique biological function of cathepsin L in secretory vesicles for biosynthesis of neuropeptides. Neuropeptides. 2010 Dec;44(6):457-66. Doi: 10.1016/j.npep.2010.08.003. Epub 2010 Nov 2. PMID: 21047684; PMCID: PMC3058267.

G

Galluzzi L, Bravo-San Pedro JM, Levine B, Green DR, Kroemer G. Pharmacological modulation of autophagy: therapeutic potential and persisting obstacles. Nat Rev Drug Discov. 2017 Jul;16(7):487-511. doi: 10.1038/nrd.2017.22. Epub 2017 May 19. PMID: 28529316; PMCID: PMC5713640.Gao FB, Almeida S, Lopez-Gonzalez R. Dysregulated molecular pathways in amyotrophic lateral sclerosis-frontotemporal dementia spectrum disorder. EMBO J. 2017 Oct 16;36(20):2931-2950. doi: 10.15252/embj.201797568. Epub 2017 Sep 15. PMID: 28916614; PMCID: PMC5641681.

Gami P, Murray C, Schottlaender L, Bettencourt C, De Pablo Fernandez E, Mudanohwo E, Mizielinska S, Polke JM, Holton JL, Isaacs AM, Houlden H, Revesz T, Lashley T. A 30-unit hexanucleotide repeat expansion in C9orf72 induces pathological lesions with dipeptide-repeat proteins and RNA foci, but not TDP-43 inclusions and clinical disease. Acta Neuropathol. 2015 Oct;130(4):599-601. doi: 10.1007/s00401-015-1473-5. Epub 2015 Sep 7. PMID: 26347457.

Gantois I, Khoutorsky A, Popic J, Aguilar-Valles A, Freemantle E, Cao R, Sharma V, Pooters T, Nagpal A, Skalecka A, Truong VT, Wiebe S, Groves IA, Jafarnejad SM, Chapat C, McCullagh EA, Gamache K, Nader K, Lacaille JC, Gkogkas CG, Sonenberg N. Metformin ameliorates core deficits in a mouse model of fragile X syndrome. Nat Med. 2017 Jun;23(6):674-677. doi: 10.1038/nm.4335. Epub 2017 May 15. PMID: 28504725.

García-Murias M, Quintáns B, Arias M, Seixas AI, Cacheiro P, Tarrío R, Pardo J, Millán MJ, Arias-Rivas S, Blanco-Arias P, Dapena D, Moreira R, Rodríguez-Trelles F, Sequeiros J, Carracedo A, Silveira I, Sobrido MJ. 'Costa da Morte' ataxia is spinocerebellar ataxia 36: clinical and genetic

characterization. Brain. 2012 May;135(Pt 5):1423-35. doi: 10.1093/brain/aws069. Epub 2012 Apr 3. PMID: 22492559; PMCID: PMC3338928.

Gattey D, Zhu AY, Stagner A, Terry MA, Jun AS. Fuchs endothelial corneal dystrophy in patients with myotonic dystrophy: a case series. Cornea. 2014 Jan;33(1):96-8. doi: 10.1097/ICO.0000000000000018. PMID: 24270677; PMCID: PMC3898337.

Gelfman S, Dugger S, de Araujo Martins Moreno C, Ren Z, Wolock CJ, Shneider NA, Phatnani H, Cirulli ET, Lasseigne BN, Harris T, Maniatis T, Rouleau GA, Brown RH Jr, Gitler AD, Myers RM, Petrovski S, Allen A, Goldstein DB, Harms MB. A new approach for rare variation collapsing on functional protein domains implicates specific genic regions in ALS. Genome Res. 2019 May;29(5):809-818. doi: 10.1101/gr.243592.118. Epub 2019 Apr 2. PMID: 30940688; PMCID: PMC6499321.

Gelpi E, Botta-Orfila T, Bodi L, Marti S, Kovacs G, Grau-Rivera O, Lozano M, Sánchez-Valle R, Muñoz E, Valldeoriola F, Pagonabarraga J, Tartaglia GG, Milà M. Neuronal intranuclear (hyaline) inclusion disease and fragile X-associated tremor/ataxia syndrome: a morphological and molecular dilemma. Brain. 2017 Aug 1;140(8):e51. doi: 10.1093/brain/awx156. PMID: 28899011.

Gendron TF, Bieniek KF, Zhang YJ, Jansen-West K, Ash PE, Caulfield T, Daughrity L, Dunmore JH, Castanedes-Casey M, Chew J, Cosio DM, van Blitterswijk M, Lee WC, Rademakers R, Boylan KB, Dickson DW, Petrucelli L. Antisense transcripts of the expanded C9ORF72 hexanucleotide repeat form nuclear RNA foci and undergo repeat-associated non-ATG translation in c9FTD/ALS. Acta Neuropathol. 2013 Dec;126(6):829-44. doi: 10.1007/s00401-013-1192-8. Epub 2013 Oct 16. PMID: 24129584; PMCID: PMC3830741.

Gendron TF, Chew J, Stankowski JN, Hayes LR, Zhang YJ, Prudencio M, Carlomagno Y, Daughrity LM, Jansen-West K, Perkerson EA, O'Raw A, Cook C, Pregent L, Belzil V, van Blitterswijk M, Tabassian LJ, Lee CW, Yue M, Tong J, Song Y, Castanedes-Casey M, Rousseau L, Phillips V, Dickson DW, Rademakers R, Fryer JD, Rush BK, Pedraza O, Caputo AM, Desaro P, Palmucci C, Robertson A, Heckman MG, Diehl NN, Wiggs E, Tierney M, Braun L, Farren J, Lacomis D, Ladha S, Fournier CN, McCluskey LF, Elman LB, Toledo JB, McBride JD, Tiloca C, Morelli C, Poletti B, Solca F, Prelle A, Wuu J, Jockel-Balsarotti J, Rigo F, Ambrose C, Datta A, Yang W, Raitcheva D, Antognetti G, McCampbell A, Van Swieten JC, Miller BL, Boxer AL, Brown RH, Bowser R, Miller TM, Trojanowski JQ, Grossman M, Berry JD, Hu WT, Ratti A, Traynor BJ, Disney MD, Benatar M, Silani V, Glass JD, Floeter MK, Rothstein JD, Boylan KB, Petrucelli L. Poly(GP) proteins are a useful pharmacodynamic marker for *C9ORF72*-associated amyotrophic lateral sclerosis. Sci Transl Med. 2017 Mar 29;9(383):eaai7866. doi: 10.1126/scitranslmed.aai7866. PMID: 28356511; PMCID: PMC5576451.

Gijselinck I, Van Mossevelde S, van der Zee J, Sieben A, Engelborghs S, De Bleecker J, Ivanoiu A, Deryck O, Edbauer D, Zhang M, Heeman B, Bäumer V, Van den Broeck M, Mattheijssens M, Peeters K, Rogaeva E, De Jonghe P, Cras P, Martin JJ, de Deyn PP, Cruts M, Van Broeckhoven C. The C9orf72 repeat size correlates with onset age of disease, DNA methylation and transcriptional downregulation of the promoter. Mol Psychiatry. 2016 Aug;21(8):1112-24. doi: 10.1038/mp.2015.159. Epub 2015 Oct 20. PMID: 26481318; PMCID: PMC4960451.

Gitler AD, Tsuiji H. There has been an awakening: Emerging mechanisms of C9orf72 mutations in FTD/ALS. Brain Res. 2016 Sep 15;1647:19-29. doi: 10.1016/j.brainres.2016.04.004. Epub 2016 Apr 6. PMID: 27059391; PMCID: PMC5003651.

Glick D, Barth S, Macleod KF. Autophagy: cellular and molecular mechanisms. J Pathol. 2010 May;221(1):3-12. doi: 10.1002/path.2697. PMID: 20225336; PMCID: PMC2990190.

Gohel D, Sripada L, Prajapati P, Singh K, Roy M, Kotadia D, Tassone F, Charlet-Berguerand N, Singh R. FMRpolyG alters mitochondrial transcripts level and respiratory chain complex assembly in Fragile X associated tremor/ataxia syndrome [FXTAS]. Biochim Biophys Acta Mol Basis Dis. 2019 Jun 1;1865(6):1379-1388. doi: 10.1016/j.bbadis.2019.02.010. Epub 2019 Feb 13. PMID: 30771487.

Gohel D, Sripada L, Prajapati P, Currim F, Roy M, Singh K, Shinde A, Mane M, Kotadia D, Tassone F, Charlet-Berguerand N, Singh R. Expression of expanded FMR1-CGG repeats alters mitochondrial miRNAs and modulates mitochondrial functions and cell death in cellular model of FXTAS. Free Radic Biol Med. 2021 Mar;165:100-110. doi: 10.1016/j.freeradbiomed.2021.01.038. Epub 2021 Jan 23. PMID: 33497798.

Gomez-Deza J, Lee YB, Troakes C, Nolan M, Al-Sarraj S, Gallo JM, Shaw CE. Dipeptide repeat protein inclusions are rare in the spinal cord and almost absent from motor neurons in C9ORF72 mutant amyotrophic lateral sclerosis and are unlikely to cause their degeneration. Acta Neuropathol Commun. 2015 Jun 25;3:38. doi: 10.1186/s40478-015-0218-y. PMID: 26108573; PMCID: PMC4479315.

Greco CM, Hagerman RJ, Tassone F, Chudley AE, Del Bigio MR, Jacquemont S, Leehey M, Hagerman PJ. Neuronal intranuclear inclusions in a new cerebellar tremor/ataxia syndrome among fragile X carriers. Brain. 2002 Aug;125(Pt 8):1760-71. doi: 10.1093/brain/awf184. PMID: 12135967.

Greco CM, Berman RF, Martin RM, Tassone F, Schwartz PH, Chang A, Trapp BD, Iwahashi C, Brunberg J, Grigsby J, Hessl D, Becker EJ, Papazian J, Leehey MA, Hagerman RJ, Hagerman PJ. Neuropathology of fragile X-associated tremor/ataxia syndrome (FXTAS). Brain. 2006 Jan;129(Pt 1):243-55. doi: 10.1093/brain/awh683. Epub 2005 Dec 5. PMID: 16332642.

Green KM, Glineburg MR, Kearse MG, Flores BN, Linsalata AE, Fedak SJ, Goldstrohm AC, Barmada SJ, Todd PK. RAN translation at C9orf72-associated repeat expansions is selectively enhanced by the integrated stress response. Nat Commun. 2017 Dec 8;8(1):2005. doi: 10.1038/s41467-017-02200-0. PMID: 29222490; PMCID: PMC5722904.

Gros-Louis F, Andersen PM, Dupre N, Urushitani M, Dion P, Souchon F, D'Amour M, Camu W, Meininger V, Bouchard JP, Rouleau GA, Julien JP. Chromogranin B P413L variant as risk factor and modifier of disease onset for amyotrophic lateral sclerosis. Proc Natl Acad Sci U S A. 2009 Dec 22;106(51):21777-82. doi: 10.1073/pnas.0902174106. Epub 2009 Dec 9. PMID: 20007371; PMCID: PMC2799865.

Grote A, Robens BK, Blümcke I, Becker AJ, Schoch S, Gembé E. LRP12 silencing during brain development results in cortical dyslamination and seizure sensitization. Neurobiol Dis. 2016 Feb;86:170-6. doi: 10.1016/j.nbd.2015.11.021. Epub 2015 Dec 2. PMID: 26639854.

Gu X, Greiner ER, Mishra R, Kodali R, Osmand A, Finkbeiner S, Steffan JS, Thompson LM, Wetzel R, Yang XW. Serines 13 and 16 are critical determinants of full-length human mutant huntingtin induced disease pathogenesis in HD mice. Neuron. 2009 Dec 24;64(6):828-40. doi: 10.1016/j.neuron.2009.11.020. PMID: 20064390; PMCID: PMC2807408.

Gu X, Cantle JP, Greiner ER, Lee CY, Barth AM, Gao F, Park CS, Zhang Z, Sandoval-Miller S, Zhang RL, Diamond M, Mody I, Coppola G, Yang XW. N17 Modifies mutant Huntingtin nuclear pathogenesis and severity of disease in HD BAC transgenic mice. Neuron. 2015 Feb 18;85(4):726-41. doi: 10.1016/j.neuron.2015.01.008. Epub 2015 Feb 5. PMID: 25661181; PMCID: PMC4386927.

Guo L, Elcioglu NH, Iida A, Demirkol YK, Aras S, Matsumoto N, Nishimura G, Miyake N, Ikegawa S. Novel and recurrent XYLT1 mutations in two Turkish families with Desbuquois dysplasia, type 2. J Hum Genet. 2017 Mar;62(3):447-451. doi: 10.1038/jhg.2016.143. Epub 2016 Nov 24. PMID: 27881841.

Guo Q, Lehmer C, Martínez-Sánchez A, Rudack T, Beck F, Hartmann H, Pérez-Berlanga M, Frottin F, Hipp MS, Hartl FU, Edbauer D, Baumeister W, Fernández-Busnadiego R. In Situ Structure of Neuronal C9orf72 Poly-GA Aggregates Reveals Proteasome Recruitment. Cell. 2018 Feb 8;172(4):696-705.e12. doi: 10.1016/j.cell.2017.12.030. Epub 2018 Feb 1. PMID: 29398115; PMCID: PMC6035389.

H

Haeusler AR, Donnelly CJ, Periz G, Simko EA, Shaw PG, Kim MS, Maragakis NJ, Troncoso JC, Pandey A, Sattler R, Rothstein JD, Wang J. C9orf72 nucleotide repeat structures initiate molecular cascades of disease. Nature. 2014 Mar 13;507(7491):195-200. Doi: 10.1038/nature13124. Epub 2014 Mar 5. PMID: 24598541; PMCID: PMC4046618.

Hagerman RJ, Leehey M, Heinrichs W, Tassone F, Wilson R, Hills J, Grigsby J, Gage B, Hagerman PJ. Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. Neurology. 2001 Jul 10;57(1):127-30. Doi: 10.1212/wnl.57.1.127. PMID: 11445641.

Haify SN, Buijsen RAM, Verwegen L, Severijnen LWFM, de Boer H, Boumeester V, Monshouwer R, Yang WY, Cameron MD, Willemsen R, Disney MD, Hukema RK. Small molecule 1a reduces

FMRpolyG-mediated toxicity in in vitro and in vivo models for FMR1 premutation. Hum Mol Genet. 2021 Aug 12;30(17):1632-1648. Doi: 10.1093/hmg/ddab143. PMID: 34077515; PMCID: PMC8369842.

Hanby MF, Scott KM, Scotton W, Wijesekera L, Mole T, Ellis CE, Leigh PN, Shaw CE, Al-Chalabi A. The risk to relatives of patients with sporadic amyotrophic lateral sclerosis. Brain. 2011 Dec;134(Pt 12):3454-7. Doi: 10.1093/brain/awr248. Epub 2011 Sep 20. PMID: 21933809; PMCID: PMC3235555.

Hansen M, Rubinsztein DC, Walker DW. Autophagy as a promoter of longevity: insights from model organisms. Nat Rev Mol Cell Biol. 2018 Sep;19(9):579-593. Doi: 10.1038/s41580-018-0033-y. Erratum in: Nat Rev Mol Cell Biol. 2018 Jul 25;: PMID: 30006559; PMCID: PMC6424591.

Hao Z, Liu L, Tao Z, Wang R, Ren H, Sun H, Lin Z, Zhang Z, Mu C, Zhou J, Wang G. Motor dysfunction and neurodegeneration in a C9orf72 mouse line expressing poly-PR. Nat Commun. 2019 Jul 2;10(1):2906. Doi: 10.1038/s41467-019-10956-w. PMID: 31266945; PMCID: PMC6606620.

Hasskarl J. Everolimus. Recent Results Cancer Res. 2018;211:101-123. Doi: 10.1007/978-3-319-91442-8_8. PMID: 30069763.

Hautbergue GM, Castelli LM, Ferraiuolo L, Sanchez-Martinez A, Cooper-Knock J, Higginbottom A, Lin YH, Bauer CS, Dodd JE, Myszczynska MA, Alam SM, Garneret P, Chandran JS, Karyka E, Stopford MJ, Smith EF, Kirby J, Meyer K, Kaspar BK, Isaacs AM, El-Khamisy SF, De Vos KJ, Ning K, Azzouz M, Whitworth AJ, Shaw PJ. SRSF1-dependent nuclear export inhibition of C9ORF72 repeat transcripts prevents neurodegeneration and associated motor deficits. Nat Commun. 2017 Jul 5;8:16063. Doi: 10.1038/ncomms16063. PMID: 28677678; PMCID: PMC5504286.

Hayes LR, Duan L, Bowen K, Kalab P, Rothstein JD. *C9orf72* arginine-rich dipeptide repeat proteins disrupt karyopherin-mediated nuclear import. Elife. 2020 Mar 2;9:e51685. Doi: 10.7554/eLife.51685. PMID: 32119645; PMCID: PMC7051184.

Hazari Y, Bravo-San Pedro JM, Hetz C, Galluzzi L, Kroemer G. Autophagy in hepatic adaptation to stress. J Hepatol. 2020 Jan;72(1):183-196. Doi: 10.1016/j.jhep.2019.08.026. PMID: 31849347.

Herranz-Martin S, Chandran J, Lewis K, Mulcahy P, Higginbottom A, Walker C, Valenzuela IMY, Jones RA, Coldicott I, Iannitti T, Akaaboune M, El-Khamisy SF, Gillingwater TH, Shaw PJ, Azzouz M. Viral delivery of *C9orf72* hexanucleotide repeat expansions in mice leads to repeat-length-dependent neuropathology and behavioural deficits. Dis Model Mech. 2017 Jul 1;10(7):859-868. Doi: 10.1242/dmm.029892. Epub 2017 May 26. PMID: 28550099; PMCID: PMC5536911.

Hirst M, Grewal P, Flannery A, Slatter R, Maher E, Barton D, Fryns JP, Davies K. Two new cases of FMR1 deletion associated with mental impairment. Am J Hum Genet. 1995 Jan;56(1):67-74. PMID: 7825604; PMCID: PMC1801332.

Ho TH, Charlet-B N, Poulos MG, Singh G, Swanson MS, Cooper TA. Muscleblind proteins regulate alternative splicing. EMBO J. 2004 Aug 4;23(15):3103-12. Doi: 10.1038/sj.emboj.7600300. Epub 2004 Jul 15. PMID: 15257297; PMCID: PMC514918.

Hoffman NJ, Parker BL, Chaudhuri R, Fisher-Wellman KH, Kleinert M, Humphrey SJ, Yang P, Holliday M, Trefely S, Fazakerley DJ, Stöckli J, Burchfield JG, Jensen TE, Jothi R, Kiens B, Wojtaszewski JF, Richter EA, James DE. Global Phosphoproteomic Analysis of Human Skeletal Muscle Reveals a Network of Exercise-Regulated Kinases and AMPK Substrates. Cell Metab. 2015 Nov 3;22(5):922-35. Doi: 10.1016/j.cmet.2015.09.001. Epub 2015 Oct 1. Erratum in: Cell Metab. 2015 Nov 3;22(5):948. PMID: 26437602; PMCID: PMC4635038.

Holmberg M, Duyckaerts C, Dürr A, Cancel G, Gourfinkel-An I, Damier P, Faucheux B, Trottier Y, Hirsch EC, Agid Y, Brice A. Spinocerebellar ataxia type 7 (SCA7): a neurodegenerative disorder with neuronal intranuclear inclusions. Hum Mol Genet. 1998 May;7(5):913-8. Doi: 10.1093/hmg/7.5.913. PMID: 9536097.

Holmes SE, O'Hearn EE, McInnis MG, Gorelick-Feldman DA, Kleiderlein JJ, Callahan C, Kwak NG, Ingersoll-Ashworth RG, Sherr M, Sumner AJ, Sharp AH, Ananth U, Seltzer WK, Boss MA, Vieria-Saecker AM, Epplen JT, Riess O, Ross CA, Margolis RL. Expansion of a novel CAG trinucleotide repeat in the 5' region of PPP2R2B is associated with SCA12. Nat Genet. 1999 Dec;23(4):391-2. Doi: 10.1038/70493. PMID: 10581021.

Holmes SE, O'Hearn E, Rosenblatt A, Callahan C, Hwang HS, Ingersoll-Ashworth RG, Fleisher A, Stevanin G, Brice A, Potter NT, Ross CA, Margolis RL. A repeat expansion in the gene encoding junctophilin-3 is associated with Huntington disease-like 2. Nat Genet. 2001 Dec;29(4):377-8. Doi: 10.1038/ng760. Erratum in: Nat Genet 2002 Jan;30(1):123. PMID: 11694876.

Holmes C, Boche D, Wilkinson D, Yadegarfar G, Hopkins V, Bayer A, Jones RW, Bullock R, Love S, Neal JW, Zotova E, Nicoll JA. Long-term effects of Abeta42 immunisation in Alzheimer's disease: follow-up of a 170andomized, placebo-controlled phase I trial. Lancet. 2008 Jul 19;372(9634):216-23. Doi: 10.1016/S0140-6736(08)61075-2. PMID: 18640458.

Hua Y, Vickers TA, Okunola HL, Bennett CF, Krainer AR. Antisense masking of an hnRNP A1/A2 intronic splicing silencer corrects SMN2 splicing in transgenic mice. Am J Hum Genet. 2008 Apr;82(4):834-48. Doi: 10.1016/j.ajhg.2008.01.014. Epub 2008 Mar 27. PMID: 18371932; PMCID: PMC2427210.

Huang HK, Yoon H, Hannig EM, Donahue TF. GTP hydrolysis controls stringent selection of the AUG start codon during translation initiation in Saccharomyces cerevisiae. Genes Dev. 1997 Sep 15;11(18):2396-413. Doi: 10.1101/gad.11.18.2396. PMID: 9308967; PMCID: PMC316512.

Hukema RK, Buijsen RA, Raske C, Severijnen LA, Nieuwenhuizen-Bakker I, Minneboo M, Maas A, de Crom R, Kros JM, Hagerman PJ, Berman RF, Willemsen R. Induced expression of expanded CGG RNA causes mitochondrial dysfunction in vivo. Cell Cycle. 2014;13(16):2600-8. Doi: 10.4161/15384101.2014.943112. PMID: 25486200; PMCID: PMC4614669.

Hunsaker MR, Greco CM, Spath MA, Smits AP, Navarro CS, Tassone F, Kros JM, Severijnen LA, Berry-Kravis EM, Berman RF, Hagerman PJ, Willemsen R, Hagerman RJ, Hukema RK. Widespread non-central nervous system organ pathology in fragile X premutation carriers with fragile X-associated tremor/ataxia syndrome and CGG knock-in mice. Acta Neuropathol. 2011 Oct;122(4):467-79. Doi: 10.1007/s00401-011-0860-9. Epub 2011 Jul 23. PMID: 21785977; PMCID: PMC3222079.

Hutton M, Lendon CL, Rizzu P, Baker M, Froelich S, Houlden H, Pickering-Brown S, Chakraverty S, Isaacs A, Grover A, Hackett J, Adamson J, Lincoln S, Dickson D, Davies P, Petersen RC, Stevens M, de Graaff E, Wauters E, van Baren J, Hillebrand M, Joosse M, Kwon JM, Nowotny P, Che LK, Norton J, Morris JC, Reed LA, Trojanowski J, Basun H, Lannfelt L, Neystat M, Fahn S, Dark F, Tannenberg T, Dodd PR, Hayward N, Kwok JB, Schofield PR, Andreadis A, Snowden J, Craufurd D, Neary D, Owen F, Oostra BA, Hardy J, Goate A, van Swieten J, Mann D, Lynch T, Heutink P. Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17. Nature. 1998 Jun 18;393(6686):702-5. Doi: 10.1038/31508. PMID: 9641683.

Huynh DP, Figueroa K, Hoang N, Pulst SM. Nuclear localization or inclusion body formation of ataxin-2 are not necessary for SCA2 pathogenesis in mouse or human. Nat Genet. 2000 Sep;26(1):44-50. Doi: 10.1038/79162. PMID: 10973246.

Hsu PP, Kang SA, Rameseder J, Zhang Y, Ottina KA, Lim D, Peterson TR, Choi Y, Gray NS, Yaffe MB, Marto JA, Sabatini DM. The mTOR-regulated phosphoproteome reveals a mechanism of mTORC1-mediated inhibition of growth factor signaling. Science. 2011 Jun 10;332(6035):1317-22. Doi: 10.1126/science.1199498. PMID: 21659604; PMCID: PMC3177140.

I

Ikeda H, Yamaguchi M, Sugai S, Aze Y, Narumiya S, Kakizuka A. Expanded polyglutamine in the Machado-Joseph disease protein induces cell death in vitro and in vivo. Nat Genet. 1996 Jun;13(2):196-202. doi: 10.1038/ng0696-196. PMID: 8640226.

Ikeda Y, Shizuka M, Watanabe M, Okamoto K, Shoji M. Molecular and clinical analyses of spinocerebellar ataxia type 8 in Japan. Neurology. 2000 Feb 22;54(4):950-5. doi: 10.1212/wnl.54.4.950. PMID: 10690991.

Imamura K, Izumi Y, Watanabe A, Tsukita K, Woltjen K, Yamamoto T, Hotta A, Kondo T, Kitaoka S, Ohta A, Tanaka A, Watanabe D, Morita M, Takuma H, Tamaoka A, Kunath T, Wray S, Furuya H, Era T, Makioka K, Okamoto K, Fujisawa T, Nishitoh H, Homma K, Ichijo H, Julien JP, Obata N, Hosokawa M, Akiyama H, Kaneko S, Ayaki T, Ito H, Kaji R, Takahashi R, Yamanaka S, Inoue H. The Src/c-Abl pathway is a potential therapeutic target in amyotrophic lateral sclerosis. Sci Transl Med. 2017 May 24;9(391):eaaf3962. doi: 10.1126/scitranslmed.aaf3962. PMID: 28539470.

Ingolia NT, Lareau LF, Weissman JS. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell. 2011 Nov 11;147(4):789-802. doi: 10.1016/j.cell.2011.10.002. Epub 2011 Nov 3. PMID: 22056041; PMCID: PMC3225288.

Ishiguro T, Sato N, Ueyama M, Fujikake N, Sellier C, Kanegami A, Tokuda E, Zamiri B, Gall-Duncan T, Mirceta M, Furukawa Y, Yokota T, Wada K, Taylor JP, Pearson CE, Charlet-Berguerand N, Mizusawa H, Nagai Y, Ishikawa K. Regulatory Role of RNA Chaperone TDP-43 for RNA Misfolding and Repeat-Associated Translation in SCA31. Neuron. 2017 Apr 5;94(1):108-124.e7. doi: 10.1016/j.neuron.2017.02.046. Epub 2017 Mar 23. PMID: 28343865; PMCID: PMC5681996.

Ishiguro T, Nagai Y, Ishikawa K. Insight Into Spinocerebellar Ataxia Type 31 (SCA31) From *Drosophila* Model. Front Neurosci. 2021 May 25;15:648133. doi: 10.3389/fnins.2021.648133. PMID: 34113230; PMCID: PMC8185138.

Ishikawa K, Fujigasaki H, Saegusa H, Ohwada K, Fujita T, Iwamoto H, Komatsuzaki Y, Toru S, Toriyama H, Watanabe M, Ohkoshi N, Shoji S, Kanazawa I, Tanabe T, Mizusawa H. Abundant expression and cytoplasmic aggregations of [alpha]1A voltage-dependent calcium channel protein associated with neurodegeneration in spinocerebellar ataxia type 6. Hum Mol Genet. 1999 Jul;8(7):1185-93. doi: 10.1093/hmg/8.7.1185. PMID: 10369863.

Ishiura H, Doi K, Mitsui J, Yoshimura J, Matsukawa MK, Fujiyama A, Toyoshima Y, Kakita A, Takahashi H, Suzuki Y, Sugano S, Qu W, Ichikawa K, Yurino H, Higasa K, Shibata S, Mitsue A, Tanaka M, Ichikawa Y, Takahashi Y, Date H, Matsukawa T, Kanda J, Nakamoto FK, Higashihara M, Abe K, Koike R, Sasagawa M, Kuroha Y, Hasegawa N, Kanesawa N, Kondo T, Hitomi T, Tada M, Takano H, Saito Y, Sanpei K, Onodera O, Nishizawa M, Nakamura M, Yasuda T, Sakiyama Y, Otsuka M, Ueki A, Kaida KI, Shimizu J, Hanajima R, Hayashi T, Terao Y, Inomata-Terada S, Hamada M, Shirota Y, Kubota A, Ugawa Y, Koh K, Takiyama Y, Ohsawa-Yoshida N, Ishiura S, Yamasaki R, Tamaoka A, Akiyama H, Otsuki T, Sano A, Ikeda A, Goto J, Morishita S, Tsuji S. Expansions of intronic TTTCA and TTTTA repeats in benign adult familial myoclonic epilepsy. Nat Genet. 2018 Apr;50(4):581-590. doi: 10.1038/s41588-018-0067-2. Epub 2018 Mar 5. PMID: 29507423.

Ishiura H, Shibata S, Yoshimura J, Suzuki Y, Qu W, Doi K, Almansour MA, Kikuchi JK, Taira M, Mitsui J, Takahashi Y, Ichikawa Y, Mano T, Iwata A, Harigaya Y, Matsukawa MK, Matsukawa T, Tanaka M, Shirota Y, Ohtomo R, Kowa H, Date H, Mitsue A, Hatsuta H, Morimoto S, Murayama S, Shiio Y, Saito Y, Mitsutake A, Kawai M, Sasaki T, Sugiyama Y, Hamada M, Ohtomo G, Terao Y, Nakazato Y, Takeda A, Sakiyama Y, Umeda-Kameyama Y, Shinmi J, Ogata K, Kohno Y, Lim SY, Tan AH, Shimizu J, Goto J, Nishino I, Toda T, Morishita S, Tsuji S. Noncoding CGG repeat expansions in neuronal intranuclear inclusion disease, oculopharyngodistal myopathy and an overlapping disease. Nat Genet. 2019 Aug;51(8):1222-1232. doi: 10.1038/s41588-019-0458-z. Epub 2019 Jul 22. PMID: 31332380.

Ivanov IP, Loughran G, Sachs MS, Atkins JF. Initiation context modulates autoregulation of eukaryotic translation initiation factor 1 (eIF1). Proc Natl Acad Sci U S A. 2010 Oct 19;107(42):18056-60. doi: 10.1073/pnas.1009269107. Epub 2010 Oct 4. PMID: 20921384; PMCID: PMC2964218.

Iwahashi CK, Yasui DH, An HJ, Greco CM, Tassone F, Nannen K, Babineau B, Lebrilla CB, Hagerman RJ, Hagerman PJ. Protein composition of the intranuclear inclusions of FXTAS. Brain. 2006 Jan;129(Pt 1):256-71. doi: 10.1093/brain/awh650. Epub 2005 Oct 24. PMID: 16246864.

J

Jackson RJ, Hellen CU, Pestova TV. The mechanism of eukaryotic translation initiation and principles of its regulation. Nat Rev Mol Cell Biol. 2010 Feb;11(2):113-27. Doi: 10.1038/nrm2838. PMID: 20094052; PMCID: PMC4461372.

Jacquemont S, Hagerman RJ, Leehey M, Grigsby J, Zhang L, Brunberg JA, Greco C, Des Portes V, Jardini T, Levine R, Berry-Kravis E, Brown WT, Schaeffer S, Kissel J, Tassone F, Hagerman PJ. Fragile X premutation tremor/ataxia syndrome: molecular, clinical, and neuroimaging correlates. Am J Hum Genet. 2003 Apr;72(4):869-78. Doi: 10.1086/374321. Epub 2003 Mar 12. PMID: 12638084; PMCID: PMC1180350.

Jamsheer A, Olech EM, Kozłowski K, Niedziela M, Sowińska-Seidler A, Obara-Moszyńska M, Latos-Bieleńska A, Karczewski M, Zemojtel T. Exome sequencing reveals two novel compound

heterozygous XYLT1 mutations in a Polish patient with Desbuquois dysplasia type 2 and growth hormone deficiency. J Hum Genet. 2016 Jul;61(7):577-83. Doi: 10.1038/jhg.2016.30. Epub 2016 Mar 31. PMID: 27030147.

Jensen BK, Schuldi MH, McAvoy K, Russell KA, Boehringer A, Curran BM, Krishnamurthy K, Wen X, Westergard T, Ma L, Haeusler AR, Edbauer D, Pasinelli P, Trotti D. Synaptic dysfunction induced by glycine-alanine dipeptides in C9orf72-ALS/FTD is rescued by SV2 replenishment. EMBO Mol Med. 2020 May 8;12(5):e10722. Doi: 10.15252/emmm.201910722. Epub 2020 Apr 29. PMID: 32347002; PMCID: PMC7207170.

Jiang J, Zhu Q, Gendron TF, Saberi S, McAlonis-Downes M, Seelman A, Stauffer JE, Jafar-Nejad P, Drenner K, Schulte D, Chun S, Sun S, Ling SC, Myers B, Engelhardt J, Katz M, Baughn M, Platoshyn O, Marsala M, Watt A, Heyser CJ, Ard MC, De Muynck L, Daughrity LM, Swing DA, Tessarollo L, Jung CJ, Delpoux A, Utzschneider DT, Hedrick SM, de Jong PJ, Edbauer D, Van Damme P, Petrucelli L, Shaw CE, Bennett CF, Da Cruz S, Ravits J, Rigo F, Cleveland DW, Lagier-Tourenne C. Gain of Toxicity from ALS/FTD-Linked Repeat Expansions in C9ORF72 Is Alleviated by Antisense Oligonucleotides Targeting GGGGCC-Containing RNAs. Neuron. 2016 May 4;90(3):535-50. Doi: 10.1016/j.neuron.2016.04.006. Epub 2016 Apr 21. PMID: 27112497; PMCID: PMC4860075.

Jin P, Duan R, Qurashi A, Qin Y, Tian D, Rosser TC, Liu H, Feng Y, Warren ST. Pur alpha binds to rCGG repeats and modulates repeat-mediated neurodegeneration in a Drosophila model of fragile X tremor/ataxia syndrome. Neuron. 2007 Aug 16;55(4):556-64. Doi: 10.1016/j.neuron.2007.07.020. PMID: 17698009; PMCID: PMC1994817.

Jung J, Nayak A, Schaeffer V, Starzetz T, Kirsch AK, Müller S, Dikic I, Mittelbronn M, Behrends C. Multiplex image-based autophagy RNAi screening identifies SMCR8 as ULK1 kinase activity and gene expression regulator. Elife. 2017 Feb 14;6:e23063. Doi: 10.7554/eLife.23063. PMID: 28195531; PMCID: PMC5323046.

K

Kato S, Horiuchi S, Nakashima K, Hirano A, Shibata N, Nakano I, Saito M, Kato M, Asayama K, Ohama E. Astrocytic hyaline inclusions contain advanced glycation endproducts in familial amyotrophic lateral sclerosis with superoxide dismutase 1 gene mutation: immunohistochemical and immunoelectron microscopical analyses. Acta Neuropathol. 1999 Mar;97(3):260-6. doi: 10.1007/s004010050983. PMID: 10090673.

Katsuno M, Adachi H, Kume A, Li M, Nakagomi Y, Niwa H, Sang C, Kobayashi Y, Doyu M, Sobue G. Testosterone reduction prevents phenotypic expression in a transgenic mouse model of spinal and bulbar muscular atrophy. Neuron. 2002 Aug 29;35(5):843-54. doi: 10.1016/s0896-6273(02)00834-6. PMID: 12372280.

Kearse MG, Green KM, Krans A, Rodriguez CM, Linsalata AE, Goldstrohm AC, Todd PK. CGG Repeat-Associated Non-AUG Translation Utilizes a Cap-Dependent Scanning Mechanism of Initiation to Produce Toxic Proteins. Mol Cell. 2016 Apr 21;62(2):314-322. doi: 10.1016/j.molcel.2016.02.034. Epub 2016 Mar 31. PMID: 27041225; PMCID: PMC4854189.

Kearse MG, Wilusz JE. Non-AUG translation: a new start for protein synthesis in eukaryotes. Genes Dev. 2017 Sep 1;31(17):1717-1731. doi: 10.1101/gad.305250.117. PMID: 28982758; PMCID: PMC5666671.

Khosravi B, Hartmann H, May S, Möhl C, Ederle H, Michaelsen M, Schludi MH, Dormann D, Edbauer D. Cytoplasmic poly-GA aggregates impair nuclear import of TDP-43 in C9orf72 ALS/FTLD. Hum Mol Genet. 2017 Feb 15;26(4):790-800. doi: 10.1093/hmg/ddw432. PMID: 28040728; PMCID: PMC5409121.

Khosravi B, LaClair KD, Riemenschneider H, Zhou Q, Frottin F, Mareljic N, Czuppa M, Farny D, Hartmann H, Michaelsen M, Arzberger T, Hartl FU, Hipp MS, Edbauer D. Cell-to-cell transmission of C9orf72 poly-(Gly-Ala) triggers key features of ALS/FTD. EMBO J. 2020 Apr 15;39(8):e102811. doi: 10.15252/embj.2019102811. Epub 2020 Mar 16. PMID: 32175624; PMCID: PMC7156967.

Kiliszek A, Kierzek R, Krzyzosiak WJ, Rypniewski W. Atomic resolution structure of CAG RNA repeats: structural insights and implications for the trinucleotide repeat expansion diseases. Nucleic

Acids Res. 2010 Dec;38(22):8370-6. doi: 10.1093/nar/gkq700. Epub 2010 Aug 11. PMID: 20702420; PMCID: PMC3001072.
Kim M, Bellini M, Ceman S. Fragile X mental retardation protein FMRP binds mRNAs in the nucleus. Mol Cell Biol. 2009 Jan;29(1):214-28. doi: 10.1128/MCB.01377-08. Epub 2008 Oct 20. PMID: 18936162; PMCID: PMC2612477.
Kim KH, Son JM, Benayoun BA, Lee C. The Mitochondrial-Encoded Peptide MOTS-c Translocates to the Nucleus to Regulate Nuclear Gene Expression in Response to Metabolic Stress. Cell Metab. 2018 Sep 4;28(3):516-524.e7. doi: 10.1016/j.cmet.2018.06.008. Epub 2018 Jul 5. PMID: 29983246; PMCID: PMC6185997.
Kimber TE, Blumbergs PC, Rice JP, Hallpike JF, Edis R, Thompson PD, Suthers G. Familial neuronal intranuclear inclusion disease with ubiquitin positive inclusions. J Neurol Sci. 1998 Sep 18;160(1):33-40. doi: 10.1016/s0022-510x(98)00169-5. PMID: 9804114.
Kingwell K. Double setback for ASO trials in Huntington disease. Nat Rev Drug Discov. 2021 Jun;20(6):412-413. doi: 10.1038/d41573-021-00088-6. PMID: 34012000.
Klement IA, Skinner PJ, Kaytor MD, Yi H, Hersch SM, Clark HB, Zoghbi HY, Orr HT. Ataxin-1 nuclear localization and aggregation: role in polyglutamine-induced disease in SCA1 transgenic mice. Cell. 1998 Oct 2;95(1):41-53. doi: 10.1016/s0092-8674(00)81781-x. PMID: 9778246.
Knight SJ, Ritchie RJ, Chakrabarti L, Cross G, Taylor GR, Mueller RF, Hurst J, Paterson J, Yates JR, Dow DJ, Davies KE. A study of FRAXE in mentally retarded individuals referred for fragile X syndrome (FRAXA) testing in the United Kingdom. Am J Hum Genet. 1996 May;58(5):906-13. PMID: 8651274; PMCID: PMC1914619.
Kobayashi H, Abe K, Matsuura T, Ikeda Y, Hitomi T, Akechi Y, Habu T, Liu W, Okuda H, Koizumi A. Expansion of intronic GGCCTG hexanucleotide repeat in NOP56 causes SCA36, a type of spinocerebellar ataxia accompanied by motor neuron involvement. Am J Hum Genet. 2011 Jul 15;89(1):121-30. doi: 10.1016/j.ajhg.2011.05.015. Epub 2011 Jun 16. PMID: 21683323; PMCID: PMC3135815.
Komar AA, Hatzoglou M. Internal ribosome entry sites in cellular mRNAs: mystery of their existence. J Biol Chem. 2005 Jun 24;280(25):23425-8. doi: 10.1074/jbc.R400041200. Epub 2005 Mar 4. PMID: 15749702.
Kotani T, Kirisako H, Koizumi M, Ohsumi Y, Nakatogawa H. The Atg2-Atg18 complex tethers pre-autophagosomal membranes to the endoplasmic reticulum for autophagosome formation. Proc Natl Acad Sci U S A. 2018 Oct 9;115(41):10363-10368. doi: 10.1073/pnas.1806727115. Epub 2018 Sep 25. PMID: 30254161; PMCID: PMC6187169.
Kozak M. Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems. Mol Cell Biol. 1989 Nov;9(11):5073-80. doi: 10.1128/mcb.9.11.5073-5080.1989. PMID: 2601709; PMCID: PMC363659.
Krans A, Kearse MG, Todd PK. Repeat-associated non-AUG translation from antisense CCG repeats in fragile X tremor/ataxia syndrome. Ann Neurol. 2016 Dec;80(6):871-881. doi: 10.1002/ana.24800. Epub 2016 Nov 26. PMID: 27761921; PMCID: PMC5177492.
Krans A, Skariah G, Zhang Y, Bayly B, Todd PK. Neuropathology of RAN translation proteins in fragile X-associated tremor/ataxia syndrome. Acta Neuropathol Commun. 2019 Oct 30;7(1):152. doi: 10.1186/s40478-019-0782-7. PMID: 31665086; PMCID: PMC6821001.
Kuliawat R, Klumperman J, Ludwig T, Arvan P. Differential sorting of lysosomal enzymes out of the regulated secretory pathway in pancreatic beta-cells. J Cell Biol. 1997 May 5;137(3):595-608. doi: 10.1083/jcb.137.3.595. PMID: 9151667; PMCID: PMC2139876.
Kwiatkowski TJ Jr, Bosco DA, Leclerc AL, Tamrazian E, Vanderburg CR, Russ C, Davis A, Gilchrist J, Kasarskis EJ, Munsat T, Valdmanis P, Rouleau GA, Hosler BA, Cortelli P, de Jong PJ, Yoshinaga Y, Haines JL, Pericak-Vance MA, Yan J, Ticozzi N, Siddique T, McKenna-Yasek D, Sapp PC, Horvitz HR, Landers JE, Brown RH Jr. Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. Science. 2009 Feb 27;323(5918):1205-8. doi: 10.1126/science.1166066. PMID: 19251627.
Kwon I, Xiang S, Kato M, Wu L, Theodoropoulos P, Wang T, Kim J, Yun J, Xie Y, McKnight SL. Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. Science. 2014 Sep 5;345(6201):1139-45. doi: 10.1126/science.1254917. Epub 2014 Jul 31. PMID: 25081482; PMCID: PMC4459787.

Kwon D. Failure of genetic therapies for Huntington's devastates community. Nature. 2021 May;593(7858):180. doi: 10.1038/d41586-021-01177-7. PMID: 33963316.

L

La Spada AR, Wilson EM, Lubahn DB, Harding AE, Fischbeck KH. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. Nature. 1991 Jul 4;352(6330):77-9. doi: 10.1038/352077a0. PMID: 2062380.

LaClair KD, Zhou Q, Michaelsen M, Wefers B, Brill MS, Janjic A, Rathkolb B, Farny D, Cygan M, de Angelis MH, Wurst W, Neumann M, Enard W, Misgeld T, Arzberger T, Edbauer D. Congenic expression of poly-GA but not poly-PR in mice triggers selective neuron loss and interferon responses found in C9orf72 ALS. Acta Neuropathol. 2020 Aug;140(2):121-142. doi: 10.1007/s00401-020-02176-0. Epub 2020 Jun 19. PMID: 32562018; PMCID: PMC7360660.

LaCroix AJ, Stabley D, Sahraoui R, Adam MP, Mehaffey M, Kernan K, Myers CT, Fagerstrom C, Anadiotis G, Akkari YM, Robbins KM, Gripp KW, Baratela WAR, Bober MB, Duker AL, Doherty D, Dempsey JC, Miller DG, Kircher M, Bamshad MJ, Nickerson DA; University of Washington Center for Mendelian Genomics, Mefford HC, Sol-Church K. GGC Repeat Expansion and Exon 1 Methylation of XYLT1 Is a Common Pathogenic Variant in Baratela-Scott Syndrome. Am J Hum Genet. 2019 Jan 3;104(1):35-44. doi: 10.1016/j.ajhg.2018.11.005. Epub 2018 Dec 13. PMID: 30554721; PMCID: PMC6323552.

Laggerbauer B, Ostareck D, Keidel EM, Ostareck-Lederer A, Fischer U. Evidence that fragile X mental retardation protein is a negative regulator of translation. Hum Mol Genet. 2001 Feb 15;10(4):329-38. doi: 10.1093/hmg/10.4.329. PMID: 11157796.

Lagier-Tourenne C, Baughn M, Rigo F, Sun S, Liu P, Li HR, Jiang J, Watt AT, Chun S, Katz M, Qiu J, Sun Y, Ling SC, Zhu Q, Polymenidou M, Drenner K, Artates JW, McAlonis-Downes M, Markmiller S, Hutt KR, Pizzo DP, Cady J, Harms MB, Baloh RH, Vandenberg SR, Yeo GW, Fu XD, Bennett CF, Cleveland DW, Ravits J. Targeted degradation of sense and antisense C9orf72 RNA foci as therapy for ALS and frontotemporal degeneration. Proc Natl Acad Sci U S A. 2013 Nov 19;110(47):E4530-9. doi: 10.1073/pnas.1318835110. Epub 2013 Oct 29. PMID: 24170860; PMCID: PMC3839752.

Lam YC, Bowman AB, Jafar-Nejad P, Lim J, Richman R, Fryer JD, Hyun ED, Duvick LA, Orr HT, Botas J, Zoghbi HY. ATAXIN-1 interacts with the repressor Capicua in its native complex to cause SCA1 neuropathology. Cell. 2006 Dec 29;127(7):1335-47. doi: 10.1016/j.cell.2006.11.038. PMID: 17190598.

Lee YB, Chen HJ, Peres JN, Gomez-Deza J, Attig J, Stalekar M, Troakes C, Nishimura AL, Scotter EL, Vance C, Adachi Y, Sardone V, Miller JW, Smith BN, Gallo JM, Ule J, Hirth F, Rogelj B, Houart C, Shaw CE. Hexanucleotide repeats in ALS/FTD form length-dependent RNA foci, sequester RNA binding proteins, and are neurotoxic. Cell Rep. 2013 Dec 12;5(5):1178-86. doi: 10.1016/j.celrep.2013.10.049. Epub 2013 Nov 27. PMID: 24290757; PMCID: PMC3898469.

Lee C, Zeng J, Drew BG, Sallam T, Martin-Montalvo A, Wan J, Kim SJ, Mehta H, Hevener AL, de Cabo R, Cohen P. The mitochondrial-derived peptide MOTS-c promotes metabolic homeostasis and reduces obesity and insulin resistance. Cell Metab. 2015 Mar 3;21(3):443-54. doi: 10.1016/j.cmet.2015.02.009. PMID: 25738459; PMCID: PMC4350682.

Lee KH, Zhang P, Kim HJ, Mitrea DM, Sarkar M, Freibaum BD, Cika J, Coughlin M, Messing J, Molliex A, Maxwell BA, Kim NC, Temirov J, Moore J, Kolaitis RM, Shaw TI, Bai B, Peng J, Kriwacki RW, Taylor JP. C9orf72 Dipeptide Repeats Impair the Assembly, Dynamics, and Function of Membrane-Less Organelles. Cell. 2016 Oct 20;167(3):774-788.e17. doi: 10.1016/j.cell.2016.10.002. PMID: 27768896; PMCID: PMC5079111.

Leehey MA, Munhoz RP, Lang AE, Brunberg JA, Grigsby J, Greco C, Jacquemont S, Tassone F, Lozano AM, Hagerman PJ, Hagerman RJ. The fragile X premutation presenting as essential tremor. Arch Neurol. 2003 Jan;60(1):117-21. doi: 10.1001/archneur.60.1.117. PMID: 12533098.

Lehmer C, Oeckl P, Weishaupt JH, Volk AE, Diehl-Schmid J, Schroeter ML, Lauer M, Kornhuber J, Levin J, Fassbender K, Landwehrmeyer B; German Consortium for Frontotemporal Lobar Degeneration, Schludi MH, Arzberger T, Kremmer E, Flatley A, Feederle R, Steinacker P, Weydt P, Ludolph AC, Edbauer D, Otto M. Poly-GP in cerebrospinal fluid links *C9orf72*-associated

dipeptide repeat expression to the asymptomatic phase of ALS/FTD. EMBO Mol Med. 2017 Jul;9(7):859-868. doi: 10.15252/emmm.201607486. PMID: 28408402; PMCID: PMC5494528.

Levine TP, Daniels RD, Gatta AT, Wong LH, Hayes MJ. The product of C9orf72, a gene strongly implicated in neurodegeneration, is structurally related to DENN Rab-GEFs. Bioinformatics. 2013 Feb 15;29(4):499-503. doi: 10.1093/bioinformatics/bts725. Epub 2013 Jan 16. PMID: 23329412; PMCID: PMC3570213.

Li WW, Li J, Bao JK. Microautophagy: lesser-known self-eating. Cell Mol Life Sci. 2012 Apr;69(7):1125-36. doi: 10.1007/s00018-011-0865-5. Epub 2011 Nov 12. PMID: 22080117.

Li S, Wu Z, Li Y, Tantray I, De Stefani D, Mattarei A, Krishnan G, Gao FB, Vogel H, Lu B. Altered MICOS Morphology and Mitochondrial Ion Homeostasis Contribute to Poly(GR) Toxicity Associated with C9-ALS/FTD. Cell Rep. 2020 Aug 4;32(5):107989. doi: 10.1016/j.celrep.2020.107989. PMID: 32755582; PMCID: PMC7433775.

Lieberman AP, Shakkottai VG, Albin RL. Polyglutamine Repeats in Neurodegenerative Diseases. Annu Rev Pathol. 2019 Jan 24;14:1-27. doi: 10.1146/annurev-pathmechdis-012418-012857. Epub 2018 Aug 8. PMID: 30089230; PMCID: PMC6387631.

Lin Y, Mori E, Kato M, Xiang S, Wu L, Kwon I, McKnight SL. Toxic PR Poly-Dipeptides Encoded by the C9orf72 Repeat Expansion Target LC Domain Polymers. Cell. 2016 Oct 20;167(3):789-802.e12. doi: 10.1016/j.cell.2016.10.003. PMID: 27768897; PMCID: PMC5076566.

Lim SY, Ishiura H, Ramli N, Shibata S, Almansour MA, Tan AH, Houlden H, Lang AE, Tsuji S. Adult-onset neuronal intranuclear inclusion disease mimicking Fragile X-associated tremor-ataxia syndrome in ethnic Chinese patients. Parkinsonism Relat Disord. 2020 May;74:25-27. doi: 10.1016/j.parkreldis.2020.03.025. Epub 2020 Apr 8. PMID: 32289521.

Lindenberg R, Rubinstein LJ, Herman MM, Haydon GB. A light and electron microscopy study of an unusual widespread nuclear inclusion body disease. A possible residuum of an old herpesvirus infection. Acta Neuropathol. 1968 Jan 2;10(1):54-73. doi: 10.1007/BF00690510. PMID: 4295804.

Liquori CL, Ricker K, Moseley ML, Jacobsen JF, Kress W, Naylor SL, Day JW, Ranum LP. Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. Science. 2001 Aug 3;293(5531):864-7. doi: 10.1126/science.1062125. PMID: 11486088.

Liu F, Liu Q, Lu CX, Cui B, Guo XN, Wang RR, Liu MS, Li XG, Cui LY, Zhang X. Identification of a novel loss-of-function C9orf72 splice site mutation in a patient with amyotrophic lateral sclerosis. Neurobiol Aging. 2016a Nov;47:219.e1-219.e5. doi: 10.1016/j.neurobiolaging.2016.07.027. Epub 2016 Aug 8. PMID: 27595458.

Liu Y, Pattamatta A, Zu T, Reid T, Bardhi O, Borchelt DR, Yachnis AT, Ranum LP. C9orf72 BAC Mouse Model with Motor Deficits and Neurodegenerative Features of ALS/FTD. Neuron. 2016b May 4;90(3):521-34. doi: 10.1016/j.neuron.2016.04.005. Epub 2016 Apr 21. PMID: 27112499.

Logroscino G, Traynor BJ, Hardiman O, Chiò A, Mitchell D, Swingler RJ, Millul A, Benn E, Beghi E; EURALS. Incidence of amyotrophic lateral sclerosis in Europe. J Neurol Neurosurg Psychiatry. 2010 Apr;81(4):385-90. doi: 10.1136/jnnp.2009.183525. Epub 2009 Aug 25. PMID: 19710046; PMCID: PMC2850819.

Loomis EW, Sanz LA, Chédin F, Hagerman PJ. Transcription-associated R-loop formation across the human FMR1 CGG-repeat region. PLoS Genet. 2014 Apr 17;10(4):e1004294. doi: 10.1371/journal.pgen.1004294. PMID: 24743386; PMCID: PMC3990486.

Lopez-Gonzalez R, Lu Y, Gendron TF, Karydas A, Tran H, Yang D, Petrucelli L, Miller BL, Almeida S, Gao FB. Poly(GR) in C9ORF72-Related ALS/FTD Compromises Mitochondrial Function and Increases Oxidative Stress and DNA Damage in iPSC-Derived Motor Neurons. Neuron. 2016 Oct 19;92(2):383-391. doi: 10.1016/j.neuron.2016.09.015. Epub 2016 Oct 6. PMID: 27720481; PMCID: PMC5111366.

Lu HC, Tan Q, Rousseaux MW, Wang W, Kim JY, Richman R, Wan YW, Yeh SY, Patel JM, Liu X, Lin T, Lee Y, Fryer JD, Han J, Chahrour M, Finnell RH, Lei Y, Zurita-Jimenez ME, Ahimaz P, Anyane-Yeboa K, Van Maldergem L, Lehalle D, Jean-Marcais N, Mosca-Boidron AL, Thevenon J, Cousin MA, Bro DE, Lanpher BC, Klee EW, Alexander N, Bainbridge MN, Orr HT, Sillitoe RV, Ljungberg MC, Liu Z, Schaaf CP, Zoghbi HY. Disruption of the ATXN1-CIC complex causes a spectrum of neurobehavioral phenotypes in mice and humans. Nat Genet. 2017 Apr;49(4):527-536. doi: 10.1038/ng.3808. Epub 2017 Mar 13. PMID: 28288114; PMCID: PMC5374026.

Lugenbeel KA, Peier AM, Carson NL, Chudley AE, Nelson DL. Intragenic loss of function mutations demonstrate the primary role of FMR1 in fragile X syndrome. Nat Genet. 1995 Aug;10(4):483-5. doi: 10.1038/ng0895-483. PMID: 7670500.

Lystad AH, Simonsen A. Mechanisms and Pathophysiological Roles of the ATG8 Conjugation Machinery. Cells. 2019 Aug 25;8(9):973. doi: 10.3390/cells8090973. PMID: 31450711; PMCID: PMC6769624.

M

Ma TC, Buescher JL, Oatis B, Funk JA, Nash AJ, Carrier RL, Hoyt KR. Metformin therapy in a transgenic mouse model of Huntington's disease. Neurosci Lett. 2007 Jan 10;411(2):98-103. Doi: 10.1016/j.neulet.2006.10.039. Epub 2006 Nov 15. PMID: 17110029.

Ma D, Tan YJ, Ng ASL, Ong HL, Sim W, Lim WK, Teo JX, Ng EYL, Lim EC, Lim EW, Chan LL, Tan LCS, Yi Z, Tan EK. Association of NOTCH2NLC Repeat Expansions With Parkinson Disease. JAMA Neurol. 2020 Dec 1;77(12):1559-1563. Doi: 10.1001/jamaneurol.2020.3023. PMID: 32852534; PMCID: PMC7445625.

Macejak DG, Sarnow P. Internal initiation of translation mediated by the 5' leader of a cellular mRNA. Nature. 1991 Sep 5;353(6339):90-4. Doi: 10.1038/353090a0. PMID: 1652694.

Mackenzie IR, Arzberger T, Kremmer E, Troost D, Lorenzl S, Mori K, Weng SM, Haass C, Kretzschmar HA, Edbauer D, Neumann M. Dipeptide repeat protein pathology in C9ORF72 mutation cases: 176harac-pathological correlations. Acta Neuropathol. 2013 Dec;126(6):859-79. Doi: 10.1007/s00401-013-1181-y. Epub 2013 Oct 6. PMID: 24096617.

Mackenzie IR, Frick P, Grässer FA, Gendron TF, Petrucelli L, Cashman NR, Edbauer D, Kremmer E, Prudlo J, Troost D, Neumann M. Quantitative analysis and 176harac-pathological correlations of different dipeptide repeat protein pathologies in C9ORF72 mutation carriers. Acta Neuropathol. 2015 Dec;130(6):845-61. Doi: 10.1007/s00401-015-1476-2. Epub 2015 Sep 15. PMID: 26374446.

Maeda S, Yamamoto H, Kinch LN, Garza CM, Takahashi S, Otomo C, Grishin NV, Forli S, Mizushima N, Otomo T. Structure, lipid scrambling activity and role in autophagosome formation of ATG9A. Nat Struct Mol Biol. 2020 Dec;27(12):1194-1201. Doi: 10.1038/s41594-020-00520-2. Epub 2020 Oct 26. PMID: 33106659; PMCID: PMC7718406.

Mahalati K, Kahan BD. Clinical pharmacokinetics of sirolimus. Clin Pharmacokinet. 2001;40(8):573-85. Doi: 10.2165/00003088-200140080-00002. PMID: 11523724.

Majounie E, Renton AE, Mok K, Dopper EG, Waite A, Rollinson S, Chiò A, Restagno G, Nicolaou N, Simon-Sanchez J, van Swieten JC, Abramzon Y, Johnson JO, Sendtner M, Pamphlett R, Orrell RW, Mead S, Sidle KC, Houlden H, Rohrer JD, Morrison KE, Pall H, Talbot K, Ansorge O; Chromosome 9-ALS/FTD Consortium; French research network on FTLD/FTLD/ALS; ITALSGEN Consortium, Hernandez DG, Arepalli S, Sabatelli M, Mora G, Corbo M, Giannini F, Calvo A, Englund E, Borghero G, Floris GL, Remes AM, Laaksovirta H, McCluskey L, Trojanowski JQ, Van Deerlin VM, Schellenberg GD, Nalls MA, Drory VE, Lu CS, Yeh TH, Ishiura H, Takahashi Y, Tsuji S, Le Ber I, Brice A, Drepper C, Williams N, Kirby J, Shaw P, Hardy J, Tienari PJ, Heutink P, Morris HR, Pickering-Brown S, Traynor BJ. Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study. Lancet Neurol. 2012 Apr;11(4):323-30. Doi: 10.1016/S1474-4422(12)70043-1. Epub 2012 Mar 9. PMID: 22406228; PMCID: PMC3322422.

Makino S, Kaji R, Ando S, Tomizawa M, Yasuno K, Goto S, Matsumoto S, Tabuena MD, Maranon E, Dantes M, Lee LV, Ogasawara K, Tooyama I, Akatsu H, Nishimura M, Tamiya G. Reduced neuron-specific expression of the TAF1 gene is associated with X-linked dystonia-parkinsonism. Am J Hum Genet. 2007 Mar;80(3):393-406. Doi: 10.1086/512129. Epub 2007 Jan 23. PMID: 17273961; PMCID: PMC1821114.

Malik I, Kelley CP, Wang ET, Todd PK. Molecular mechanisms underlying nucleotide repeat expansion disorders. Nat Rev Mol Cell Biol. 2021 Sep;22(9):589-607. Doi: 10.1038/s41580-021-00382-6. Epub 2021 Jun 17. Erratum in: Nat Rev Mol Cell Biol. 2021 Jul 6;: PMID: 34140671.

Man L, Lekovich J, Rosenwaks Z, Gerhardt J. Fragile X-Associated Diminished Ovarian Reserve and Primary Ovarian Insufficiency from Molecular Mechanisms to Clinical Manifestations. Front Mol

Neurosci. 2017 Sep 12;10:290. Doi: 10.3389/fnmol.2017.00290. PMID: 28955201; PMCID: PMC5600956.

Mangiarini L, Sathasivam K, Seller M, Cozens B, Harper A, Hetherington C, Lawton M, Trottier Y, Lehrach H, Davies SW, Bates GP. Exon 1 of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. Cell. 1996 Nov 1;87(3):493-506. Doi: 10.1016/s0092-8674(00)81369-0. PMID: 8898202.

Mankodi A, Takahashi MP, Jiang H, Beck CL, Bowers WJ, Moxley RT, Cannon SC, Thornton CA. Expanded CUG repeats trigger aberrant splicing of ClC-1 chloride channel pre-mRNA and hyperexcitability of skeletal muscle in myotonic dystrophy. Mol Cell. 2002 Jul;10(1):35-44. Doi: 10.1016/s1097-2765(02)00563-4. PMID: 12150905.

Maor-Nof M, Shipony Z, Lopez-Gonzalez R, Nakayama L, Zhang YJ, Couthouis J, Blum JA, Castruita PA, Linares GR, Ruan K, Ramaswami G, Simon DJ, Nof A, Santana M, Han K, Sinnott-Armstrong N, Bassik MC, Geschwind DH, Tessier-Lavigne M, Attardi LD, Lloyd TE, Ichida JK, Gao FB, Greenleaf WJ, Yokoyama JS, Petrucelli L, Gitler AD. P53 is a central regulator driving neurodegeneration caused by C9orf72 poly(PR). Cell. 2021 Feb 4;184(3):689-708.e20. doi: 10.1016/j.cell.2020.12.025. Epub 2021 Jan 21. PMID: 33482083; PMCID: PMC7886018.

Marrone L, Poser I, Casci I, Japtok J, Reinhardt P, Janosch A, Andree C, Lee HO, Moebius C, Koerner E, Reinhardt L, Cicardi ME, Hackmann K, Klink B, Poletti A, Alberti S, Bickle M, Hermann A, Pandey UB, Hyman AA, Sterneckert JL. Isogenic FUS-eGFP iPSC Reporter Lines Enable Quantification of FUS Stress Granule Pathology that Is Rescued by Drugs Inducing Autophagy. Stem Cell Reports. 2018 Feb 13;10(2):375-389. Doi: 10.1016/j.stemcr.2017.12.018. Epub 2018 Jan 18. PMID: 29358088; PMCID: PMC5857889.

Marrone L, Drexler HCA, Wang J, Tripathi P, Distler T, Heisterkamp P, Anderson EN, Kour S, Moraiti A, Maharana S, Bhatnagar R, Belgard TG, Tripathy V, Kalmbach N, Hosseinzadeh Z, Crippa V, Abo-Rady M, Wegner F, Poletti A, Troost D, Aronica E, Busskamp V, Weis J, Pandey UB, Hyman AA, Alberti S, Goswami A, Sterneckert J. FUS pathology in ALS is linked to alterations in multiple ALS-associated proteins and rescued by drugs stimulating autophagy. Acta Neuropathol. 2019 Jul;138(1):67-84. Doi: 10.1007/s00401-019-01998-x. Epub 2019 Apr 1. PMID: 30937520; PMCID: PMC6570784.

Martinez TF, Chu Q, Donaldson C, Tan D, Shokhirev MN, Saghatelian A. Accurate annotation of human protein-coding small open reading frames. Nat Chem Biol. 2020 Apr;16(4):458-468. Doi: 10.1038/s41589-019-0425-0. Epub 2019 Dec 9. PMID: 31819274; PMCID: PMC7085969.

Maruyama H, Morino H, Ito H, Izumi Y, Kato H, Watanabe Y, Kinoshita Y, Kamada M, Nodera H, Suzuki H, Komure O, Matsuura S, Kobatake K, Morimoto N, Abe K, Suzuki N, Aoki M, Kawata A, Hirai T, Kato T, Ogasawara K, Hirano A, Takumi T, Kusaka H, Hagiwara K, Kaji R, Kawakami H. Mutations of optineurin in amyotrophic lateral sclerosis. Nature. 2010 May 13;465(7295):223-6. Doi: 10.1038/nature08971. Epub 2010 Apr 28. PMID: 20428114.

Matoba K, Kotani T, Tsutsumi A, Tsuji T, Mori T, Noshiro D, Sugita Y, Nomura N, Iwata S, Ohsumi Y, Fujimoto T, Nakatogawa H, Kikkawa M, Noda NN. Atg9 is a lipid scramblase that mediates autophagosomal membrane expansion. Nat Struct Mol Biol. 2020 Dec;27(12):1185-1193. Doi: 10.1038/s41594-020-00518-w. Epub 2020 Oct 26. Erratum in: Nat Struct Mol Biol. 2020 Nov 12;: PMID: 33106658.

Matoba K, Noda NN. Secret of Atg9: lipid scramblase activity drives de novo autophagosome biogenesis. Cell Death Differ. 2020 Dec;27(12):3386-3388. Doi: 10.1038/s41418-020-00663-1. Epub 2020 Nov 11. PMID: 33177618; PMCID: PMC7852674.

Matthijs G, Schollen E, Legius E, Devriendt K, Goemans N, Kayserili H, Apäk MY, Cassiman JJ. Unusual molecular findings in autosomal recessive spinal muscular atrophy. J Med Genet. 1996 Jun;33(6):469-74. Doi: 10.1136/jmg.33.6.469. PMID: 8782046; PMCID: PMC1050632.

May S, Hornburg D, Schludi MH, Arzberger T, Rentzsch K, Schwenk BM, Grässer FA, Mori K, Kremmer E, Banzhaf-Strathmann J, Mann M, Meissner F, Edbauer D. C9orf72 FTLD/ALS-associated Gly-Ala dipeptide repeat proteins cause neuronal toxicity and Unc119 sequestration. Acta Neuropathol. 2014 Oct;128(4):485-503. Doi: 10.1007/s00401-014-1329-4. Epub 2014 Aug 14. PMID: 25120191; PMCID: PMC4159571.

McEachin ZT, Gendron TF, Raj N, García-Murias M, Banerjee A, Purcell RH, Ward PJ, Todd TW, Merritt-Garza ME, Jansen-West K, Hales CM, García-Sobrino T, Quintáns B, Holler CJ, Taylor G,

San Millán B, Teijeira S, Yamashita T, Ohkubo R, Boulis NM, Xu C, Wen Z, Streichenberger N; Neuro–CEB Neuropathology Network, Fogel BL, Kukar T, Abe K, Dickson DW, Arias M, Glass JD, Jiang J, Tansey MG, Sobrido MJ, Petrucelli L, Rossoll W, Bassell GJ. Chimeric Peptide Species Contribute to Divergent Dipeptide Repeat Pathology in c9ALS/FTD and SCA36. Neuron. 2020 Jul 22;107(2):292-305.e6. doi: 10.1016/j.neuron.2020.04.011. Epub 2020 May 5. PMID: 32375063; PMCID: PMC8138626.

McGoldrick P, Zhang M, van Blitterswijk M, Sato C, Moreno D, Xiao S, Zhang AB, McKeever PM, Weichert A, Schneider R, Keith J, Petrucelli L, Rademakers R, Zinman L, Robertson J, Rogaeva E. Unaffected mosaic *C9orf72* case: RNA foci, dipeptide proteins, but upregulated C9orf72 expression. Neurology. 2018 Jan 23;90(4):e323-e331. Doi: 10.1212/WNL.0000000000004865. Epub 2017 Dec 27. PMID: 29282338; PMCID: PMC5798652.

Meeter LHH, Gendron TF, Sias AC, Jiskoot LC, Russo SP, Donker Kaat L, Papma JM, Panman JL, van der Ende EL, Dopper EG, Franzen S, Graff C, Boxer AL, Rosen HJ, Sanchez-Valle R, Galimberti D, Pijnenburg YAL, Benussi L, Ghidoni R, Borroni B, Laforce R Jr, Del Campo M, Teunissen CE, van Minkelen R, Rojas JC, Coppola G, Geschwind DH, Rademakers R, Karydas AM, Öijerstedt L, Scarpini E, Binetti G, Padovani A, Cash DM, Dick KM, Bocchetta M, Miller BL, Rohrer JD, Petrucelli L, van Swieten JC, Lee SE. Poly(GP), neurofilament and grey matter deficits in *C9orf72* expansion carriers. Ann Clin Transl Neurol. 2018 Apr 6;5(5):583-597. Doi: 10.1002/I3.559. PMID: 29761121; PMCID: PMC5945959.

Mercer TJ, Ohashi Y, Boeing S, Jefferies HBJ, De Tito S, Flynn H, Tremel S, Zhang W, Wirth M, Frith D, Snijders AP, Williams RL, Tooze SA. Phosphoproteomic identification of ULK substrates reveals VPS15-dependent ULK/VPS34 interplay in the regulation of autophagy. EMBO J. 2021 Jul 15;40(14):e105985. Doi: 10.15252/embj.2020105985. Epub 2021 Jun 14. PMID: 34121209; PMCID: PMC8280838.

Mercy L, Hodges JR, Dawson K, Barker RA, Brayne C. Incidence of early-onset dementias in Cambridgeshire, United Kingdom. Neurology. 2008 Nov 4;71(19):1496-9. Doi: 10.1212/01.wnl.0000334277.16896.fa. PMID: 18981371.

Messaed C, Rouleau GA. Molecular mechanisms underlying polyalanine diseases. Neurobiol Dis. 2009 Jun;34(3):397-405. Doi: 10.1016/j.nbd.2009.02.013. Epub 2009 Mar 6. PMID: 19269323.

Miller JW, Urbinati CR, Teng-Umnuay P, Stenberg MG, Byrne BJ, Thornton CA, Swanson MS. Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy. EMBO J. 2000 Sep 1;19(17):4439-48. Doi: 10.1093/emboj/19.17.4439. PMID: 10970838; PMCID: PMC302046.

Mizielinska S, Lashley T, Norona FE, Clayton EL, Ridler CE, Fratta P, Isaacs AM. C9orf72 frontotemporal lobar degeneration is 178haracterized by frequent neuronal sense and antisense RNA foci. Acta Neuropathol. 2013 Dec;126(6):845-57. Doi: 10.1007/s00401-013-1200-z. Epub 2013 Oct 30. PMID: 24170096; PMCID: PMC3830745.

Mizielinska S, Grönke S, Niccoli T, Ridler CE, Clayton EL, Devoy A, Moens T, Norona FE, Woollacott IOC, Pietrzyk J, Cleverley K, Nicoll AJ, Pickering-Brown S, Dols J, Cabecinha M, Hendrich O, Fratta P, Fisher EMC, Partridge L, Isaacs AM. C9orf72 repeat expansions cause neurodegeneration in Drosophila through arginine-rich proteins. Science. 2014 Sep 5;345(6201):1192-1194. Doi: 10.1126/science.1256800. Epub 2014 Aug 7. PMID: 25103406; PMCID: PMC4944841.

Mootha VV, Hansen B, Rong Z, Mammen PP, Zhou Z, Xing C, Gong X. Fuchs' Endothelial Corneal Dystrophy and RNA Foci in Patients With Myotonic Dystrophy. Invest Ophthalmol Vis Sci. 2017 Sep 1;58(11):4579-4585. Doi: 10.1167/iovs.17-22350. PMID: 28886202; PMCID: PMC5590687.

Mordes DA, Prudencio M, Goodman LD, Klim JR, Moccia R, Limone F, Pietilainen O, Chowdhary K, Dickson DW, Rademakers R, Bonini NM, Petrucelli L, Eggan K. Dipeptide repeat proteins activate a heat shock response found in C9ORF72-ALS/FTLD patients. Acta Neuropathol Commun. 2018 Jul 4;6(1):55. Doi: 10.1186/s40478-018-0555-8. PMID: 29973287; PMCID: PMC6031111.

Mordes DA, Morrison BM, Ament XH, Cantrell C, Mok J, Eggan P, Xue C, Wang JY, Eggan K, Rothstein JD. Absence of Survival and Motor Deficits in 500 Repeat C9ORF72 BAC Mice. Neuron. 2020 Nov 25;108(4):775-783.e4. doi: 10.1016/j.neuron.2020.08.009. Epub 2020 Oct 5. PMID: 33022228.

Mori K, Weng SM, Arzberger T, May S, Rentzsch K, Kremmer E, Schmid B, Kretzschmar HA, Cruts M, Van Broeckhoven C, Haass C, Edbauer D. The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. Science. 2013a Mar 15;339(6125):1335-8. Doi: 10.1126/science.1232927. Epub 2013 Feb 7. PMID: 23393093.

Mori K, Arzberger T, Grässer FA, Gijselinck I, May S, Rentzsch K, Weng SM, Schludi MH, van der Zee J, Cruts M, Van Broeckhoven C, Kremmer E, Kretzschmar HA, Haass C, Edbauer D. Bidirectional transcripts of the expanded C9orf72 hexanucleotide repeat are translated into aggregating dipeptide repeat proteins. Acta Neuropathol. 2013b Dec;126(6):881-93. Doi: 10.1007/s00401-013-1189-3. Epub 2013 Oct 17. PMID: 24132570.

Mori K, Gotoh S, Yamashita T, Uozumi R, Kawabe Y, Tagami S, Kamp F, Nuscher B, Edbauer D, Haass C, Nagai Y, Ikeda M. The porphyrin TMPyP4 inhibits elongation during the noncanonical translation of the FTLD/ALS-associated GGGGCC repeat in the C9orf72 gene. J Biol Chem. 2021 Oct;297(4):101120. Doi: 10.1016/j.jbc.2021.101120. Epub 2021 Aug 25. PMID: 34450161; PMCID: PMC8446798.

Morimoto S, Hatsuta H, Komiya T, Kanemaru K, Tokumaru AM, Murayama S. Simultaneous skin-nerve-muscle biopsy and abnormal mitochondrial inclusions in intranuclear hyaline inclusion body disease. J Neurol Sci. 2017 Jan 15;372:447-449. Doi: 10.1016/j.jns.2016.10.042. Epub 2016 Oct 27. PMID: 27823834.

Moseley ML, Zu T, Ikeda Y, Gao W, Mosemiller AK, Daughters RS, Chen G, Weatherspoon MR, Clark HB, Ebner TJ, Day JW, Ranum LP. Bidirectional expression of CUG and CAG expansion transcripts and intranuclear polyglutamine inclusions in spinocerebellar ataxia type 8. Nat Genet. 2006 Jul;38(7):758-69. Doi: 10.1038/ng1827. Epub 2006 Jun 25. PMID: 16804541.

Mrakovic A, Kay JG, Furuya W, Brumell JH, Botelho RJ. Rab7 and Arl8 GTPases are necessary for lysosome tubulation in macrophages. Traffic. 2012 Dec;13(12):1667-79. Doi: 10.1111/tra.12003. Epub 2012 Sep 13. PMID: 22909026.

Mullard A. FDA approval for Biogen's aducanumab sparks Alzheimer disease firestorm. Nat Rev Drug Discov. 2021 Jul;20(7):496. Doi: 10.1038/d41573-021-00099-3. PMID: 34112945.

Müller S, Schöttler M, Schön S, Prante C, Brinkmann T, Kuhn J, Götting C, Kleesiek K. Human xylosyltransferase I: functional and biochemical characterization of cysteine residues required for enzymic activity. Biochem J. 2005 Mar 1;386(Pt 2):227-36. Doi: 10.1042/BJ20041206. PMID: 15461586; PMCID: PMC1134786.

Murray A, Webb J, Grimley S, Conway G, Jacobs P. Studies of FRAXA and FRAXE in women with premature ovarian failure. J Med Genet. 1998 Aug;35(8):637-40. Doi: 10.1136/jmg.35.8.637. PMID: 9719368; PMCID: PMC1051387.

Murray A, Schoemaker MJ, Bennett CE, Ennis S, Macpherson JN, Jones M, Morris DH, Orr N, Ashworth A, Jacobs PA, Swerdlow AJ. Population-based estimates of the prevalence of FMR1 expansion mutations in women with early menopause and primary ovarian insufficiency. Genet Med. 2014 Jan;16(1):19-24. Doi: 10.1038/gim.2013.64. Epub 2013 May 23. PMID: 23703681; PMCID: PMC3914024.

N

Naccache SN, Hasson T, Horowitz A. Binding of internalized receptors to the PDZ domain of GIPC/synectin recruits myosin VI to endocytic vesicles. Proc Natl Acad Sci U S A. 2006 Aug 22;103(34):12735-40. doi: 10.1073/pnas.0605317103. Epub 2006 Aug 14. Erratum in: Proc Natl Acad Sci U S A. 2006 Oct 10;103(41):15272. PMID: 16908842; PMCID: PMC1568917.

Nakamori M, Panigrahi GB, Lanni S, Gall-Duncan T, Hayakawa H, Tanaka H, Luo J, Otabe T, Li J, Sakata A, Caron MC, Joshi N, Prasolava T, Chiang K, Masson JY, Wold MS, Wang X, Lee MYWT, Huddleston J, Munson KM, Davidson S, Layeghifard M, Edward LM, Gallon R, Santibanez-Koref M, Murata A, Takahashi MP, Eichler EE, Shlien A, Nakatani K, Mochizuki H, Pearson CE. A slipped-CAG DNA-binding small molecule induces trinucleotide-repeat contractions in vivo. Nat Genet. 2020 Feb;52(2):146-159. doi: 10.1038/s41588-019-0575-8. Epub 2020 Feb 14. PMID: 32060489; PMCID: PMC7043212.

Nasir J, Floresco SB, O'Kusky JR, Diewert VM, Richman JM, Zeisler J, Borowski A, Marth JD, Phillips AG, Hayden MR. Targeted disruption of the Huntington's disease gene results in embryonic

lethality and behavioral and morphological changes in heterozygotes. Cell. 1995 Jun 2;81(5):811-23. doi: 10.1016/0092-8674(95)90542-1. PMID: 7774020.

Nelson BR, Makarewich CA, Anderson DM, Winders BR, Troupes CD, Wu F, Reese AL, McAnally JR, Chen X, Kavalali ET, Cannon SC, Houser SR, Bassel-Duby R, Olson EN. A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. Science. 2016 Jan 15;351(6270):271-5. doi: 10.1126/science.aad4076. PMID: 26816378; PMCID: PMC4892890.

Neumann M, Sampathu DM, Kwong LK, Truax AC, Micsenyi MC, Chou TT, Bruce J, Schuck T, Grossman M, Clark CM, McCluskey LF, Miller BL, Masliah E, Mackenzie IR, Feldman H, Feiden W, Kretzschmar HA, Trojanowski JQ, Lee VM. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Science. 2006 Oct 6;314(5796):130-3. doi: 10.1126/science.1134108. PMID: 17023659.

Ng ASL, Lim WK, Xu Z, Ong HL, Tan YJ, Sim WY, Ng EYL, Teo JX, Foo JN, Lim TCC, Yu WY, Chan LL, Lee HY, Chen Z, Lim EW, Ting SKS, Prakash KM, Tan LCS, Yi Z, Tan EK. NOTCH2NLC GGC Repeat Expansions Are Associated with Sporadic Essential Tremor: Variable Disease Expressivity on Long-Term Follow-up. Ann Neurol. 2020 Sep;88(3):614-618. doi: 10.1002/ana.25803. Epub 2020 Jul 16. PMID: 32495371.

Nguyen L, Montrasio F, Pattamatta A, Tusi SK, Bardhi O, Meyer KD, Hayes L, Nakamura K, Banez-Coronel M, Coyne A, Guo S, Laboissonniere LA, Gu Y, Narayanan S, Smith B, Nitsch RM, Kankel MW, Rushe M, Rothstein J, Zu T, Grimm J, Ranum LPW. Antibody Therapy Targeting RAN Proteins Rescues C9 ALS/FTD Phenotypes in C9orf72 Mouse Model. Neuron. 2020 Feb 19;105(4):645-662.e11. doi: 10.1016/j.neuron.2019.11.007. Epub 2019 Dec 9. PMID: 31831332; PMCID: PMC7391607.

Nicolas A, Kenna KP, Renton AE, Ticozzi N, Faghri F, Chia R, Dominov JA, Kenna BJ, Nalls MA, Keagle P, Rivera AM, van Rheenen W, Murphy NA, van Vugt JJFA, Geiger JT, Van der Spek RA, Pliner HA, Shankaracharya, Smith BN, Marangi G, Topp SD, Abramzon Y, Gkazi AS, Eicher JD, Kenna A; ITALSGEN Consortium, Mora G, Calvo A, Mazzini L, Riva N, Mandrioli J, Caponnetto C, Battistini S, Volanti P, La Bella V, Conforti FL, Borghero G, Messina S, Simone IL, Trojsi F, Salvi F, Logullo FO, D'Alfonso S, Corrado L, Capasso M, Ferrucci L; Genomic Translation for ALS Care (GTAC) Consortium, Moreno CAM, Kamalakaran S, Goldstein DB; ALS Sequencing Consortium, Gitler AD, Harris T, Myers RM; NYGC ALS Consortium, Phatnani H, Musunuri RL, Evani US, Abhyankar A, Zody MC; Answer ALS Foundation, Kaye J, Finkbeiner S, Wyman SK, LeNail A, Lima L, Fraenkel E, Svendsen CN, Thompson LM, Van Eyk JE, Berry JD, Miller TM, Kolb SJ, Cudkowicz M, Baxi E; Clinical Research in ALS and Related Disorders for Therapeutic Development (CReATe) Consortium, Benatar M, Taylor JP, Rampersaud E, Wu G, Wuu J; SLAGEN Consortium, Lauria G, Verde F, Fogh I, Tiloca C, Comi GP, Sorarù G, Cereda C; French ALS Consortium, Corcia P, Laaksovirta H, Myllykangas L, Jansson L, Valori M, Ealing J, Hamdalla H, Rollinson S, Pickering-Brown S, Orrell RW, Sidle KC, Malaspina A, Hardy J, Singleton AB, Johnson JO, Arepalli S, Sapp PC, McKenna-Yasek D, Polak M, Asress S, Al-Sarraj S, King A, Troakes C, Vance C, de Belleroche J, Baas F, Ten Asbroek ALMA, Muñoz-Blanco JL, Hernandez DG, Ding J, Gibbs JR, Scholz SW, Floeter MK, Campbell RH, Landi F, Bowser R, Pulst SM, Ravits JM, MacGowan DJL, Kirby J, Pioro EP, Pamphlett R, Broach J, Gerhard G, Dunckley TL, Brady CB, Kowall NW, Troncoso JC, Le Ber I, Mouzat K, Lumbroso S, Heiman-Patterson TD, Kamel F, Van Den Bosch L, Baloh RH, Strom TM, Meitinger T, Shatunov A, Van Eijk KR, de Carvalho M, Kooyman M, Middelkoop B, Moisse M, McLaughlin RL, Van Es MA, Weber M, Boylan KB, Van Blitterswijk M, Rademakers R, Morrison KE, Basak AN, Mora JS, Drory VE, Shaw PJ, Turner MR, Talbot K, Hardiman O, Williams KL, Fifita JA, Nicholson GA, Blair IP, Rouleau GA, Esteban-Pérez J, García-Redondo A, Al-Chalabi A; Project MinE ALS Sequencing Consortium, Rogaeva E, Zinman L, Ostrow LW, Maragakis NJ, Rothstein JD, Simmons Z, Cooper-Knock J, Brice A, Goutman SA, Feldman EL, Gibson SB, Taroni F, Ratti A, Gellera C, Van Damme P, Robberecht W, Fratta P, Sabatelli M, Lunetta C, Ludolph AC, Andersen PM, Weishaupt JH, Camu W, Trojanowski JQ, Van Deerlin VM, Brown RH Jr, van den Berg LH, Veldink JH, Harms MB, Glass JD, Stone DJ, Tienari P, Silani V, Chiò A, Shaw CE, Traynor BJ, Landers JE. Genome-wide Analyses Identify KIF5A as a Novel ALS Gene. Neuron. 2018 Mar 21;97(6):1268-1283.e6. doi: 10.1016/j.neuron.2018.02.027. PMID: 29566793; PMCID: PMC5867896.

Niimi Y, Takahashi M, Sugawara E, Umeda S, Obayashi M, Sato N, Ishiguro T, Higashi M, Eishi Y, Mizusawa H, Ishikawa K. Abnormal RNA structures (RNA foci) containing a penta-nucleotide repeat (UGGAA)n in the Purkinje cell nucleus is associated with spinocerebellar ataxia type 31 pathogenesis. Neuropathology. 2013 Dec;33(6):600-11. doi: 10.1111/neup.12032. Epub 2013 Apr 22. PMID: 23607545.

O

Oberlé I, Rousseau F, Heitz D, Kretz C, Devys D, Hanauer A, Boué J, Bertheas MF, Mandel JL. Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. Science. 1991 May 24;252(5009):1097-102. Doi: 10.1126/science.252.5009.1097. PMID: 2031184.

Ofer N, Weisman-Shomer P, Shklover J, Fry M. The quadruplex r(CGG)n destabilizing cationic porphyrin TMPyP4 cooperates with hnRNPs to increase the translation efficiency of fragile X premutation mRNA. Nucleic Acids Res. 2009 May;37(8):2712-22. Doi: 10.1093/nar/gkp130. Epub 2009 Mar 9. PMID: 19273535; PMCID: PMC2677883.

Ogasawara M, Iida A, Kumutpongpanich T, Ozaki A, Oya Y, Konishi H, Nakamura A, Abe R, Takai H, Hanajima R, Doi H, Tanaka F, Nakamura H, Nonaka I, Wang Z, Hayashi S, Noguchi S, Nishino I. CGG expansion in NOTCH2NLC is associated with oculopharyngodistal myopathy with neurological manifestations. Acta Neuropathol Commun. 2020 Nov 25;8(1):204. Doi: 10.1186/s40478-020-01084-4. PMID: 33239111; PMCID: PMC7690190.

Oh SY, He F, Krans A, Frazer M, Taylor JP, Paulson HL, Todd PK. RAN translation at CGG repeats induces ubiquitin proteasome system impairment in models of fragile X-associated tremor ataxia syndrome. Hum Mol Genet. 2015 Aug 1;24(15):4317-26. Doi: 10.1093/hmg/ddv165. Epub 2015 May 7. PMID: 25954027; PMCID: PMC4492395.

O'Hearn EE, Hwang HS, Holmes SE, Rudnicki DD, Chung DW, Seixas AI, Cohen RL, Ross CA, Trojanowski JQ, Pletnikova O, Troncoso JC, Margolis RL. Neuropathology and Cellular Pathogenesis of Spinocerebellar Ataxia Type 12. Mov Disord. 2015 Nov;30(13):1813-1824. Doi: 10.1002/mds.26348. Epub 2015 Sep 4. PMID: 26340331; PMCID: PMC5127409.

Ohki Y, Wenninger-Weinzierl A, Hruscha A, Asakawa K, Kawakami K, Haass C, Edbauer D, Schmid B. Glycine-alanine dipeptide repeat protein contributes to toxicity in a zebrafish model of C9orf72 associated neurodegeneration. Mol Neurodegener. 2017 Jan 14;12(1):6. Doi: 10.1186/s13024-016-0146-8. PMID: 28088213; PMCID: PMC5237533.

Olney NT, Spina S, Miller BL. Frontotemporal Dementia. Neurol Clin. 2017 May;35(2):339-374. Doi: 10.1016/j.ncl.2017.01.008. PMID: 28410663; PMCID: PMC5472209.

Olszewska DA, Lonergan R, Fallon EM, Lynch T. Genetics of Frontotemporal Dementia. Curr Neurol Neurosci Rep. 2016 Dec;16(12):107. Doi: 10.1007/s11910-016-0707-9. PMID: 27878525.

Ordway JM, Tallaksen-Greene S, Gutekunst CA, Bernstein EM, Cearley JA, Wiener HW, Dure LS 4th, Lindsey R, Hersch SM, Jope RS, Albin RL, Detloff PJ. Ectopically expressed CAG repeats cause intranuclear inclusions and a progressive late onset neurological phenotype in the mouse. Cell. 1997 Dec 12;91(6):753-63. Doi: 10.1016/s0092-8674(00)80464-x. PMID: 9413985.

O'Rawe JA, Wu Y, Dörfel MJ, Rope AF, Au PY, Parboosingh JS, Moon S, Kousi M, Kosma K, Smith CS, Tzetis M, Schuette JL, Hufnagel RB, Prada CE, Martinez F, Orellana C, Crain J, Caro-Llopis A, Oltra S, Monfort S, Jiménez-Barrón LT, Swensen J, Ellingwood S, Smith R, Fang H, Ospina S, Stegmann S, Den Hollander N, Mittelman D, Highnam G, Robison R, Yang E, Faivre L, Roubertie A, Rivière JB, Monaghan KG, Wang K, Davis EE, Katsanis N, Kalscheuer VM, Wang EH, Metcalfe K, Kleefstra T, Innes AM, Kitsiou-Tzeli S, Rosello M, Keegan CE, Lyon GJ. TAF1 Variants Are Associated with Dysmorphic Features, Intellectual Disability, and Neurological Manifestations. Am J Hum Genet. 2015 Dec 3;97(6):922-32. Doi: 10.1016/j.ajhg.2015.11.005. PMID: 26637982; PMCID: PMC4678794.

O'Rourke JG, Bogdanik L, Muhammad AKMG, Gendron TF, Kim KJ, Austin A, Cady J, Liu EY, Zarrow J, Grant S, Ho R, Bell S, Carmona S, Simpkinson M, Lall D, Wu K, Daughrity L, Dickson DW, Harms MB, Petrucelli L, Lee EB, Lutz CM, Baloh RH. C9orf72 BAC Transgenic Mice Display Typical Pathologic Features of ALS/FTD. Neuron. 2015 Dec 2;88(5):892-901. Doi: 10.1016/j.neuron.2015.10.027. PMID: 26637796; PMCID: PMC4672384.

O'Rourke JG, Bogdanik L, Yáñez A, Lall D, Wolf AJ, Muhammad AK, Ho R, Carmona S, Vit JP, Zarrow J, Kim KJ, Bell S, Harms MB, Miller TM, Dangler CA, Underhill DM, Goodridge HS, Lutz CM, Baloh RH. C9orf72 is required for proper macrophage and microglial function in mice. Science. 2016 Mar 18 ;351(6279) :1324-9. Doi : 10.1126/science.aaf1064. PMID : 26989253 ; PMCID : PMC5120541.

Osaki T, Uzel SGM, Kamm RD. Microphysiological 3D model of amyotrophic lateral sclerosis (ALS) from human iPS-derived muscle cells and optogenetic motor neurons. Sci Adv. 2018 Oct 10;4(10):eaat5847. Doi: 10.1126/sciadv.aat5847. PMID: 30324134; PMCID: PMC6179377.

Osawa T, Kotani T, Kawaoka T, Hirata E, Suzuki K, Nakatogawa H, Ohsumi Y, Noda NN. Atg2 mediates direct lipid transfer between membranes for autophagosome formation. Nat Struct Mol Biol. 2019 Apr;26(4):281-288. Doi: 10.1038/s41594-019-0203-4. Epub 2019 Mar 25. PMID: 30911189.

P

Pal A, Kretner B, Abo-Rady M, Glaβ H, Dash BP, Naumann M, Japtok J, Kreiter N, Dhingra A, Heutink P, Böckers TM, Günther R, Sterneckert J, Hermann A. Concomitant gain and loss of function pathomechanisms in C9ORF72 amyotrophic lateral sclerosis. Life Sci Alliance. 2021 Feb 22;4(4):e202000764. doi: 10.26508/lsa.202000764. PMID: 33619157; PMCID: PMC7918691.

Park H, Kang JH, Lee S. Autophagy in Neurodegenerative Diseases: A Hunter for Aggregates. Int J Mol Sci. 2020 May 10;21(9):3369. doi: 10.3390/ijms21093369. PMID: 32397599; PMCID: PMC7247013.

Parzych KR, Klionsky DJ. An overview of autophagy: morphology, mechanism, and regulation. Antioxid Redox Signal. 2014 Jan 20;20(3):460-73. doi: 10.1089/ars.2013.5371. Epub 2013 Aug 2. PMID: 23725295; PMCID: PMC3894687.

Patten SA, Aggad D, Martinez J, Tremblay E, Petrillo J, Armstrong GA, La Fontaine A, Maios C, Liao M, Ciura S, Wen XY, Rafuse V, Ichida J, Zinman L, Julien JP, Kabashi E, Robitaille R, Korngut L, Parker JA, Drapeau P. Neuroleptics as therapeutic compounds stabilizing neuromuscular transmission in amyotrophic lateral sclerosis. JCI Insight. 2017 Nov 16;2(22):e97152. doi: 10.1172/jci.insight.97152. PMID: 29202456; PMCID: PMC5752378.

Peabody DS. Translation initiation at an ACG triplet in mammalian cells. J Biol Chem. 1987 Aug 25;262(24):11847-51. PMID: 3040720.

Peabody DS. Translation initiation at non-AUG triplets in mammalian cells. J Biol Chem. 1989 Mar 25;264(9):5031-5. PMID: 2538469.

Perez BA, Shorrock HK, Banez-Coronel M, Zu T, Romano LE, Laboissonniere LA, Reid T, Ikeda Y, Reddy K, Gomez CM, Bird T, Ashizawa T, Schut LJ, Brusco A, Berglund JA, Hasholt LF, Nielsen JE, Subramony SH, Ranum LP. CCG•CGG interruptions in high-penetrance SCA8 families increase RAN translation and protein toxicity. EMBO Mol Med. 2021 Nov 8;13(11):e14095. doi: 10.15252/emmm.202114095. Epub 2021 Oct 11. PMID: 34632710; PMCID: PMC8573593.

Peters OM, Cabrera GT, Tran H, Gendron TF, McKeon JE, Metterville J, Weiss A, Wightman N, Salameh J, Kim J, Sun H, Boylan KB, Dickson D, Kennedy Z, Lin Z, Zhang YJ, Daughrity L, Jung C, Gao FB, Sapp PC, Horvitz HR, Bosco DA, Brown SP, de Jong P, Petrucelli L, Mueller C, Brown RH Jr. Human C9ORF72 Hexanucleotide Expansion Reproduces RNA Foci and Dipeptide Repeat Proteins but Not Neurodegeneration in BAC Transgenic Mice. Neuron. 2015 Dec 2;88(5):902-909. doi: 10.1016/j.neuron.2015.11.018. PMID: 26637797; PMCID: PMC4828340.

Philips AV, Timchenko LT, Cooper TA. Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. Science. 1998 May 1;280(5364):737-41. doi: 10.1126/science.280.5364.737. PMID: 9563950.

Philips T, Robberecht W. Neuroinflammation in amyotrophic lateral sclerosis: role of glial activation in motor neuron disease. Lancet Neurol. 2011 Mar;10(3):253-63. doi: 10.1016/S1474-4422(11)70015-1. PMID: 21349440.

Pisarev AV, Kolupaeva VG, Pisareva VP, Merrick WC, Hellen CU, Pestova TV. Specific functional interactions of nucleotides at key -3 and +4 positions flanking the initiation codon with components

of the mammalian 48S translation initiation complex. Genes Dev. 2006 Mar 1;20(5):624-36. doi: 10.1101/gad.1397906. PMID: 16510876; PMCID: PMC1410799.

Q

Qu L, Pan C, He SM, Lang B, Gao GD, Wang XL, Wang Y. The Ras Superfamily of Small GTPases in Non-neoplastic Cerebral Diseases. Front Mol Neurosci. 2019 May 21;12:121. Doi: 10.3389/fnmol.2019.00121. PMID: 31213978; PMCID: PMC6555388.

Quaegebeur A, Glaria I, Lashley T, Isaacs AM. Soluble and insoluble dipeptide repeat protein measurements in C9orf72-frontotemporal dementia brains show regional differential solubility and correlation of poly-GR with clinical severity. Acta Neuropathol Commun. 2020 Nov 9;8(1):184. Doi: 10.1186/s40478-020-01036-y. PMID: 33168090; PMCID: PMC7650212.

R

Rafehi H, Szmulewicz DJ, Bennett MF, Sobreira NLM, Pope K, Smith KR, Gillies G, Diakumis P, Dolzhenko E, Eberle MA, Barcina MG, Breen DP, Chancellor AM, Cremer PD, Delatycki MB, Fogel BL, Hackett A, Halmagyi GM, Kapetanovic S, Lang A, Mossman S, Mu W, Patrikios P, Perlman SL, Rosemergy I, Storey E, Watson SRD, Wilson MA, Zee DS, Valle D, Amor DJ, Bahlo M, Lockhart PJ. Bioinformatics-Based Identification of Expanded Repeats: A Non-reference Intronic Pentamer Expansion in RFC1 Causes CANVAS. Am J Hum Genet. 2019 Jul 3;105(1):151-165. doi: 10.1016/j.ajhg.2019.05.016. Epub 2019 Jun 20. PMID: 31230722; PMCID: PMC6612533.

Rathore A, Chu Q, Tan D, Martinez TF, Donaldson CJ, Diedrich JK, Yates JR 3rd, Saghatelian A. MIEF1 Microprotein Regulates Mitochondrial Translation. Biochemistry. 2018 Sep 25;57(38):5564-5575. doi: 10.1021/acs.biochem.8b00726. Epub 2018 Sep 14. PMID: 30215512; PMCID: PMC6443411.

Rau F, Lainé J, Ramanoudjame L, Ferry A, Arandel L, Delalande O, Jollet A, Dingli F, Lee KY, Peccate C, Lorain S, Kabashi E, Athanasopoulos T, Koo T, Loew D, Swanson MS, Le Rumeur E, Dickson G, Allamand V, Marie J, Furling D. Abnormal splicing switch of DMD's penultimate exon compromises muscle fibre maintenance in myotonic dystrophy. Nat Commun. 2015 May 28;6:7205. doi: 10.1038/ncomms8205. PMID: 26018658; PMCID: PMC4458869.

Reddy A, Caler EV, Andrews NW. Plasma membrane repair is mediated by Ca(2+)-regulated exocytosis of lysosomes. Cell. 2001 Jul 27;106(2):157-69. doi: 10.1016/s0092-8674(01)00421-4. PMID: 11511344.

Reddy K, Zamiri B, Stanley SYR, Macgregor RB Jr, Pearson CE. The disease-associated r(GGGGCC)n repeat from the C9orf72 gene forms tract length-dependent uni- and multimolecular RNA G-quadruplex structures. J Biol Chem. 2013 Apr 5;288(14):9860-9866. doi: 10.1074/jbc.C113.452532. Epub 2013 Feb 19. PMID: 23423380; PMCID: PMC3617286.

Renko M, Quarto N, Morimoto T, Rifkin DB. Nuclear and cytoplasmic localization of different basic fibroblast growth factor species. J Cell Physiol. 1990 Jul;144(1):108-14. doi: 10.1002/jcp.1041440114. PMID: 2195042.

Renton AE, Majounie E, Waite A, Simón-Sánchez J, Rollinson S, Gibbs JR, Schymick JC, Laaksovirta H, van Swieten JC, Myllykangas L, Kalimo H, Paetau A, Abramzon Y, Remes AM, Kaganovich A, Scholz SW, Duckworth J, Ding J, Harmer DW, Hernandez DG, Johnson JO, Mok K, Ryten M, Trabzuni D, Guerreiro RJ, Orrell RW, Neal J, Murray A, Pearson J, Jansen IE, Sondervan D, Seelaar H, Blake D, Young K, Halliwell N, Callister JB, Toulson G, Richardson A, Gerhard A, Snowden J, Mann D, Neary D, Nalls MA, Peuralinna T, Jansson L, Isoviita VM, Kaivorinne AL, Hölttä-Vuori M, Ikonen E, Sulkava R, Benatar M, Wuu J, Chiò A, Restagno G, Borghero G, Sabatelli M; ITALSGEN Consortium, Heckerman D, Rogaeva E, Zinman L, Rothstein JD, Sendtner M, Drepper C, Eichler EE, Alkan C, Abdullaev Z, Pack SD, Dutra A, Pak E, Hardy J, Singleton A, Williams NM, Heutink P, Pickering-Brown S, Morris HR, Tienari PJ, Traynor BJ. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron. 2011 Oct 20;72(2):257-68. doi: 10.1016/j.neuron.2011.09.010. Epub 2011 Sep 21. PMID: 21944779; PMCID: PMC3200438.

Restituito S, Thompson RM, Eliet J, Raike RS, Riedl M, Charnet P, Gomez CM. The polyglutamine expansion in spinocerebellar ataxia type 6 causes a beta subunit-specific enhanced activation of P/Q-type calcium channels in Xenopus oocytes. J Neurosci. 2000 Sep 1;20(17):6394-403. doi: 10.1523/JNEUROSCI.20-17-06394.2000. PMID: 10964945; PMCID: PMC6772973.

Richard P, Trollet C, Gidaro T, Demay L, Brochier G, Malfatti E, Tom FM, Fardeau M, Lafor P, Romero N, Martin-N ML, Sol G, Ferrer-Monasterio X, Saint-Guily JL, Eymard B. PABPN1 (GCN)11 as a Dominant Allele in Oculopharyngeal Muscular Dystrophy -Consequences in Clinical Diagnosis and Genetic Counselling. J Neuromuscul Dis. 2015 Jun 4;2(2):175-180. doi: 10.3233/JND-140060. PMID: 27858728; PMCID: PMC5271460.

Riemslagh FW, Lans H, Seelaar H, Severijnen LWFM, Melhem S, Vermeulen W, Aronica E, Pasterkamp RJ, van Swieten JC, Willemsen R. HR23B pathology preferentially co-localizes with p62, pTDP-43 and poly-GA in C9ORF72-linked frontotemporal dementia and amyotrophic lateral sclerosis. Acta Neuropathol Commun. 2019 Mar 13;7(1):39. doi: 10.1186/s40478-019-0694-6. PMID: 30867060; PMCID: PMC6416930.

Riemslagh FW, Verhagen RFM, van der Toorn EC, Smits DJ, Quint WH, van der Linde HC, van Ham TJ, Willemsen R. Reduction of oxidative stress suppresses poly-GR mediated toxicity in zebrafish embryos. Dis Model Mech. 2021 Oct 25:dmm.049092. doi: 10.1242/dmm.049092. Epub ahead of print. PMID: 34693978.

Ringholz GM, Appel SH, Bradshaw M, Cooke NA, Mosnik DM, Schulz PE. Prevalence and patterns of cognitive impairment in sporadic ALS. Neurology. 2005 Aug 23;65(4):586-90. doi: 10.1212/01.wnl.0000172911.39167.b6. PMID: 16116120.

Robin G, López JR, Espinal GM, Hulsizer S, Hagerman PJ, Pessah IN. Calcium dysregulation and Cdk5-ATM pathway involved in a mouse model of fragile X-associated tremor/ataxia syndrome. Hum Mol Genet. 2017 Jul 15;26(14):2649-2666. doi: 10.1093/hmg/ddx148. PMID: 28444183; PMCID: PMC5886271.

Rodriguez CM, Wright SE, Kearse MG, Haenfler JM, Flores BN, Liu Y, Ifrim MF, Glineburg MR, Krans A, Jafar-Nejad P, Sutton MA, Bassell GJ, Parent JM, Rigo F, Barmada SJ, Todd PK. A native function for RAN translation and CGG repeats in regulating fragile X protein synthesis. Nat Neurosci. 2020 Mar;23(3):386-397. doi: 10.1038/s41593-020-0590-1. Epub 2020 Feb 17. PMID: 32066985; PMCID: PMC7668390.

Rosen DR. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. Nature. 1993 Jul 22;364(6435):362. doi: 10.1038/364362c0. Erratum for: Nature. 1993 Mar 4;362(6415):59-62. PMID: 8332197.

Rousseaux MWC, Tschumperlin T, Lu HC, Lackey EP, Bondar VV, Wan YW, Tan Q, Adamski CJ, Friedrich J, Twaroski K, Chen W, Tolar J, Henzler C, Sharma A, Bajić A, Lin T, Duvick L, Liu Z, Sillitoe RV, Zoghbi HY, Orr HT. ATXN1-CIC Complex Is the Primary Driver of Cerebellar Pathology in Spinocerebellar Ataxia Type 1 through a Gain-of-Function Mechanism. Neuron. 2018 Mar 21;97(6):1235-1243.e5. doi: 10.1016/j.neuron.2018.02.013. Epub 2018 Mar 8. PMID: 29526553; PMCID: PMC6422678.

Rudelli RD, Brown WT, Wisniewski K, Jenkins EC, Laure-Kamionowska M, Connell F, Wisniewski HM. Adult fragile X syndrome. Clinico-neuropathologic findings. Acta Neuropathol. 1985;67(3-4):289-95. doi: 10.1007/BF00687814. PMID: 4050344.

S

Saberi S, Stauffer JE, Schulte DJ, Ravits J. Neuropathology of Amyotrophic Lateral Sclerosis and Its Variants. Neurol Clin. 2015 Nov;33(4):855-76. doi: 10.1016/j.ncl.2015.07.012. PMID: 26515626; PMCID: PMC4628785.

Sancak Y, Bar-Peled L, Zoncu R, Markhard AL, Nada S, Sabatini DM. Ragulator-Rag complex targets mTORC1 to the lysosomal surface and is necessary for its activation by amino acids. Cell. 2010 Apr 16;141(2):290-303. doi: 10.1016/j.cell.2010.02.024. Epub 2010 Apr 8. PMID: 20381137; PMCID: PMC3024592.

Sareen D, O'Rourke JG, Meera P, Muhammad AK, Grant S, Simpkinson M, Bell S, Carmona S, Ornelas L, Sahabian A, Gendron T, Petrucelli L, Baughn M, Ravits J, Harms MB, Rigo F, Bennett CF, Otis TS, Svendsen CN, Baloh RH. Targeting RNA foci in iPSC-derived motor neurons from

ALS patients with a C9ORF72 repeat expansion. Sci Transl Med. 2013 Oct 23;5(208):208ra149. doi: 10.1126/scitranslmed.3007529. PMID: 24154603; PMCID: PMC4090945.

Sasaki T, Hideyama T, Saito Y, Shimizu J, Maekawa R, Shiio Y. Neuronal intranuclear inclusion disease presenting with recurrent cerebral infarct-like lesions. Neurol. Clin. Neurosci. 2015 Feb;3: 185-187. Doi: 10.1111/ncn3.178

Sato N, Amino T, Kobayashi K, Asakawa S, Ishiguro T, Tsunemi T, Takahashi M, Matsuura T, Flanigan KM, Iwasaki S, Ishino F, Saito Y, Murayama S, Yoshida M, Hashizume Y, Takahashi Y, Tsuji S, Shimizu N, Toda T, Ishikawa K, Mizusawa H. Spinocerebellar ataxia type 31 is associated with "inserted" penta-nucleotide repeats containing (TGGAA)n. Am J Hum Genet. 2009 Nov;85(5):544-57. doi: 10.1016/j.ajhg.2009.09.019. Epub 2009 Oct 29. PMID: 19878914; PMCID: PMC2775824.

Satoyoshi E, Kinoshita M. Oculopharyngodistal myopathy. Arch Neurol. 1977 Feb;34(2):89-92. doi: 10.1001/archneur.1977.00500140043007. PMID: 836191.

Saudou F, Finkbeiner S, Devys D, Greenberg ME. Huntingtin acts in the nucleus to induce apoptosis but death does not correlate with the formation of intranuclear inclusions. Cell. 1998 Oct 2;95(1):55-66. doi: 10.1016/s0092-8674(00)81782-1. PMID: 9778247.

Savkur RS, Philips AV, Cooper TA. Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy. Nat Genet. 2001 Sep;29(1):40-7. doi: 10.1038/ng704. PMID: 11528389.

Sawa-Makarska J, Baumann V, Coudevylle N, von Bülow S, Nogellova V, Abert C, Schuschnig M, Graef M, Hummer G, Martens S. Reconstitution of autophagosome nucleation defines Atg9 vesicles as seeds for membrane formation. Science. 2020 Sep 4;369(6508):eaaz7714. doi: 10.1126/science.aaz7714. PMID: 32883836; PMCID: PMC7610778.

Saxton RA, Sabatini DM. mTOR Signaling in Growth, Metabolism, and Disease. Cell. 2017 Mar 9;168(6):960-976. doi: 10.1016/j.cell.2017.02.004. Erratum in: Cell. 2017 Apr 6;169(2):361-371. PMID: 28283069; PMCID: PMC5394987.

Schiffer D, Cordera S, Giordana MT, Attanasio A, Pezzulo T. Synaptic vesicle proteins, synaptophysin and chromogranin A in amyotrophic lateral sclerosis. J Neurol Sci. 1995 May;129 Suppl:68-74. doi: 10.1016/0022-510x(95)00068-d. PMID: 7595626.

Schludi MH, May S, Grässer FA, Rentzsch K, Kremmer E, Küpper C, Klopstock T; German Consortium for Frontotemporal Lobar Degeneration; Bavarian Brain Banking Alliance, Arzberger T, Edbauer D. Distribution of dipeptide repeat proteins in cellular models and C9orf72 mutation cases suggests link to transcriptional silencing. Acta Neuropathol. 2015 Oct;130(4):537-55. doi: 10.1007/s00401-015-1450-z. Epub 2015 Jun 18. Erratum in: Acta Neuropathol. 2015 Oct;130(4):557-8. PMID: 26085200; PMCID: PMC4575390.

Schludi MH, Becker L, Garrett L, Gendron TF, Zhou Q, Schreiber F, Popper B, Dimou L, Strom TM, Winkelmann J, von Thaden A, Rentzsch K, May S, Michaelsen M, Schwenk BM, Tan J, Schoser B, Dieterich M, Petrucelli L, Hölter SM, Wurst W, Fuchs H, Gailus-Durner V, de Angelis MH, Klopstock T, Arzberger T, Edbauer D. Spinal poly-GA inclusions in a C9orf72 mouse model trigger motor deficits and inflammation without neuron loss. Acta Neuropathol. 2017 Aug;134(2):241-254. doi: 10.1007/s00401-017-1711-0. Epub 2017 Apr 13. PMID: 28409281; PMCID: PMC5508040.

Schneider S, Gulacsi A, Hatten ME. Lrp12/Mig13a reveals changing patterns of preplate neuronal polarity during corticogenesis that are absent in reeler mutant mice. Cereb Cortex. 2011 Jan;21(1):134-44. doi: 10.1093/cercor/bhq070. Epub 2010 May 3. PMID: 20439316; PMCID: PMC3000567.

Sellier C, Rau F, Liu Y, Tassone F, Hukema RK, Gattoni R, Schneider A, Richard S, Willemsen R, Elliott DJ, Hagerman PJ, Charlet-Berguerand N. Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients. EMBO J. 2010 Apr 7;29(7):1248-61. doi: 10.1038/emboj.2010.21. Epub 2010 Feb 25. PMID: 20186122; PMCID: PMC2857464.

Sellier C, Campanari ML, Julie Corbier C, Gaucherot A, Kolb-Cheynel I, Oulad-Abdelghani M, Ruffenach F, Page A, Ciura S, Kabashi E, Charlet-Berguerand N. Loss of C9ORF72 impairs autophagy and synergizes with polyQ Ataxin-2 to induce motor neuron dysfunction and cell death. EMBO J. 2016 Jun 15;35(12):1276-97. doi: 10.15252/embj.201593350. Epub 2016 Apr 21. PMID: 27103069; PMCID: PMC4910533.

Sellier C, Buijsen RAM, He F, Natla S, Jung L, Tropel P, Gaucherot A, Jacobs H, Meziane H, Vincent A, Champy MF, Sorg T, Pavlovic G, Wattenhofer-Donze M, Birling MC, Oulad-Abdelghani M, Eberling P, Ruffenach F, Joint M, Anheim M, Martinez-Cerdeno V, Tassone F, Willemsen R, Hukema RK, Viville S, Martinat C, Todd PK, Charlet-Berguerand N. Translation of Expanded CGG Repeats into FMRpolyG Is Pathogenic and May Contribute to Fragile X Tremor Ataxia Syndrome. Neuron. 2017 Jan 18;93(2):331-347. doi: 10.1016/j.neuron.2016.12.016. Epub 2017 Jan 5. PMID: 28065649; PMCID: PMC5263258.

Sellier C, Cerro-Herreros E, Blatter M, Freyermuth F, Gaucherot A, Ruffenach F, Sarkar P, Puymirat J, Udd B, Day JW, Meola G, Bassez G, Fujimura H, Takahashi MP, Schoser B, Furling D, Artero R, Allain FHT, Llamusi B, Charlet-Berguerand N. rbFOX1/MBNL1 competition for CCUG RNA repeats binding contributes to myotonic dystrophy type 1/type 2 differences. Nat Commun. 2018 May 22;9(1):2009. doi: 10.1038/s41467-018-04370-x. PMID: 29789616; PMCID: PMC5964235.

Sevigny J, Chiao P, Bussière T, Weinreb PH, Williams L, Maier M, Dunstan R, Salloway S, Chen T, Ling Y, O'Gorman J, Qian F, Arastu M, Li M, Chollate S, Brennan MS, Quintero-Monzon O, Scannevin RH, Arnold HM, Engber T, Rhodes K, Ferrero J, Hang Y, Mikulskis A, Grimm J, Hock C, Nitsch RM, Sandrock A. The antibody aducanumab reduces Aβ plaques in Alzheimer's disease. Nature. 2016 Sep 1;537(7618):50-6. doi: 10.1038/nature19323. Update in: Nature. 2017 Jun 21;546(7659):564. PMID: 27582220.

Shaid S, Brandts CH, Serve H, Dikic I. Ubiquitination and selective autophagy. Cell Death Differ. 2013 Jan;20(1):21-30. doi: 10.1038/cdd.2012.72. Epub 2012 Jun 22. PMID: 22722335; PMCID: PMC3524631.

Shaikh SA, Sahoo SK, Periasamy M. Phospholamban and sarcolipin: Are they functionally redundant or distinct regulators of the Sarco(Endo)Plasmic Reticulum Calcium ATPase? J Mol Cell Cardiol. 2016 Feb;91:81-91. doi: 10.1016/j.yjmcc.2015.12.030. Epub 2015 Dec 29. PMID: 26743715; PMCID: PMC4843517.

Shao Q, Yang M, Liang C, Ma L, Zhang W, Jiang Z, Luo J, Lee JK, Liang C, Chen JF. *C9orf72* and *smcr8* mutant mice reveal MTORC1 activation due to impaired lysosomal degradation and exocytosis. Autophagy. 2020 Sep;16(9):1635-1650. doi: 10.1080/15548627.2019.1703353. Epub 2019 Dec 26. PMID: 31847700; PMCID: PMC8386604.

Shaw PJ, Ince PG. Glutamate, excitotoxicity and amyotrophic lateral sclerosis. J Neurol. 1997 May;244 Suppl 2:S3-14. doi: 10.1007/BF03160574. PMID: 9178165.

Shelly KE, Candelaria NR, Li Z, Allen EG, Jin P, Nelson DL. Ectopic expression of CGG-repeats alters ovarian response to gonadotropins and leads to infertility in a murine FMR1 premutation model. Hum Mol Genet. 2021 May 29;30(10):923-938. doi: 10.1093/hmg/ddab083. PMID: 33856019; PMCID: PMC8165648.

Shi Y, Lin S, Staats KA, Li Y, Chang WH, Hung ST, Hendricks E, Linares GR, Wang Y, Son EY, Wen X, Kisler K, Wilkinson B, Menendez L, Sugawara T, Woolwine P, Huang M, Cowan MJ, Ge B, Koutsodendris N, Sandor KP, Komberg J, Vangoor VR, Senthilkumar K, Hennes V, Seah C, Nelson AR, Cheng TY, Lee SJ, August PR, Chen JA, Wisniewski N, Hanson-Smith V, Belgard TG, Zhang A, Coba M, Grunseich C, Ward ME, van den Berg LH, Pasterkamp RJ, Trotti D, Zlokovic BV, Ichida JK. Haploinsufficiency leads to neurodegeneration in C9ORF72 ALS/FTD human induced motor neurons. Nat Med. 2018 Mar;24(3):313-325. doi: 10.1038/nm.4490. Epub 2018 Feb 5. PMID: 29400714; PMCID: PMC6112156.

Simone R, Balendra R, Moens TG, Preza E, Wilson KM, Heslegrave A, Woodling NS, Niccoli T, Gilbert-Jaramillo J, Abdelkarim S, Clayton EL, Clarke M, Konrad MT, Nicoll AJ, Mitchell JS, Calvo A, Chio A, Houlden H, Polke JM, Ismail MA, Stephens CE, Vo T, Farahat AA, Wilson WD, Boykin DW, Zetterberg H, Partridge L, Wray S, Parkinson G, Neidle S, Patani R, Fratta P, Isaacs AM. G-quadruplex-binding small molecules ameliorate *C9orf72* FTD/ALS pathology *in vitro* and *in vivo*. EMBO Mol Med. 2018 Jan;10(1):22-31. doi: 10.15252/emmm.201707850. PMID: 29113975; PMCID: PMC5760849.

Sofola OA, Jin P, Qin Y, Duan R, Liu H, de Haro M, Nelson DL, Botas J. RNA-binding proteins hnRNP A2/B1 and CUGBP1 suppress fragile X CGG premutation repeat-induced neurodegeneration in a Drosophila model of FXTAS. Neuron. 2007 Aug 16;55(4):565-71. doi: 10.1016/j.neuron.2007.07.021. PMID: 17698010; PMCID: PMC2215388.

Sone J, Tanaka F, Koike H, Inukai A, Katsuno M, Yoshida M, Watanabe H, Sobue G. Skin biopsy is useful for the antemortem diagnosis of neuronal intranuclear inclusion disease. Neurology. 2011 Apr 19;76(16):1372-6. doi: 10.1212/WNL.0b013e3182166e13. Epub 2011 Mar 16. PMID: 21411744.

Sone J, Mori K, Inagaki T, Katsumata R, Takagi S, Yokoi S, Araki K, Kato T, Nakamura T, Koike H, Takashima H, Hashiguchi A, Kohno Y, Kurashige T, Kuriyama M, Takiyama Y, Tsuchiya M, Kitagawa N, Kawamoto M, Yoshimura H, Suto Y, Nakayasu H, Uehara N, Sugiyama H, Takahashi M, Kokubun N, Konno T, Katsuno M, Tanaka F, Iwasaki Y, Yoshida M, Sobue G. Clinicopathological features of adult-onset neuronal intranuclear inclusion disease. Brain. 2016 Dec;139(Pt 12):3170-3186. doi: 10.1093/brain/aww249. Epub 2016 Oct 25. PMID: 27797808; PMCID: PMC5382941.

Sone J, Mitsuhashi S, Fujita A, Mizuguchi T, Hamanaka K, Mori K, Koike H, Hashiguchi A, Takashima H, Sugiyama H, Kohno Y, Takiyama Y, Maeda K, Doi H, Koyano S, Takeuchi H, Kawamoto M, Kohara N, Ando T, Ieda T, Kita Y, Kokubun N, Tsuboi Y, Katoh K, Kino Y, Katsuno M, Iwasaki Y, Yoshida M, Tanaka F, Suzuki IK, Frith MC, Matsumoto N, Sobue G. Long-read sequencing identifies GGC repeat expansions in NOTCH2NLC associated with neuronal intranuclear inclusion disease. Nat Genet. 2019 Aug;51(8):1215-1221. doi: 10.1038/s41588-019-0459-y. Epub 2019 Jul 22. PMID: 31332381.

Sonobe Y, Ghadge G, Masaki K, Sendoel A, Fuchs E, Roos RP. Translation of dipeptide repeat proteins from the C9ORF72 expanded repeat is associated with cellular stress. Neurobiol Dis. 2018 Aug;116:155-165. doi: 10.1016/j.nbd.2018.05.009. Epub 2018 May 22. PMID: 29792928; PMCID: PMC6331062.

Sonobe Y, Aburas J, Krishnan G, Fleming AC, Ghadge G, Islam P, Warren EC, Gu Y, Kankel MW, Brown AEX, Kiskinis E, Gendron TF, Gao FB, Roos RP, Kratsios P. A C. elegans model of C9orf72-associated ALS/FTD uncovers a conserved role for eIF2D in RAN translation. Nat Commun. 2021 Oct 15;12(1):6025. doi: 10.1038/s41467-021-26303-x. PMID: 34654821; PMCID: PMC8519953.

Soragni E, Petrosyan L, Rinkoski TA, Wieben ED, Baratz KH, Fautsch MP, Gottesfeld JM. Repeat-Associated Non-ATG (RAN) Translation in Fuchs' Endothelial Corneal Dystrophy. Invest Ophthalmol Vis Sci. 2018 Apr 1;59(5):1888-1896. doi: 10.1167/iovs.17-23265. PMID: 29677349; PMCID: PMC5886103.

Sproviero W, Shatunov A, Stahl D, Shoai M, van Rheenen W, Jones AR, Al-Sarraj S, Andersen PM, Bonini NM, Conforti FL, Van Damme P, Daoud H, Del Mar Amador M, Fogh I, Forzan M, Gaastra B, Gellera C, Gitler AD, Hardy J, Fratta P, La Bella V, Le Ber I, Van Langenhove T, Lattante S, Lee YC, Malaspina A, Meininger V, Millecamps S, Orrell R, Rademakers R, Robberecht W, Rouleau G, Ross OA, Salachas F, Sidle K, Smith BN, Soong BW, Sorarù G, Stevanin G, Kabashi E, Troakes C, van Broeckhoven C, Veldink JH, van den Berg LH, Shaw CE, Powell JF, Al-Chalabi A. ATXN2 trinucleotide repeat length correlates with risk of ALS. Neurobiol Aging. 2017 Mar;51:178.e1-178.e9. doi: 10.1016/j.neurobiolaging.2016.11.010. Epub 2016 Nov 24. PMID: 28017481; PMCID: PMC5302215.

Stein CS, Jadiya P, Zhang X, McLendon JM, Abouassaly GM, Witmer NH, Anderson EJ, Elrod JW, Boudreau RL. Mitoregulin: A lncRNA-Encoded Microprotein that Supports Mitochondrial Supercomplexes and Respiratory Efficiency. Cell Rep. 2018 Jun 26;23(13):3710-3720.e8. doi: 10.1016/j.celrep.2018.06.002. PMID: 29949756; PMCID: PMC6091870.

Stenmark H, Olkkonen VM. The Rab GTPase family. Genome Biol. 2001;2(5):REVIEWS3007. doi: 10.1186/gb-2001-2-5-reviews3007. Epub 2001 Apr 27. PMID: 11387043; PMCID: PMC138937.

Su Z, Zhang Y, Gendron TF, Bauer PO, Chew J, Yang WY, Fostvedt E, Jansen-West K, Belzil VV, Desaro P, Johnston A, Overstreet K, Oh SY, Todd PK, Berry JD, Cudkowicz ME, Boeve BF, Dickson D, Floeter MK, Traynor BJ, Morelli C, Ratti A, Silani V, Rademakers R, Brown RH, Rothstein JD, Boylan KB, Petrucelli L, Disney MD. Discovery of a biomarker and lead small molecules to target r(GGGGCC)-associated defects in c9FTD/ALS. Neuron. 2014 Sep 3;83(5):1043-50. doi: 10.1016/j.neuron.2014.07.041. Epub 2014 Aug 14. Erratum in: Neuron. 2014 Oct 1;84(1):239. PMID: 25132468; PMCID: PMC4232217.

Su MY, Fromm SA, Zoncu R, Hurley JH. Structure of the C9orf72 ARF GAP complex that is haploinsufficient in ALS and FTD. Nature. 2020 Sep;585(7824):251-255. doi: 10.1038/s41586-020-2633-x. Epub 2020 Aug 26. PMID: 32848248; PMCID: PMC8054479.

Su MY, Fromm SA, Remis J, Toso DB, Hurley JH. Structural basis for the ARF GAP activity and specificity of the C9orf72 complex. Nat Commun. 2021 Jun 18;12(1):3786. doi: 10.1038/s41467-021-24081-0. PMID: 34145292; PMCID: PMC8213707.

Sudria-Lopez E, Koppers M, de Wit M, van der Meer C, Westeneng HJ, Zundel CA, Youssef SA, Harkema L, de Bruin A, Veldink JH, van den Berg LH, Pasterkamp RJ. Full ablation of C9orf72 in mice causes immune system-related pathology and neoplastic events but no motor neuron defects. Acta Neuropathol. 2016 Jul;132(1):145-7. doi: 10.1007/s00401-016-1581-x. Epub 2016 May 20. PMID: 27206760; PMCID: PMC4911370.

Suh E, Lee EB, Neal D, Wood EM, Toledo JB, Rennert L, Irwin DJ, McMillan CT, Krock B, Elman LB, McCluskey LF, Grossman M, Xie SX, Trojanowski JQ, Van Deerlin VM. Semi-automated quantification of C9orf72 expansion size reveals inverse correlation between hexanucleotide repeat number and disease duration in frontotemporal degeneration. Acta Neuropathol. 2015 Sep;130(3):363-72. doi: 10.1007/s00401-015-1445-9. Epub 2015 May 29. PMID: 26022924; PMCID: PMC4545720.

Sullivan PM, Zhou X, Robins AM, Paushter DH, Kim D, Smolka MB, Hu F. The ALS/FTLD associated protein C9orf72 associates with SMCR8 and WDR41 to regulate the autophagy-lysosome pathway. Acta Neuropathol Commun. 2016 May 18;4(1):51. doi: 10.1186/s40478-016-0324-5. PMID: 27193190; PMCID: PMC4870812.

Sundin OH, Broman KW, Chang HH, Vito EC, Stark WJ, Gottsch JD. A common locus for late-onset Fuchs corneal dystrophy maps to 18q21.2-q21.32. Invest Ophthalmol Vis Sci. 2006 Sep;47(9):3919-26. doi: 10.1167/iovs.05-1619. PMID: 16936105.

Suzuki IK, Gacquer D, Van Heurck R, Kumar D, Wojno M, Bilheu A, Herpoel A, Lambert N, Cheron J, Polleux F, Detours V, Vanderhaeghen P. Human-Specific NOTCH2NL Genes Expand Cortical Neurogenesis through Delta/Notch Regulation. Cell. 2018a May 31;173(6):1370-1384.e16. doi: 10.1016/j.cell.2018.03.067. Epub 2018 May 31. PMID: 29856955; PMCID: PMC6092419.

Suzuki H, Shibagaki Y, Hattori S, Matsuoka M. The proline-arginine repeat protein linked to C9-ALS/FTD causes neuronal toxicity by inhibiting the DEAD-box RNA helicase-mediated ribosome biogenesis. Cell Death Dis. 2018b Sep 24;9(10):975. doi: 10.1038/s41419-018-1028-5. PMID: 30250194; PMCID: PMC6155127.

Swaminathan A, Bouffard M, Liao M, Ryan S, Callister JB, Pickering-Brown SM, Armstrong GAB, Drapeau P. Expression of C9orf72-related dipeptides impairs motor function in a vertebrate model. Hum Mol Genet. 2018 May 15;27(10):1754-1762. doi: 10.1093/hmg/ddy083. PMID: 29528390; PMCID: PMC5932562.

Swash M, Desai J. Motor neuron disease: classification and nomenclature. Amyotroph Lateral Scler Other Motor Neuron Disord. 2000 Mar;1(2):105-12. doi: 10.1080/14660820050515403. PMID: 11467047.

Swinnen B, Robberecht W. The phenotypic variability of amyotrophic lateral sclerosis. Nat Rev Neurol. 2014 Nov;10(11):661-70. doi: 10.1038/nrneurol.2014.184. Epub 2014 Oct 14. PMID: 25311585.

Switonski PM, Szlachcic WJ, Krzyzosiak WJ, Figiel M. A new humanized ataxin-3 knock-in mouse model combines the genetic features, pathogenesis of neurons and glia and late disease onset of SCA3/MJD. Neurobiol Dis. 2015 Jan;73:174-88. doi: 10.1016/j.nbd.2014.09.020. Epub 2014 Oct 7. PMID: 25301414.

T

Tabet R, Moutin E, Becker JA, Heintz D, Fouillen L, Flatter E, Krężel W, Alunni V, Koebel P, Dembélé D, Tassone F, Bardoni B, Mandel JL, Vitale N, Muller D, Le Merrer J, Moine H. Fragile X Mental Retardation Protein (FMRP) controls diacylglycerol kinase activity in neurons. Proc Natl Acad Sci U S A. 2016 Jun 28;113(26):E3619-28. Doi: 10.1073/pnas.1522631113. Epub 2016 May 27. PMID: 27233938; PMCID: PMC4932937.

Tabet R, Schaeffer L, Freyermuth F, Jambeau M, Workman M, Lee CZ, Lin CC, Jiang J, Jansen-West K, Abou-Hamdan H, Désaubry L, Gendron T, Petrucelli L, Martin F, Lagier-Tourenne C. CUG initiation and frameshifting enable production of dipeptide repeat proteins from ALS/FTD C9ORF72 transcripts. Nat Commun. 2018 Jan 11;9(1):152. Doi: 10.1038/s41467-017-02643-5. PMID: 29323119; PMCID: PMC5764992.

Takumida H, Yakabe M, Mori H, Shibasaki K, Umeda-Kameyama Y, Urano T, Mano T, Hayashi A, Ikemura M, Ogawa S, Akishita M. Case of a 78-year-old woman with a neuronal intranuclear inclusion disease. Geriatr Gerontol Int. 2017 Dec;17(12):2623-2625. Doi: 10.1111/ggi.13174. PMID: 29265760.

Talaia G, Amick J, Ferguson SM. Receptor-like role for PQLC2 amino acid transporter in the lysosomal sensing of cationic amino acids. Proc Natl Acad Sci U S A. 2021 Feb 23;118(8):e2014941118. Doi: 10.1073/pnas.2014941118. PMID: 33597295; PMCID: PMC7923529.

Tang D, Sheng J, Xu L, Zhan X, Liu J, Jiang H, Shu X, Liu X, Zhang T, Jiang L, Zhou C, Li W, Cheng W, Li Z, Wang K, Lu K, Yan C, Qi S. Cryo-EM structure of C9ORF72-SMCR8-WDR41 reveals the role as a GAP for Rab8a and Rab11a. Proc Natl Acad Sci U S A. 2020 May 5;117(18):9876-9883. Doi: 10.1073/pnas.2002110117. Epub 2020 Apr 17. PMID: 32303654; PMCID: PMC7211967.

Tao Z, Wang H, Xia Q, Li K, Li K, Jiang X, Xu G, Wang G, Ying Z. Nucleolar stress and impaired stress granule formation contribute to C9orf72 RAN translation-induced cytotoxicity. Hum Mol Genet. 2015 May 1;24(9):2426-41. Doi: 10.1093/hmg/ddv005. Epub 2015 Jan 9. PMID: 25575510.

Tarleton J, Kenneson A, Taylor AK, Crandall K, Fletcher R, Casey R, Hart PS, Hatton D, Fisch G, Warren ST. A single base alteration in the CGG repeat region of FMR1: possible effects on gene expression and phenotype. J Med Genet. 2002 Mar;39(3):196-200. Doi: 10.1136/jmg.39.3.196. PMID: 11897823; PMCID: PMC1735062.

Tassone F, Hagerman RJ, Loesch DZ, Lachiewicz A, Taylor AK, Hagerman PJ. Fragile X males with unmethylated, full mutation trinucleotide repeat expansions have elevated levels of FMR1 messenger RNA. Am J Med Genet. 2000 Sep 18;94(3):232-6. Doi: 10.1002/1096-8628(20000918). PMID: 10995510.

Tekirdag K, Cuervo AM. Chaperone-mediated autophagy and endosomal microautophagy: Joint by a chaperone. J Biol Chem. 2018 Apr 13;293(15):5414-5424. Doi: 10.1074/jbc.R117.818237. Epub 2017 Dec 15. PMID: 29247007; PMCID: PMC5900761.

Tian Y, Wang JL, Huang W, Zeng S, Jiao B, Liu Z, Chen Z, Li Y, Wang Y, Min HX, Wang XJ, You Y, Zhang RX, Chen XY, Yi F, Zhou YF, Long HY, Zhou CJ, Hou X, Wang JP, Xie B, Liang F, Yang ZY, Sun QY, Allen EG, Shafik AM, Kong HE, Guo JF, Yan XX, Hu ZM, Xia K, Jiang H, Xu HW, Duan RH, Jin P, Tang BS, Shen L. Expansion of Human-Specific GGC Repeat in Neuronal Intranuclear Inclusion Disease-Related Disorders. Am J Hum Genet. 2019 Jul 3;105(1):166-176. Doi: 10.1016/j.ajhg.2019.05.013. Epub 2019 Jun 6. PMID: 31178126; PMCID: PMC6612530.

Todd PK, Oh SY, Krans A, He F, Sellier C, Frazer M, Renoux AJ, Chen KC, Scaglione KM, Basrur V, Elenitoba-Johnson K, Vonsattel JP, Louis ED, Sutton MA, Taylor JP, Mills RE, Charlet-Berguerand N, Paulson HL. CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. Neuron. 2013 May 8;78(3):440-55. Doi: 10.1016/j.neuron.2013.03.026. Epub 2013 Apr 18. Erratum in: Neuron. 2013 Jul 24;79(2):402. PMID: 23602499; PMCID: PMC3831531.

Todd TW, McEachin ZT, Chew J, Burch AR, Jansen-West K, Tong J, Yue M, Song Y, Castanedes-Casey M, Kurti A, Dunmore JH, Fryer JD, Zhang YJ, San Millan B, Teijeira Bautista S, Arias M, Dickson D, Gendron TF, Sobrido MJ, Disney MD, Bassell GJ, Rossoll W, Petrucelli L. Hexanucleotide Repeat Expansions in c9FTD/ALS and SCA36 Confer Selective Patterns of Neurodegeneration In Vivo. Cell Rep. 2020 May 5;31(5):107616. Doi: 10.1016/j.celrep.2020.107616. PMID: 32375043; PMCID: PMC7480900.

Toko M, Ohshita T, Kurashige T, Morino H, Kume K, Yamashita H, Sobue G, Iwasaki Y, Sone J, Kawakami H, Maruyama H. FXTAS is difficult to differentiate from neuronal intranuclear inclusion disease through skin biopsy: a case report. BMC Neurol. 2021 Oct 12;21(1):396. Doi: 10.1186/s12883-021-02425-z. Erratum in: BMC Neurol. 2021 Oct 27;21(1):413. PMID: 34641814; PMCID: PMC8513318.

Tomé FM, Fardeau M. Nuclear inclusions in oculopharyngeal dystrophy. Acta Neuropathol. 1980;49(1):85-7. Doi: 10.1007/BF00692226. PMID: 6243839.

Tran H, Almeida S, Moore J, Gendron TF, Chalasani U, Lu Y, Du X, Nickerson JA, Petrucelli L, Weng Z, Gao FB. Differential Toxicity of Nuclear RNA Foci versus Dipeptide Repeat Proteins in a Drosophila Model of C9ORF72 FTD/ALS. Neuron. 2015 Sep 23;87(6):1207-1214. Doi: 10.1016/j.neuron.2015.09.015. PMID: 26402604; PMCID: PMC4589299.

Trollet C, Anvar SY, Venema A, Hargreaves IP, Foster K, Vignaud A, Ferry A, Negroni E, Hourde C, Baraibar MA, 't Hoen PA, Davies JE, Rubinsztein DC, Heales SJ, Mouly V, van der Maarel SM, Butler-Browne G, Raz V, Dickson G. Molecular and phenotypic characterization of a mouse model of oculopharyngeal muscular dystrophy reveals severe muscular atrophy restricted to fast glycolytic fibres. Hum Mol Genet. 2010 Jun 1;19(11):2191-207. Doi: 10.1093/hmg/ddq098. Epub 2010 Mar 5. PMID: 20207626.

Trottier Y, Lutz Y, Stevanin G, Imbert G, Devys D, Cancel G, Saudou F, Weber C, David G, Tora L, et al. Polyglutamine expansion as a pathological epitope in Huntington's disease and four dominant cerebellar ataxias. Nature. 1995 Nov 23;378(6555):403-6. Doi: 10.1038/378403a0. PMID: 7477379.

Tsvetkov AS, Miller J, Arrasate M, Wong JS, Pleiss MA, Finkbeiner S. A small-molecule scaffold induces autophagy in primary neurons and protects against toxicity in a Huntington disease model. Proc Natl Acad Sci U S A. 2010 Sep 28;107(39):16982-7. Doi: 10.1073/pnas.1004498107. Epub 2010 Sep 10. PMID: 20833817; PMCID: PMC2947884.

Turner MR, Al-Chalabi A, Chio A, Hardiman O, Kiernan MC, Rohrer JD, Rowe J, Seeley W, Talbot K. Genetic screening in sporadic ALS and FTD. J Neurol Neurosurg Psychiatry. 2017 Dec;88(12):1042-1044. Doi: 10.1136/jnnp-2017-315995. Epub 2017 Jun 22. PMID: 28642287; PMCID: PMC5740553.

U

Ugolino J, Ji YJ, Conchina K, Chu J, Nirujogi RS, Pandey A, Brady NR, Hamacher-Brady A, Wang J. Loss of C9orf72 Enhances Autophagic Activity via Deregulated mTOR and TFEB Signaling. PLoS Genet. 2016 Nov 22;12(11):e1006443. doi: 10.1371/journal.pgen.1006443. PMID: 27875531; PMCID: PMC5119725.

V

Vagner S, Touriol C, Galy B, Audigier S, Gensac MC, Amalric F, Bayard F, Prats H, Prats AC. Translation of CUG- but not AUG-initiated forms of human fibroblast growth factor 2 is activated in transformed and stressed cells. J Cell Biol. 1996 Dec;135(5):1391-402. Doi: 10.1083/jcb.135.5.1391. PMID: 8947560; PMCID: PMC2121090.

Valverde DP, Yu S, Boggavarapu V, Kumar N, Lees JA, Walz T, Reinisch KM, Melia TJ. ATG2 transports lipids to promote autophagosome biogenesis. J Cell Biol. 2019 Jun 3;218(6):1787-1798. Doi: 10.1083/jcb.201811139. Epub 2019 Apr 5. PMID: 30952800; PMCID: PMC6548141.

Van Eijk RPA, Jones AR, Sproviero W, Shatunov A, Shaw PJ, Leigh PN, Young CA, Shaw CE, Mora G, Mandrioli J, Borghero G, Volanti P, Diekstra FP, van Rheenen W, Verstraete E, Eijkemans MJC, Veldink JH, Chio A, Al-Chalabi A, van den Berg LH, van Es MA; For UKMND-LiCALS and LITALS Study Group. Meta-analysis of pharmacogenetic interactions in amyotrophic lateral sclerosis clinical trials. Neurology. 2017 Oct 31;89(18):1915-1922. Doi: 10.1212/WNL.0000000000004606. Epub 2017 Oct 4. Erratum in: Neurology. 2017 Nov 28;89(22):2303. PMID: 28978660; PMCID: PMC5664299.

Vance C, Rogelj B, Hortobágyi T, De Vos KJ, Nishimura AL, Sreedharan J, Hu X, Smith B, Ruddy D, Wright P, Ganesalingam J, Williams KL, Tripathi V, Al-Saraj S, Al-Chalabi A, Leigh PN, Blair IP, Nicholson G, de Belleroche J, Gallo JM, Miller CC, Shaw CE. Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. Science. 2009 Feb 27;323(5918):1208-1211. Doi: 10.1126/science.1165942. PMID: 19251628; PMCID: PMC4516382.

Vattem KM, Wek RC. Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. Proc Natl Acad Sci U S A. 2004 Aug 3;101(31):11269-74. Doi: 10.1073/pnas.0400541101. Epub 2004 Jul 26. PMID: 15277680; PMCID: PMC509193.

Verkerk AJ, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DP, Pizzuti A, Reiner O, Richards S, Victoria MF, Zhang FP, et al. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell. 1991 May 31;65(5):905-14. Doi: 10.1016/0092-8674(91)90397-h. PMID: 1710175.

Vieira RT, Caixeta L, Machado S, Silva AC, Nardi AE, Arias-Carrión O, Carta MG. Epidemiology of early-onset dementia: a review of the literature. Clin Pract Epidemiol Ment Health. 2013 Jun 14;9:88-95. Doi: 10.2174/1745017901309010088. PMID: 23878613; PMCID: PMC3715758.

Villacé P, Marión RM, Ortín J. The composition of Staufen-containing RNA granules from human cells indicates their role in the regulated transport and translation of messenger RNAs. Nucleic Acids Res. 2004 Apr 30;32(8):2411-20. Doi: 10.1093/nar/gkh552. PMID: 15121898; PMCID: PMC419443.

W

Waite AJ, Bäumer D, East S, Neal J, Morris HR, Ansorge O, Blake DJ. Reduced C9orf72 protein levels in frontal cortex of amyotrophic lateral sclerosis and frontotemporal degeneration brain with the C9ORF72 hexanucleotide repeat expansion. Neurobiol Aging. 2014 Jul;35(7):1779.e5-1779.e13. doi: 10.1016/j.neurobiolaging.2014.01.016. Epub 2014 Jan 17. PMID: 24559645; PMCID: PMC3988882.

Wang IF, Guo BS, Liu YC, Wu CC, Yang CH, Tsai KJ, Shen CK. Autophagy activators rescue and alleviate pathogenesis of a mouse model with proteinopathies of the TAR DNA-binding protein 43. Proc Natl Acad Sci U S A. 2012 Sep 11;109(37):15024-9. doi: 10.1073/pnas.1206362109. Epub 2012 Aug 29. PMID: 22932872; PMCID: PMC3443184.

Wang M, Wang H, Tao Z, Xia Q, Hao Z, Prehn JHM, Zhen X, Wang G, Ying Z. C9orf72 associates with inactive Rag GTPases and regulates mTORC1-mediated autophagosomal and lysosomal biogenesis. Aging Cell. 2020 Apr;19(4):e13126. doi: 10.1111/acel.13126. Epub 2020 Feb 25. PMID: 32100453; PMCID: PMC7189992.

Webster CP, Smith EF, Bauer CS, Moller A, Hautbergue GM, Ferraiuolo L, Myszczynska MA, Higginbottom A, Walsh MJ, Whitworth AJ, Kaspar BK, Meyer K, Shaw PJ, Grierson AJ, De Vos KJ. The C9orf72 protein interacts with Rab1a and the ULK1 complex to regulate initiation of autophagy. EMBO J. 2016 Aug 1;35(15):1656-76. doi: 10.15252/embj.201694401. Epub 2016 Jun 22. PMID: 27334615; PMCID: PMC4969571.

Wen X, Tan W, Westergard T, Krishnamurthy K, Markandaiah SS, Shi Y, Lin S, Shneider NA, Monaghan J, Pandey UB, Pasinelli P, Ichida JK, Trotti D. Antisense proline-arginine RAN dipeptides linked to C9ORF72-ALS/FTD form toxic nuclear aggregates that initiate in vitro and in vivo neuronal death. Neuron. 2014 Dec 17;84(6):1213-25. doi: 10.1016/j.neuron.2014.12.010. PMID: 25521377; PMCID: PMC4632245.

Westenberger A, Reyes CJ, Saranza G, Dobricic V, Hanssen H, Domingo A, Laabs BH, Schaake S, Pozojevic J, Rakovic A, Grütz K, Begemann K, Walter U, Dressler D, Bauer P, Rolfs A, Münchau A, Kaiser FJ, Ozelius LJ, Jamora RD, Rosales RL, Diesta CCE, Lohmann K, König IR, Brüggemann N, Klein C. A hexanucleotide repeat modifies expressivity of X-linked dystonia parkinsonism. Ann Neurol. 2019 Jun;85(6):812-822. doi: 10.1002/ana.25488. Epub 2019 May 3. PMID: 30973967.

Westergard T, Jensen BK, Wen X, Cai J, Kropf E, Iacovitti L, Pasinelli P, Trotti D. Cell-to-Cell Transmission of Dipeptide Repeat Proteins Linked to C9orf72-ALS/FTD. Cell Rep. 2016 Oct 11;17(3):645-652. doi: 10.1016/j.celrep.2016.09.032. PMID: 27732842; PMCID: PMC5078984.

Wheeler VC, White JK, Gutekunst CA, Vrbanac V, Weaver M, Li XJ, Li SH, Yi H, Vonsattel JP, Gusella JF, Hersch S, Auerbach W, Joyner AL, MacDonald ME. Long glutamine tracts cause nuclear localization of a novel form of huntingtin in medium spiny striatal neurons in HdhQ92 and HdhQ111 knock-in mice. Hum Mol Genet. 2000 Mar 1;9(4):503-13. doi: 10.1093/hmg/9.4.503. PMID: 10699173.

White MR, Mitrea DM, Zhang P, Stanley CB, Cassidy DE, Nourse A, Phillips AH, Tolbert M, Taylor JP, Kriwacki RW. C9orf72 Poly(PR) Dipeptide Repeats Disturb Biomolecular Phase Separation and Disrupt Nucleolar Function. Mol Cell. 2019 May 16;74(4):713-728.e6. doi: 10.1016/j.molcel.2019.03.019. Epub 2019 Apr 10. PMID: 30981631; PMCID: PMC6525025.

Wieben ED, Aleff RA, Tang X, Butz ML, Kalari KR, Highsmith EW, Jen J, Vasmatzis G, Patel SV, Maguire LJ, Baratz KH, Fautsch MP. Trinucleotide Repeat Expansion in the Transcription Factor 4 (TCF4) Gene Leads to Widespread mRNA Splicing Changes in Fuchs' Endothelial Corneal Dystrophy. Invest Ophthalmol Vis Sci. 2017 Jan 1;58(1):343-352. doi: 10.1167/iovs.16-20900. PMID: 28118661; PMCID: PMC5270622.

Wilburn B, Rudnicki DD, Zhao J, Weitz TM, Cheng Y, Gu X, Greiner E, Park CS, Wang N, Sopher BL, La Spada AR, Osmand A, Margolis RL, Sun YE, Yang XW. An antisense CAG repeat transcript at JPH3 locus mediates expanded polyglutamine protein toxicity in Huntington's disease-like 2 mice. Neuron. 2011 May 12;70(3):427-40. doi: 10.1016/j.neuron.2011.03.021. PMID: 21555070; PMCID: PMC3107122.

Wu MM, Grabe M, Adams S, Tsien RY, Moore HP, Machen TE. Mechanisms of pH regulation in the regulated secretory pathway. J Biol Chem. 2001 Aug 31;276(35):33027-35. doi: 10.1074/jbc.M103917200. Epub 2001 Jun 11. PMID: 11402049.

X

Xi Z, Zinman L, Moreno D, Schymick J, Liang Y, Sato C, Zheng Y, Ghani M, Dib S, Keith J, Robertson J, Rogaeva E. Hypermethylation of the CpG island near the G4C2 repeat in ALS with a C9orf72 expansion. Am J Hum Genet. 2013 Jun 6;92(6):981-9. Doi: 10.1016/j.ajhg.2013.04.017. Epub 2013 May 23. PMID: 23731538; PMCID: PMC3675239.

Xi Z, Rainero I, Rubino E, Pinessi L, Bruni AC, Maletta RG, Nacmias B, Sorbi S, Galimberti D, Surace EI, Zheng Y, Moreno D, Sato C, Liang Y, Zhou Y, Robertson J, Zinman L, Tartaglia MC, St George-Hyslop P, Rogaeva E. Hypermethylation of the CpG-island near the C9orf72 G_4C_2-repeat expansion in FTLD patients. Hum Mol Genet. 2014 Nov 1;23(21):5630-7. Doi: 10.1093/hmg/ddu279. Epub 2014 Jun 6. PMID: 24908669.

Xi Z, van Blitterswijk M, Zhang M, McGoldrick P, McLean JR, Yunusova Y, Knock E, Moreno D, Sato C, McKeever PM, Schneider R, Keith J, Petrescu N, Fraser P, Tartaglia MC, Baker MC, Graff-Radford NR, Boylan KB, Dickson DW, Mackenzie IR, Rademakers R, Robertson J, Zinman L, Rogaeva E. Jump from pre-mutation to pathologic expansion in C9orf72. Am J Hum Genet. 2015 Jun 4;96(6):962-70. Doi: 10.1016/j.ajhg.2015.04.016. Epub 2015 May 21. PMID: 26004200; PMCID: PMC4457950.

Xi J, Wang X, Yue D, Dou T, Wu Q, Lu J, Liu Y, Yu W, Qiao K, Lin J, Luo S, Li J, Du A, Dong J, Chen Y, Luo L, Yang J, Niu Z, Liang Z, Zhao C, Lu J, Zhu W, Zhou Y. 5’ UTR CGG repeat expansion in GIPC1 is associated with oculopharyngodistal myopathy. Brain. 2021 Mar 3;144(2):601-614. Doi: 10.1093/brain/awaa426. PMID: 33374016.

Xiao JH, Davidson I, Matthes H, Garnier JM, Chambon P. Cloning, expression, and transcriptional properties of the human enhancer factor TEF-1. Cell. 1991 May 17;65(4):551-68. Doi: 10.1016/0092-8674(91)90088-g. PMID: 1851669.

Xiao S, MacNair L, McGoldrick P, McKeever PM, McLean JR, Zhang M, Keith J, Zinman L, Rogaeva E, Robertson J. Isoform-specific antibodies reveal distinct subcellular localizations of C9orf72 in amyotrophic lateral sclerosis. Ann Neurol. 2015 Oct;78(4):568-83. Doi: 10.1002/ana.24469. Epub 2015 Aug 29. PMID: 26174152.

Xu Z, Poidevin M, Li X, Li Y, Shu L, Nelson DL, Li H, Hales CM, Gearing M, Wingo TS, Jin P. Expanded GGGGCC repeat RNA associated with amyotrophic lateral sclerosis and frontotemporal dementia causes neurodegeneration. Proc Natl Acad Sci U S A. 2013 May 7;110(19):7778-83. Doi: 10.1073/pnas.1219643110. Epub 2013 Apr 3. PMID: 23553836; PMCID: PMC3651485.

Y

Yamaguchi N, Mano T, Ohtomo R, Ishiura H, Almansour MA, Mori H, Kanda J, Shirota Y, Taira K, Morikawa T, Ikemura M, Yanagi Y, Murayama S, Shimizu J, Sakurai Y, Tsuji S, Iwata A. An

Autopsy Case of Familial Neuronal Intranuclear Inclusion Disease with Dementia and Neuropathy. Intern Med. 2018 Dec 1;57(23):3459-3462. doi: 10.2169/internalmedicine.1141-18. Epub 2018 Aug 10. PMID: 30101925; PMCID: PMC6306544.

Yamakawa M, Ito D, Honda T, Kubo K, Noda M, Nakajima K, Suzuki N. Characterization of the dipeptide repeat protein in the molecular pathogenesis of c9FTD/ALS. Hum Mol Genet. 2015 Mar 15;24(6):1630-45. doi: 10.1093/hmg/ddu576. Epub 2014 Nov 14. PMID: 25398948.

Yang D, Abdallah A, Li Z, Lu Y, Almeida S, Gao FB. FTD/ALS-associated poly(GR) protein impairs the Notch pathway and is recruited by poly(GA) into cytoplasmic inclusions. Acta Neuropathol. 2015 Oct;130(4):525-35. doi: 10.1007/s00401-015-1448-6. Epub 2015 Jun 2. PMID: 26031661; PMCID: PMC4575383.

Yang M, Liang C, Swaminathan K, Herrlinger S, Lai F, Shiekhattar R, Chen JF. A C9ORF72/SMCR8-containing complex regulates ULK1 and plays a dual role in autophagy. Sci Adv. 2016 Sep 2;2(9):e1601167. doi: 10.1126/sciadv.1601167. PMID: 27617292; PMCID: PMC5010369.

Yang X, Zhang D, Li P, Niu J, Xu D, Guo X, Wang Z, Zhao Y, Ren H, Ling C, Wang Y, Shen J, Zhu Y, Wang D, Cui L, Chen L, Dai Y. Expansion of 5' UTR CGG repeat in RILPL1 is associated with oculopharyngodistal myopathy. 2021. doi: 10.1101/2021.09.18.21263669

Yeetong P, Pongpanich M, Srichomthong C, Assawapitaksakul A, Shotelersuk V, Tantirukdham N, Chunharas C, Suphapeetiporn K, Shotelersuk V. TTTCA repeat insertions in an intron of YEATS2 in benign adult familial myoclonic epilepsy type 4. Brain. 2019 Nov 1;142(11):3360-3366. doi: 10.1093/brain/awz267. PMID: 31539032.

Yu L, McPhee CK, Zheng L, Mardones GA, Rong Y, Peng J, Mi N, Zhao Y, Liu Z, Wan F, Hailey DW, Oorschot V, Klumperman J, Baehrecke EH, Lenardo MJ. Termination of autophagy and reformation of lysosomes regulated by mTOR. Nature. 2010 Jun 17;465(7300):942-6. doi: 10.1038/nature09076. Epub 2010 Jun 6. PMID: 20526321; PMCID: PMC2920749.

Yu J, Deng J, Guo X, Shan J, Luan X, Cao L, Zhao J, Yu M, Zhang W, Lv H, Xie Z, Meng L, Zheng Y, Zhao Y, Gang Q, Wang Q, Liu J, Zhu M, Zhou B, Li P, Liu Y, Wang Y, Yan C, Hong D, Yuan Y, Wang Z. The GGC repeat expansion in NOTCH2NLC is associated with oculopharyngodistal myopathy type 3. Brain. 2021 Jul 28;144(6):1819-1832. doi: 10.1093/brain/awab077. PMID: 33693509; PMCID: PMC8320266.

Z

Zamiri B, Reddy K, Macgregor RB Jr, Pearson CE. TMPyP4 porphyrin distorts RNA G-quadruplex structures of the disease-associated r(GGGGCC)n repeat of the C9orf72 gene and blocks interaction of RNA-binding proteins. J Biol Chem. 2014 Feb 21;289(8):4653-9. doi: 10.1074/jbc.C113.502336. Epub 2013 Dec 26. PMID: 24371143; PMCID: PMC3931028.

Zeitlin S, Liu JP, Chapman DL, Papaioannou VE, Efstratiadis A. Increased apoptosis and early embryonic lethality in mice nullizygous for the Huntington's disease gene homologue. Nat Genet. 1995 Oct;11(2):155-63. doi: 10.1038/ng1095-155. PMID: 7550343.

Zhang X, Li L, Chen S, Yang D, Wang Y, Zhang X, Wang Z, Le W. Rapamycin treatment augments motor neuron degeneration in SOD1(G93A) mouse model of amyotrophic lateral sclerosis. Autophagy. 2011 Apr;7(4):412-25. doi: 10.4161/auto.7.4.14541. Epub 2011 Apr 1. PMID: 21193837.

Zhang D, Iyer LM, He F, Aravind L. Discovery of Novel DENN Proteins: Implications for the Evolution of Eukaryotic Intracellular Membrane Structures and Human Disease. Front Genet. 2012 Dec 13;3:283. doi: 10.3389/fgene.2012.00283. PMID: 23248642; PMCID: PMC3521125.

Zhang YJ, Jansen-West K, Xu YF, Gendron TF, Bieniek KF, Lin WL, Sasaguri H, Caulfield T, Hubbard J, Daughrity L, Chew J, Belzil VV, Prudencio M, Stankowski JN, Castanedes-Casey M, Whitelaw E, Ash PE, DeTure M, Rademakers R, Boylan KB, Dickson DW, Petrucelli L. Aggregation-prone c9FTD/ALS poly(GA) RAN-translated proteins cause neurotoxicity by inducing ER stress. Acta Neuropathol. 2014 Oct;128(4):505-24. doi: 10.1007/s00401-014-1336-5. Epub 2014 Aug 31. PMID: 25173361; PMCID: PMC4159567.

Zhang K, Donnelly CJ, Haeusler AR, Grima JC, Machamer JB, Steinwald P, Daley EL, Miller SJ, Cunningham KM, Vidensky S, Gupta S, Thomas MA, Hong I, Chiu SL, Huganir RL, Ostrow LW, Matunis MJ, Wang J, Sattler R, Lloyd TE, Rothstein JD. The C9orf72 repeat expansion disrupts

nucleocytoplasmic transport. Nature. 2015 Sep 3;525(7567):56-61. doi: 10.1038/nature14973. Epub 2015 Aug 26. PMID: 26308891; PMCID: PMC4800742.

Zhang YJ, Gendron TF, Grima JC, Sasaguri H, Jansen-West K, Xu YF, Katzman RB, Gass J, Murray ME, Shinohara M, Lin WL, Garrett A, Stankowski JN, Daughrity L, Tong J, Perkerson EA, Yue M, Chew J, Castanedes-Casey M, Kurti A, Wang ZS, Liesinger AM, Baker JD, Jiang J, Lagier-Tourenne C, Edbauer D, Cleveland DW, Rademakers R, Boylan KB, Bu G, Link CD, Dickey CA, Rothstein JD, Dickson DW, Fryer JD, Petrucelli L. C9ORF72 poly(GA) aggregates sequester and impair HR23 and nucleocytoplasmic transport proteins. Nat Neurosci. 2016 May;19(5):668-677. doi: 10.1038/nn.4272. Epub 2016 Mar 21. PMID: 26998601; PMCID: PMC5138863.

Zhang Q, Vashisht AA, O'Rourke J, Corbel SY, Moran R, Romero A, Miraglia L, Zhang J, Durrant E, Schmedt C, Sampath SC, Sampath SC. The microprotein Minion controls cell fusion and muscle formation. Nat Commun. 2017a Jun 1;8:15664. doi: 10.1038/ncomms15664. PMID: 28569745; PMCID: PMC5461507.

Zhang M, Xi Z, Misquitta K, Sato C, Moreno D, Liang Y, Slow E, Rogaeva E, Tartaglia MC. C9orf72 and ATXN2 repeat expansions coexist in a family with ataxia, dementia, and parkinsonism. Mov Disord. 2017b Jan;32(1):158-162. doi: 10.1002/mds.26841. Epub 2016 Nov 7. PMID: 28124431.

Zhang YJ, Gendron TF, Ebbert MTW, O'Raw AD, Yue M, Jansen-West K, Zhang X, Prudencio M, Chew J, Cook CN, Daughrity LM, Tong J, Song Y, Pickles SR, Castanedes-Casey M, Kurti A, Rademakers R, Oskarsson B, Dickson DW, Hu W, Gitler AD, Fryer JD, Petrucelli L. Poly(GR) impairs protein translation and stress granule dynamics in C9orf72-associated frontotemporal dementia and amyotrophic lateral sclerosis. Nat Med. 2018 Aug;24(8):1136-1142. doi: 10.1038/s41591-018-0071-1. Epub 2018 Jun 25. PMID: 29942091; PMCID: PMC6520050.

Zhang K, Wang A, Zhong K, Qi S, Wei C, Shu X, Tu WY, Xu W, Xia C, Xiao Y, Chen A, Bai L, Zhang J, Luo B, Wang W, Shen C. UBQLN2-HSP70 axis reduces poly-Gly-Ala aggregates and alleviates behavioral defects in the C9ORF72 animal model. Neuron. 2021 Jun 16;109(12):1949-1962.e6. doi: 10.1016/j.neuron.2021.04.023. Epub 2021 May 14. PMID: 33991504.

Zhao XN, Usdin K. The Repeat Expansion Diseases: The dark side of DNA repair. DNA Repair (Amst). 2015a Aug;32:96-105. doi: 10.1016/j.dnarep.2015.04.019. Epub 2015 Apr 30. PMID: 26002199; PMCID: PMC4522390.

Zhao J, Liu J, Xiao J, Du J, Que C, Shi X, Liang W, Sun W, Zhang W, Lv H, Yuan Y, Wang Z. Clinical and muscle imaging findings in 14 mainland chinese patients with oculopharyngodistal myopathy. PLoS One. 2015b Jun 3;10(6):e0128629. doi: 10.1371/journal.pone.0128629. PMID: 26039504; PMCID: PMC4454561.

Zheng Z, Li A, Holmes BB, Marasa JC, Diamond MI. An N-terminal nuclear export signal regulates trafficking and aggregation of Huntingtin (Htt) protein exon 1. J Biol Chem. 2013 Mar 1;288(9):6063-71. doi: 10.1074/jbc.M112.413575. Epub 2013 Jan 14. PMID: 23319588; PMCID: PMC3585045.

Zhong S, Lian Y, Luo W, Luo R, Wu X, Ji J, Ji Y, Ding J, Wang X. Upstream open reading frame with NOTCH2NLC GGC expansion generates polyglycine aggregates and disrupts nucleocytoplasmic transport: implications for polyglycine diseases. Acta Neuropathol. 2021 Oct 25. doi: 10.1007/s00401-021-02375-3. Epub ahead of print. PMID: 34694469.

Zhou D, Palam LR, Jiang L, Narasimhan J, Staschke KA, Wek RC. Phosphorylation of eIF2 directs ATF5 translational control in response to diverse stress conditions. J Biol Chem. 2008 Mar 14;283(11):7064-73. doi: 10.1074/jbc.M708530200. Epub 2008 Jan 14. PMID: 18195013.

Zhou Q, Lehmer C, Michaelsen M, Mori K, Alterauge D, Baumjohann D, Schludi MH, Greiling J, Farny D, Flatley A, Feederle R, May S, Schreiber F, Arzberger T, Kuhm C, Klopstock T, Hermann A, Haass C, Edbauer D. Antibodies inhibit transmission and aggregation of *C9orf72* poly-GA dipeptide repeat proteins. EMBO Mol Med. 2017 May;9(5):687-702. doi: 10.15252/emmm.201607054. PMID: 28351931; PMCID: PMC5412769.

Zhu Q, Jiang J, Gendron TF, McAlonis-Downes M, Jiang L, Taylor A, Diaz Garcia S, Ghosh Dastidar S, Rodriguez MJ, King P, Zhang Y, La Spada AR, Xu H, Petrucelli L, Ravits J, Da Cruz S, Lagier-Tourenne C, Cleveland DW. Reduced C9ORF72 function exacerbates gain of toxicity from ALS/FTD-causing repeat expansion in C9orf72. Nat Neurosci. 2020 May;23(5):615-624. doi: 10.1038/s41593-020-0619-5. Epub 2020 Apr 13. PMID: 32284607; PMCID: PMC7384305.

Zu T, Gibbens B, Doty NS, Gomes-Pereira M, Huguet A, Stone MD, Margolis J, Peterson M, Markowski TW, Ingram MA, Nan Z, Forster C, Low WC, Schoser B, Somia NV, Clark HB, Schmechel S, Bitterman PB, Gourdon G, Swanson MS, Moseley M, Ranum LP. Non-ATG-initiated translation directed by microsatellite expansions. Proc Natl Acad Sci U S A. 2011 Jan 4;108(1):260-5. doi: 10.1073/pnas.1013343108. Epub 2010 Dec 20. PMID: 21173221; PMCID: PMC3017129.

Zu T, Liu Y, Bañez-Coronel M, Reid T, Pletnikova O, Lewis J, Miller TM, Harms MB, Falchook AE, Subramony SH, Ostrow LW, Rothstein JD, Troncoso JC, Ranum LP. RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. Proc Natl Acad Sci U S A. 2013 Dec 17;110(51):E4968-77. doi: 10.1073/pnas.1315438110. Epub 2013 Nov 18. PMID: 24248382; PMCID: PMC3870665.

Zu T, Cleary JD, Liu Y, Bañez-Coronel M, Bubenik JL, Ayhan F, Ashizawa T, Xia G, Clark HB, Yachnis AT, Swanson MS, Ranum LPW. RAN Translation Regulated by Muscleblind Proteins in Myotonic Dystrophy Type 2. Neuron. 2017 Sep 13;95(6):1292-1305.e5. doi: 10.1016/j.neuron.2017.08.039. PMID: 28910618; PMCID: PMC5951173.

Zu T, Guo S, Bardhi O, Ryskamp DA, Li J, Khoramian Tusi S, Engelbrecht A, Klippel K, Chakrabarty P, Nguyen L, Golde TE, Sonenberg N, Ranum LPW. Metformin inhibits RAN translation through PKR pathway and mitigates disease in *C9orf72* ALS/FTD mice. Proc Natl Acad Sci U S A. 2020 Aug 4;117(31):18591-18599. doi: 10.1073/pnas.2005748117. Epub 2020 Jul 20. PMID: 32690681; PMCID: PMC7414156.

Manon BOVIN

ETUDE DES MECANISMES MOLECULAIRES RESPONSABLES DE MALADIES DUES A DES EXPANSIONS DE REPETITIONS SITUEES DANS DES REGIONS DU GENOME DITES « NON-CODANTES »

Résumé

Plus de 50 maladies génétiques sont causées par des expansions de répétitions de nucléotides. Ma thèse a porté sur deux de ces pathologies : la maladie à inclusions intranucléaires neuronales (NIID) et la sclérose latérale amyotrophique (SLA). Ces deux maladies neurodégénératives sont dues respectivement à des expansions de répétitions CGG (NIID) ou GGGGCC (SLA). Nous avons montré que, bien que situées dans des régions décrites initialement comme non-codantes, ces répétitions CGG et GGGGCC sont traduites en protéines toxiques. L'expression de ces protéines en modèle cellulaire ou murin reproduisent les caractéristiques histopathologiques et la mort neuronale typique de ces maladies (Boivin et al., EMBO J. 2020 ; Boivin et al., Neuron 2021). Ce travail souligne ainsi l'existence d'un nouveau mécanisme pathologique où des répétitions de nucléotides situées dans des régions dites « non-codantes » sont néanmoins exprimées en protéines toxiques.

Mots clés : maladies génétiques ; expansion de répétitions ; maladies neurodégénératives.

Résumé en anglais

Over 50 human genetic diseases are caused by expansions of tri-, tetra-, penta- or hexa-nucleotide. My PhD work focused on two of these diseases, Neuronal Intranuclear Inclusion Disease (NIID) and Amyotrophic Lateral Sclerosis (ALS). These two neurodegenerative diseases are caused by expansions of CGG or GGGGCC repeats, respectively. Interestingly, we found that despite being embedded in genetic regions initially described as "non-coding", these repeats are nonetheless translated into toxic proteins. Expression of these proteins in cell or animal models recapitulates the histopathological features and neuronal cell death characteristic of these diseases (Boivin et al., EMBO J. 2020; Boivin et al., Neuron 2021). Overall, these results support the existence of a novel pathogenic mechanism in human genetic diseases where expanded repeats located in "non-coding" genetic regions are nevertheless translated into toxic proteins.

Keywords: Genetic diseases; neurodegeneration; Repeat expansions; Microsatellite diseases.