Université de Strasbourg

École doctorale Sciences de la vie et de la santé | ED 414 Université de Strasbourg

UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ

Institute of Genetics and Molecular and Cellular Biology (IGBMC)

THÈSE présentée par :

Perrine KRETZ

soutenue le : **10 Mars 2021**

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

Discipline/ Spécialité : Neurosciences

Analyses des causes génétiques du syndrome de micro-délétion du 16p11.2 et de l'impact de la protéine majeure de la voute (MVP) et de son interaction avec MAPK3 dans la physiologie cérébrale

THÈSE dirigée par :

Dr. Binnaz YALCIN	CR HDR, IGBMC, Université de Strasbourg

RAPPORTEURS :

Pr. Christophe PHILIPPE	PUPH, CHU Dijon,
Dr. Laurence GOUTEBROZE	DR, IFM, Paris

AUTRES MEMBRES DU JURY :

Pr. Claire GAVERIAUX-RUFF	PU, IGBMC, Illkirch
Pr. Pierre VEINANTE	PU,INCI, Strasbourg
Dr. Pierre BILLUART	CR, IPNP, Paris

Table of content

REMERCIEMENTS **6**

RESUME (IN FRENCH) **9**

DISPLAY ITEMS **13**

List of Figures **13**

List of Tables **15**

LIST OF ABBREVIATIONS **16**

1. INTRODUCTION **20**

1.1 Autism spectrum disorders **21**
1.1.1 Definition and history 21
1.1.2 Epidemiology and social impact 21
1.1.3 Spectrum, diagnosis and comorbidities 22
1.1.4 Sexual dimorphism 25
1.1.5 Neuroanatomy of autism 26
1.1.5.1 Anterior cingulate cortex 27
1.1.5.2 Corpus callosum 28
1.1.5.3 Functional connectivity 30
1.1.5.4 Synaptic plasticity 31
1.1.6 Genetic etiology 32
1.1.6.1 Monogenic forms 33
1.1.6.2 Polygenic risk 33
1.1.6.3 Copy Number Variants 34
1.1.6.4 Gene networks 34

1.2 Autism-associated 16p11.2 microdeletion syndrome **37**
1.2.1 Identification of the locus 37
1.2.2 Genomic architecture 38
1.2.3 Symptoms of 16p11.2 microdeletion carriers 39
1.2.3.1 Body weight 39
1.2.3.2 Neuropsychiatric impairments 40
1.2.3.3 Neuroanatomical defects 42
1.2.4 Molecular consequences 45
1.2.5 Mouse models of the 16p11.2 microdeletion syndrome 46
1.2.5.1 *Slx1b-Sept1* deletion model 48
1.2.5.2 *Coro1a-Spn* deletion model 49
1.2.5.3 *Sult1a-Spn* deletion model 50
1.2.5.4 Comparative study between mouse models 51
1.2.6 Similarities between 16p11.2 mouse models and human microdeletion carriers 53

1.3 Mitogen-activated protein kinase 3 (MAPK3) **54**
1.3.1 Mitogen-activated protein kinases (MAPKs) and origins 54
1.3.2 Roles of MAPK1 and MAPK3 55
1.3.2.1 *Mapk1*-deficient mice 55
1.3.2.2 *Mapk3*-deficient mice 55
1.3.2.3 Double *Mapk1::Mapk3*-deficient mice 56
1.3.3 ERK1/2 signaling pathway 56

1.3.4 Relevance of ERK signaling in human diseases 58
1.3.4.1 Rasopathies and autism 59
1.3.4.2 Disease models for ERK pathway 59
1.3.4.3 Modulation of ERK signaling as a therapeutic target of autism 60

1.4 The Major Vault Protein 61
1.4.1 History of vault research 61
1.4.2 Vault evolutionary biology 63
1.4.3 MVP/Vault structure and composition 65
1.4.3.1 General structure of the vault 65
1.4.3.2 Description of the vault-shell element: MVP structure 67
1.4.3.3 Description of the other vault components: vPARP, TEP1 and vRNA 68
1.4.4 MVP/Vault life cycle 69
1.4.4.1 MVP transcription 69
1.4.4.2 MVP translation and vault-shell assembly 69
1.4.4.3 Vault components integration 71
1.4.4.4 MVP/Vault degradation 71
1.4.5 MVP/Vault function 72
1.4.5.1 MVP/vault expression pattern and cellular distribution 72
1.4.5.2 MVP/vault in multi-drug resistance 72
1.4.5.3 Vault act as a scaffold protein for MAPKinase pathway 73
1.4.5.4 MVP/Vault function in the nervous system 74

2. THESIS RATIONALE 77

3. SCIENTIFIC CONTRIBUTION 79

4. MAIN PUBLICATION 80

4.1 Abstract 81

4.2 Introduction 82

4.3 Results 83
4.3.1 Mouse genetic studies unravel the implications of multiple drivers in NeuroAnatomical Phenotypes at the 16p11.2 locus 83
4.3.2 The major vault protein is expressed in the limbic system and is implicated in neuronal morphology 86
4.3.3 The double deletion of Mvp and Mapk3 rescues neuroanatomical phenotypes and alters behavioral performances in mice 88

4.4 Discussion 90

4.5 Figures and legends 95

4.6 Tables and legends 120

4.7 Dataset legends 129

4.8 Supplementary Results 137
4.8.1 Male NAP genes implicated in two or more brain categories 137
4.8.2 Male NAP genes implicated in one brain category 138
4.8.3 Female NAP genes involved in two or more brain categories 139
4.8.4 Female NAP genes involved in one brain category 140

5. MATERIAL AND METHODS 141

5.1 Study samples and mouse knockout constructions ... 141
5.1.1 Single-gene knockout ... 141
5.1.2 *Mvp*::*Mapk3* double knockout model ... 143
5.1.3 *16p11.2*$^{+/Del}$ mouse models ... 144

5.2 Animal welfare ... 144

5.3 Genotyping and sexing of study samples ... 145

5.4 Brain histological screen ... 146
5.4.1 Sample processing ... 146
5.4.2 Histomorphometric analyses and quality control ... 148
5.4.3 Statistics ... 149

5.5 Follow-up histology studies of MVP/vault ... 149
5.5.1 Embryonic stage ... 149
5.5.2 Sagittal orientation ... 150
5.5.3 Determination of layers in somatosensory and motor cortices ... 151
5.5.4 Measurement of cellular features ... 151

5.6 Evaluation of MVP/vault transcript level ... 152
5.6.1 Tissue dissection ... 152
5.6.2 RNA extraction ... 152
5.6.3 Qualitative RT-PCR ... 153
5.6.4 Quantitative RT-PCR ... 153

5.7 Immunohistofluorescence ... 154
5.7.1 Sample processing ... 154
5.7.2 MVP/Vault immunostaining ... 154
5.7.3 phospho-ERK immunohistofluorescence ... 155

5.8 Western Blot Analysis ... 155

5.9 Immunocytofluorescence ... 156
5.9.1 Primary neuronal cultures ... 156
5.9.2 Immunocytostaining ... 157
5.9.3 Cell morphological assessment ... 157

5.10 RNA sequencing ... 157

5.11 Golgi staining ... 158

5.12 Electrophysiology ... 159
5.12.1 Preparation of acute slices ... 159
5.12.2 Electrophysiological recordings ... 159

5.13 Electronic microscopy ... 160

5.14 Behavioral Analysis ... 160
5.14.1 Behavioral pipeline ... 161
5.14.2 Open field ... 161
5.14.3 Elevated plus maze ... 162
5.14.4 Fear conditioning ... 162
5.14.5 Forced swim ... 163
5.14.6 Tail suspension ... 163
5.14.7 Pentylenetetrazol ... 163
5.14.8 Grip strength ... 163
5.14.9 Rotarod ... 163
5.14.10 Sucrose preference ... 164

5.14.11 Y-maze 164
5.14.12 Novel object recognition 164
5.14.13 Three-chamber 165
5.14.14 Social interaction 165
5.14.15 Marbles burying 165
5.14.16 Circadian activity 166
5.14.17 Acoustic startle response and pre-pulse inhibition 166
5.14.18 Statistics for behavior study 167

5.15 Statistics 168

5.16 Summary of key resources table 169

6. CONCLUDING REMARKS 170

7. REFERENCES 171

Remerciements

Je souhaitais tout d'abord remercier les membres de mon jury. Merci aux Pr. Christophe PHILIPPE et Dr. Laurence GOUTEBROZE d'avoir accepté d'être rapporteurs de mon travail doctoral, ainsi que Pr. Claire GAVERIAUX-RUFF, Pr. Pierre VEINANTE et Dr. Pierre BILLUART de les accompagner dans son évaluation. C'est pour moi un grand honneur de vous présenter mes recherches et de les soumettre à votre appréciation.

Je suis également d'une grande gratitude envers ma directrice de thèse Dr. Binnaz YALCIN. Merci pour le potentiel que tu as vu en moi il y a cinq ans, la confiance et le soutien que tu m'as apportés, même dans les moments les plus incertains, ainsi que la merveilleuse opportunité que d'avoir pu travailler à tes cotés sur un sujet aussi passionnant. Cette quête n'aurait pas vu le jour sans ta lucidité, ton expertise et ton implication.

J'ai eu l'immense joie d'évoluer au sein d'une équipe bienveillante et efficace, qui m'ont enrichie tout au long de mon parcours. Dr Meghna KANNAN, Dr Stephan COLLINS, Christel WAGNER et Marie-Christine FISCHER, la formation et le travail à vos côtés ont été un enrichissement permanent et un bonheur que j'espère partager. Merci également aux stagiaires qui ont participé à mon projet : Clara BONNET, Leo GAGLIARDI, Maxence MILHAU, Jules ROUSSEY, Paula HAHN, Emeline AGUILAR et Marion GABORIT.

Je tenais également à remercier les membres de l'équipe du Dr Juliette GODIN, en compagnie de qui cela a été un plaisir de travailler au quotidien, ainsi que les différentes équipes et services administratif de l'IGBMC avec lesquels j'ai pu échanger.

Un grand merci à l'équipe du GAD de Dijon, qui m'ont accueillie au sein de leurs murs lors de la rédaction de ma thèse, et tout particulièrement à Charlotte MONTILLOT, Sylvie NGUYEN et Vincent LHUSSIEZ pour nos riches échanges.

Je souhaitais également remercier tous les collaborateurs et chercheurs avec qui j'ai eu l'honneur d'échanger au cours de ma formation doctorale que ce soit à propos d'un projet particulier ou lors de conférence.

Merci à Maxime, Xavière, Marion, Christophe, Mar et tous les autres doctorants de l'IGBMC avec qui j'ai pu échanger tant sur le plan scientifique qu'humain au cours de ces années de thèse. Le partage d'expérience et de moments chaleureux ont été un vrai moteur pour la construction de mon projet.

J'ai une grande fierté également d'avoir été entourée de personnes magnifiques au sein de Doctoneuro, qu'il s'agisse des membres du bureau fondateur qui m'ont suivi dans ce projet ambitieux de cohésion et de propagation des savoirs, mais aussi des nouveaux membres qui nous ont rejoint, en nous donnant leur énergie et leur motivation afin de concrétiser nos projets. Cela a été une part intégrante de la conception que je me suis faite de mon travail en tant que scientifique, et il a pu être mené à bien de par votre implication. Bisous-Caresses sur vos cerveaux.

Je suis heureuse d'avoir pu évoluée tout au long de mon parcours universitaire au sein de l'Université de Strasbourg. Cette grande dame a su me fournir toutes les ressources nécessaires à l'épanouissement de ma vie étudiante au cours des dix dernières années et une dynamique d'entreprendre et de confiance qui m'a entouré tout du long. Merci également au Neuropole et à Neurex pour l'organisation mise en place afin de coordonner les différentes équipes de recherche et ainsi créer un espace de d'échange et de valorisation des neurosciences au niveau local et européen.

Il me tient aussi de remercier l'ensemble de personnes travaillant à l'école doctorale Vie et Santé, ainsi qu'au Collège doctoral, qui ont été un support compréhensif et attentif dans l'ensemble de mes démarches administratives et pédagogiques.

Vient enfin le moment d'exprimer ma gratitude envers les membres de ma famille, qui depuis ma naissance m'ont donné les clefs afin de pouvoir entreprendre la vie qui est la mienne. Merci à toi Maman pour ton éducation, ta confiance et ton soutien infaillible. Merci à toi Guillaume pour nos moments partagés, ta présence et le merveilleux frère que tu es. Merci également au reste de mon entourage, qui a suivi avec curiosité et bonté mon avancée dans la belle aventure qu'a été ma thèse.

Une pensée toute particulière également pour mon père et mamie Rose, qui malheureusement ne pourront pas assister à ce moment si important pour moi. Vous êtes avec moi dans les gènes et avec toutes les valeurs que vous m'avez transmises.

Merci également à mes sœurs de cœur, mes amours, la Family, la team Cocoben, les BFF, l'Amicale des Sciences, mes amis proches et tous ceux qui m'ont donné de l'espoir, de l'affection, du réconfort et de la motivation. Avoir un tel réseau de personnes fidèles, attentionnées et compréhensives a été un atout pour ma santé mentale, la valorisation de mon être et mon accomplissement personnel.

Un grand merci également à mes thérapeutes, aux verres de gin tonic après le labo, aux soirées endiablées jusqu'aux aurores, au bouton Snooze sur le réveil, aux pauses clopes et à tous ces petits vices du quotidien qui m'ont permis de relâcher la pression de temps à autre.

Pour finir, j'adresse mes profonds respects à tous ces penseurs et ces scientifiques qui m'ont un jour inspirée, que ce soit par leur travaux, leur philosophie ou leur impact sur le monde. Je n'aurais certainement pas suivi cette voie sans eux : des profils atypiques et multiples qui ont eu la merveilleuse capacité d'avancer de front sur plusieurs plans de la connaissance afin d'élargir les perspectives de la Science et sa place au sein de la société : Leonard de Vinci, Jules Hoffman, Leonard Rome, Marie Curie, Etienne Klein, Idriss Aberkane, Hildegarde Von Bingen, Averroès, Jean-Louis Mandel, António Damásio, Patrick Burensteinas, Spinoza,…

Ce n'est pas la folie qui est capable de bouleverser le monde, c'est la conscience. – Baruch Spinoza –

Dans la vie rien n'est à craindre, tout est à comprendre. – Marie Curie -

Sachez vous éloigner car, lorsque vous reviendrez à votre travail, votre jugement sera plus sûr – Leonard de Vinci –

Vive les Scientifiques, et Vive la Science.

Résumé (in French)

Les troubles du spectre de l'autisme (TSA) sont une maladie du développement du cerveau qui peut causer de sévères handicaps pour les patients ainsi que pour leur famille.

Les symptômes se caractérisent par un déficit de l'apprentissage et de la communication verbale et non-verbale ainsi que la présence de comportements stéréotypés et persévératifs. Cette pathologie touche près d'un enfant sur 150 avec une prévalence quatre fois plus importante chez les sujets de sexe masculin que féminin. Outre la variation des degrés de sévérité des symptômes cardinaux de la maladie, les patients atteints présentent dans la plupart des cas des symptômes associés qui peuvent être très variables, tels que l'épilepsie, la déficience intellectuelle, des troubles du sommeil et de l'activité ou sensoriels. Plusieurs mécanismes cérébraux ont été mis en évidence comme pouvant être à l'origine de la maladie, tels que le transport axonal, l'homéostasie neuronale ou la plasticité synaptique.

Les causes de la maladie sont également diverses et de nombreux facteurs de risques ont pu être identifiés, notamment l'exposition à certains stress ou produits chimiques au court de la grossesse (cause environnementales), ainsi que des mutations ponctuelles ou des réarrangements du génome (causes génétiques).

Afin de diagnostiquer certaines de ces causes, des études à l'échelle génomique ont permis d'identifier que la variation du nombre de copie d'un groupe de 30 gènes présents sur le locus 16p11.2 pouvait entrainer la manifestation d'un syndrome autistique. Les symptômes associés avec la délétion et la duplication de ces gènes engendrent des troubles cognitifs (déficience intellectuelle ou schizophrénie), métaboliques (obésité ou maigreur) et neuroanatomiques (macro- ou microcéphalie, respectivement). Ces symptômes résultent d'une convergence fonctionnelle déficitaire de l'ensemble de la région, cependant, l'impact individuel de chacun des gènes du locus n'a pas encore été clairement établi.

C'est pourquoi mon projet de recherche s'est orienté sur la résolution de ces interactions au niveau neuroanatomique, afin d'identifier quel(s) pourraient être les éléments majeurs à l'origine de ces déficits et d'en comprendre les mécanismes.

Mon laboratoire d'accueil a développé au cours des dernières années un protocole d'analyse histologique du cerveau murin, qui permet d'analyser de manière quantitative les variations neuroanatomiques de 20 régions cérébrales, au niveau morphologique et cellulaire[1]. Cette méthode nous a permis d'établir l'impact de la délétion (homozygote et/ou hétérozygote) de 20 des gènes de l'intervalle du 16p11.2. Parmi eux, le gène codant pour la protéine majeure de la voute (MVP) a été identifié comme le principal acteur à l'origine des phénotypes de taille cérébrale associés au syndrome du 16p11.2.

Cette protéine s'auto agrège afin de former le plus grand complexe ribo-nucléoprotéique décrit jusqu'à présent au sein d'une cellule : la voute. Elle possède une taille de 13MDa et est composée de 78 copies de MVP, nécessaire à sa formation, ainsi que de la protéine télomérase 1, de 8 copies de Poly-ADP-ribose Polymérase 4 et d'ARN non codant qui lui sont spécifiques. Malgré plus de 30 ans de recherche à son sujet, la fonction physiologique de la voute n'a toujours pas été clairement identifiée. Des publications antérieures ont suggéré un rôle dans la pharmacorésistance de certains cancers[2,3], ou en tant que plateforme d'interactions, notamment pour la voie de signalisation des MAPKinases[4].

Notre criblage neuroanatomique a permis de mettre en évidence que la délétion du gène *Mvp* entraine chez le mâle une microcéphalie caractérisée par une diminution de la taille du corps calleux, de l'hippocampe et des cortex cingulaire, somatosensoriel, moteur et piriforme. Une étude plus approfondie des sections de cerveaux a permis de déterminer que cette diminution était liée à une réduction de la taille des neurones dans ces différentes structures. Cela a également été confirmé dans des cultures de neurones hippocampaux, pour lesquels on observe une diminution de la taille du soma, ainsi que du cône de croissance, une structure cruciale pour la perception de l'environnement cellulaire au moment du développement neuronal. Ces défauts neuroanatomiques sont présents uniquement chez le mâle, et aucun phénotype n'est visible chez la femelle. La même étude réalisée à différents âges a permis de déterminer que ces défauts n'apparaissaient qu'après la naissance.

Jusqu'alors, aucune cartographie du patron d'expression de *Mvp* au sein des différentes structures cérébrales n'était disponible chez la souris. Chez l'adulte, j'ai pu observer que l'expression de transcrits de *Mvp* est ubiquitaire, mais cependant plus

faible au sein du cerveau (cortex, hippocampe, striatum) que dans le cervelet ou les organes périphériques (foie, rein, poumon, cœur). Une analyse d'histo-immunofluorescence sur des coupes de cerveaux a permis de confirmer une expression neuronale de MVP dans l'ensemble des structures cérébrales. Cependant, j'ai pu observer que certaines d'entre elles présentent des niveaux d'expression bien supérieurs aux autres structures, notamment dans la couche externe du cortex piriforme, plusieurs noyaux du thalamus et de l'hypothalamus, dont les corps mamillaires, le raphé et le noyau du nerf vague. Les faisceaux de fibres de la substance blanche sont également marqués. Ce modèle d'expression caractéristique est observable aussi bien chez le mâle que chez la femelle.

Afin d'identifier si la délétion de *Mvp* pouvait entrainer des désordres de la connectivité synaptique, j'ai également réalisé une analyse des épines dendritiques grâce à une coloration de Golgi-Cox. Ces données montrent une diminution du nombre d'épines dendritiques sur les neurones pyramidaux. En collaboration avec un laboratoire expert en électrophysiologie, nous avons effectué des mesures de l'activité fonctionnelle des neurones du cortex cingulaire. La diminution des mesures de l'amplitude des potentiels post-synaptiques excitateurs miniatures (mEPSCs) a permis de mettre en évidence le rôle de la voute dans les processus de communication neuronale *in vivo*.

Après la caractérisation des défauts observés chez les souris mutantes, la deuxième partie de mon projet a été d'identifier la voie de signalisation cellulaire dans laquelle MVP est impliquée.

Le séquençage d'ARN effectué sur des échantillons de cortex cingulaire n'a révélé aucune modification du transcriptome entre les souris $Mvp^{+/+}$ et $Mvp^{-/-}$, signifiant que les phénotypes de tailles cellulaires n'étaient pas médiés par des modifications drastiques de l'expression des gènes.

Etant donné la structure génomique complexe étudiée dans le cas du syndrome du 16p11.2, nous avons également effectué des investigations afin de mieux comprendre les interactions de *Mvp* avec d'autres partenaires du locus, notamment *Mapk3* qui a été précédemment identifiés chez le poisson zèbre comme pouvant mener à des défauts neuroanatomiques[5]. Chez la souris, l'analyse neuroanatomique n'a pas réussi à démontrer un impact de la délétion individuelle de *Mapk3* sur la morphologie cérébrale. J'ai donc produits des souris double-hétérozygotes pour ces gènes

combinés afin de voir si leurs interactions pouvaient avoir un effet épistatique. La combinaison *Mvp*$^{+/-}$ ::*Mapk3*$^{+/-}$ permet de remédier aux phénotypes neuroanatomiques observés chez les souris *Mvp*$^{+/-}$ simple. De plus, les souris *Mvp*$^{+/-}$ ::*Mapk3*$^{+/-}$ présentent également une altération des comportements anxieux, ainsi qu'une résistance lors des crises épileptiques, qui n'avait pas été reportés chez les mutants simples. Cela démontre bien la présence d'interactions entre *Mvp* et *Mapk3* ayant un impact sur l'homéostasie neuronale.

Mon travail de thèse est une démonstration inédite de déficits liés à une réduction de la quantité de MVP. Cette première caractérisation de son patron d'expression chez la souris ainsi que l'identification des structures cérébrales spécifiquement atteintes par sa délétion ont permis de poser les premiers jalons menant vers la compréhension des mécanismes fondamentaux de la voute. La présence d'un dimorphisme franc entre les deux sexes ainsi que son action dans la maintenance de l'homéostasie neuronale notamment via la voie des MAPKinases ouvrent la voie à de nouvelles perspectives en adéquation avec les caractéristiques des syndromes autistiques, ce qui présage également de nouvelles avancées dans ce domaine.

Display items

List of Figures

Figure 1 Schematic representation of the developmental stages the human brain. 20
Figure 2 Manifestations of the cardinal symptoms in ASDs. 23
Figure 3 Prevalence of ASD-related comorbidities. 24
Figure 4 Empathization-systematization score in autism. 26
Figure 5 Abnormal gyrification in the anterior cingulate cortex of ASD patients. 28
Figure 6 Changes in the size of the corpus callosum in young ASD patients. 29
Figure 7 Whole-brain hyper-connectivity in children with ASD. 30
Figure 8 Structure of the pre- and post-synaptic compartments. 32
Figure 9 Models of ASD-risk variants. 33
Figure 10 Models of gene network action in ASDs. 34
Figure 11 Main functional networks in autism. 35
Figure 12 Key convergence in cellular functions among ASD genes. 36
Figure 13 Pedigrees presenting CNVs at the 16p11.2 locus. 37
Figure 14 Localization of the 16p11.2 BP4-BP5 locus in the human genome. 38
Figure 15 Distribution of body mass index in 16p11.2 deletion carriers. 40
Figure 16 Neuropsychiatric disorders in children with 16p11.2 deletion. 41
Figure 17 Distribution of IQ scores in deletion carriers. 41
Figure 18 Head circumference in 16p11.2 deletion carriers. 42
Figure 19 Brain MRI measurements of 16p11.2 deletion carriers. 43
Figure 20 MRI pictures of 16p11.2 deletion carriers. 43
Figure 21 Voxel-based differences of 16p11.2 deletion carriers. 44
Figure 22 Expression profile of genes within the 16p11.2 proximal region. 45
Figure 23 Enrichment of differentially expressed genes in deletion carriers. 46
Figure 24 Mouse models for the 16p11.2 rearrangements. 47
Figure 25 Brain structure in mice deleted for 16p11.2 *Slx1b-Sept1*. 48
Figure 26 Neuroanatomical abnormalities in 16p11.2 Coro1a-Spn deletion model. 50
Figure 27 The origins of human mitogen-activated protein kinases. 54
Figure 28 Representation of the Ras-Raf-MEK-ERK signal transduction cascade. 57
Figure 29 Timeline of major publications in vault biology. 62
Figure 30 MVP/Vault evolutionary predictions. 64

Figure 31 Overall structure of the vault. 65
Figure 32 Model showing the opening of a vault into flower-like structures. 66
Figure 33 Stereoscopic ribbon drawing of the overall fold of an MVP monomer. 67
Figure 34 A vertically sliced section of a half-vault structure. 69
Figure 35 Polyribosomes contributes to vault assembly from *Mvp* mRNA. 70
Figure 36 MVP cytoplasmic expression and association with microtubules. 72
Figure 37 SHP-2 regulation of MVP phosphorylation in EGF-mediated signalling. 74
Figure 38 MVP expression in rat neurons. 75
Figure 39 Mouse neuroanatomical screen at the 16p11.2 autism-associated locus. 95
Figure 40 Involvement of the *Mvp* gene in male brain morphology. 96
Figure 41 Analysis of double deletion of *Mvp* and *Mapk3* genes. 98
Figure 42 Standardized workflow for neurophenotyping. 99
Figure 43 Heat map of gene association with NAPs in males. 101
Figure 44 Heat map of gene association with NAPs in females. 102
Figure 45 Comparative analysis of *Mvp*, *Mapk3* and *Kctd13* mice. 103
Figure 46 *Mvp* expression profiling. 105
Figure 47 *Mvp* knockout model validation and viability screen. 107
Figure 48 Impact of *Mvp* deletion on neuromorphology in females. 109
Figure 49 Impact of *Mvp* deletion on neuromorphology in males. 111
Figure 50 Analysis of *Mvp* loss in the cingulate gyrus. 113
Figure 51 Behavioral characterization of *Mvp*::*Mapk3* double heterozygous. 114
Figure 52 Behavioral studies of *Mvp* homozygous mice. 116
Figure 53 Behavioral studies of *Mapk3* heterozygous mice. 118
Figure 54 Mutant allele design. 142
Figure 55 Double knockout breeding strategy. 144
Figure 56 Coronal section of interest at E18.5. 150
Figure 57 Cortical layer determination. 151
Figure 58 Acoustic startle response and pre-pulse inhibition pipeline design. 167

List of Tables

Table 1 Comparative analysis between 16p11.2 microdeletion mouse models. 52
Table 2 Dysregulation of ERK activity in Rasopathies. 59
Table 3 Description of mouse models used in the study. 120
Table 4 List of 67 coronal neuroanatomical coronal measurements. 121
Table 5 MVP-positive brain structures and link to limbic system 123
Table 6 Expression scores of MVP in murine brain 125
Table 7 List of 42 coronal neuroanatomical measurements in E18.5 embryos. 127
Table 8 List of 40 sagittal neuroanatomical measurements in postnatal mice. 128
Table 9 Genotyping and sexing primers. 146
Table 10 Sequences of RT-PCR primers. 153
Table 11 Sequences of qRT-PCR primers and probes 154
Table 12 Behavioral pipeline. 161

List of abbreviations

4V: fourth ventricle
7N: Facial nerve nucleus
10N: Vagus nerve nucleus
ABIDE: Autism Brain Imaging Data Exchange
aca: anterior commissure - anterior part
ACC: Anterior cingulate cortex
AGRE: Autism Genetic Resource Exchange
AL: Axonal length
AP: Area postrema
Arc: Arcuate hypothalamic nucleus
ASDs: Autism Spectrum Disorders
AV: arbor vitae
BMI: Body Mass Index
BP: Breakpoint
Cc: corpus callosum
cDNA: complementary DNA
CDC: Centres for Disease Control and Prevention
CE: Critical exon
Cg: Cingulate gyrus
CMT: Central medial thalamic nucleus
CN: Cuneate nucleus
CNR: National research council (Italy)
CNS: Central nervous system
CNV: Copy Number Variant
CP: Choroid plexus
CPu: Caudate Putamen
CRB: Cerebellum
CryoEM: Cryo-electronic microscopy
CTX: Cortex
CTXsp: cortical subplate
DBB: Diagonal Band of Broca
Dk: Darkschewitsch nucleus
DMH: Dorsal medial hypothalamic nucleus
DMT: Dorsal medial thalamic nucleus
DNA: Deoxyribonucleic acid
DSM: Diagnostic and Statistical Manual of Mental Disorders
DTI: Diffusion-Tensor Imaging
DUSP: dual-specificity phosphatase
EGF: Epidermal growth factor
EM: electronic microscopy
EPSC: Excitatory post-synaptic current
ERK: extracellular signal- regulated kinase
ESESC: extreme state of the empathization-systematization continuum
f: fornix
FDR: False Discovery Rate
fMRI: functional magnetic resonance imaging
FSQI: full-scale intelligence quotient

GC: Growth cone
Gnas: Guanine nucleotide-binding protein G - subunit alpha
GO: Gene Ontology
GP: Globus Pallidus
GTP: Guanosine triphosphate
GWAS: Genome Wide Association Study
HAH: High Authority of Health (France)
HC: Head circumference
HPC: Hippocampal formation
Hvg: Human vault RNA gene
HYPO: Hypothalamus
ICD: International Classification of Diseases
ICF: Immunocytofluorescence
ICS: Institut Clinique de la souris – mouse clinical institute
ID: Intellectual disability
IHF: Immunohistofluorescence
IMPC: International Mouse Phenotyping Consortium
IQ: Intellectual Quotient
JNK: c-Jun N-terminal Kinase
kDa: kilo Dalton
LCL: lymphoblastoid cell line
LCR: Low Copy Repeat
Lith: Lithoid nucleus
LPO: Lateral preoptic area
LUN: Lung
MAPKinases: mitogen-activated protein kinase
MBP: Mouse Biology Program
MDR: Multi-Drug Resistance
MED: Medulla
mEPSCs: miniature excitatory post-synaptic currents
mESC: mouse embryonic stem cell
MidB: Midbrain
MS: Medial septal nucleus
MRI: magnetic resonance imaging
mTOR: mammalian Target Of Rapamycin
MVe: Medial vestibular nucleus
MVP: Major Vault Protein
NAHR: Non-Allelic Homologous Recombination
NeuN: Neuronal Nuclei
NGF: Nerve growth factor
NOR: novel-object recognition
ODD/CD: oppositional defiant disorder/conduct disorder
OLF: Olfactory area
OMIM: Online Mendelian Inheritance in Man
PAG: Periaqueductal gray
PAL: Pallidum
PC: Purkinje cell
PCR: Polymerase Chain Reaction
Pir: Piriform cortex
PTP: protein tyrosine phosphatase

PVH: Paraventricular hypothalamic nucleus
qRT-PCR: quantitative RT-PCR
Raf: rapidly accelerated fibrosarcoma
Ras: Rat sarcoma
RM: Retromammillary nucleus
RNA: ribonucleic acid
RNP: ribonucleoprotein
RSGc: Retrosplenial cortex
RTK: Receptor Tyrosine Kinase
RT-PCR: Reverse Transcription Polymerase Chain Reaction
S2: secondary somatosensory cortex
Sb: Subiculum
SD: Standard Deviation
SFO: Subfornical organ
SH2: Src Homology 2
SHANK3: SH3 and multiple ankyrin repeat domains 3
SHP-2: Src homology region 2 domain-containing phosphatase-2
Sol: Solitary nucleus
SPRY: sprout
STAT: signal transducer and activator of transcription
STR: Striatum
TEP1: Telomerase protein 1
THL: Thalamus
TIGM: Texas A&M Institute for Genomic Medicine#
TS: Triangular septal nucleus
vPARP: Poly-ADP Polymerase 4
WSI: Welcome Sanger Institute
Xi: Xiphoid thalamic nucleus
ZI: Zona incerta

16p11.2 locus genes:

SPN: sialophorin
QPRT: quinolinate phosphoribosyltransferase
C16ORF54: chromosome 16 open reading frame 54
ZG16: zymogen granule protein 16
KIF22: kinesin family member 22
MAZ: MYC-associated zinc finger protein (purine-binding transcription factor)
PRRT2: proline-rich transmembrane protein 2
PAGR1: PAXIP1 associated glutamate rich protein 1A
MVP: major vault protein
CDIPT: CDP-diacylglycerol--inositol 3-phosphatidyltransferase
SEZ6L2: seizure related 6 homolog like 2
ASPHD1: aspartate beta-hydroxylase domain containing 1
KCTD13: potassium channel tetramerisation domain containing 13
TMEM219: transmembrane protein 219
TAOK2: TAO kinase 2
HIRIP3: HIRA interacting protein 3
INO80E: INO80 complex subunit E
DOC2A: double C2, alpha
C16ORF92: chromosome 16 open reading frame 92
FAM57B: family with sequence similarity 57, member B
ALDOA: aldolase A, fructose-bisphosphate
PPP4C: protein phosphatase 4, catalytic subunit
TBX6: T-box 6
YPEL3: yippee like 3
GDPD3: glycerophosphodiester phosphodiesterase domain containing 3
MAPK3: Mitogen-activated protein kinase 3
CORO1A: Coronin, actin-binding protein 1A

1. Introduction

The central nervous system (CNS) is the conductor of many essential mechanisms in our body and is governed by a system of extreme complexity. The development of the central nervous system can be however summarized into three main stages[6]. During the first half of gestation, stem cells proliferate and differentiate into young neurons or glial cells. During the second stage from mid- to late-gestation, neurons migrate from their place of origin to different brain areas where they begin to make connections with other neurons, establishing neuronal connections. The third phase involves the final organization of the brain associated with synaptogenesis and myelination up to adulthood (**Figure 1**).

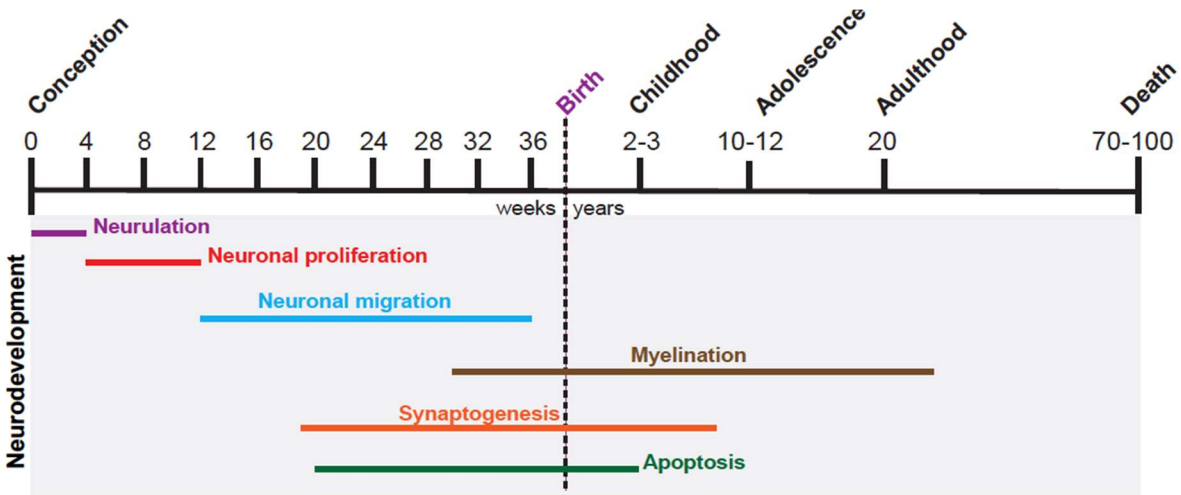


Figure 1 Schematic representation of the developmental stages the human brain. Adapted from *Chiurazzi* and Pirozzi 2016.

In order to understand these fundamental neurobiological processes during CNS development, it is often virtuous to explore the disturbances that are causing severe neurodevelopmental disorders such as autism spectrum disorders. These devastating disorders affect 1 to 3% of the population worldwide. They are a source of major distress for the patients and their entourage and, from a societal point of view, represent 5 to 10% of public health cost[7].

1.1 Autism spectrum disorders

1.1.1 Definition and history

Autism Spectrum Disorders (ASDs) constitute a clinically heterogeneous class of neurodevelopmental disorders characterized by a core set of impairments including social and interpersonal disabilities. The term "autism" derived from a Greek word meaning “self”, was first utilized by the Swiss psychiatrist Eugen Bleuler (1857-1939) to describe limitations in patients diagnosed with schizophrenia in establishing social relationships. Later on, the Austrian psychiatrist Hans Asperger (1906-1980) re-used this term to describe patients, so called ‘autistic psychopaths’, presenting social disabilities but he also pointed out that the patients did not present any hallucination nor delirium, ruling out the genealogy of this disease with schizophrenia[8].

In 1943, the American father of child psychiatry, Leo Kanner (1894-1981), described for the first time, autism spectrum disorders as distinct neurodevelopmental disorders, resulting from "autistic disturbances of emotional contact"[9]. However, it is only half a century later that ASDs were considered as a distinct group of diseases and the medical term autism first appeared in the Diagnostic and Statistical Manual of Mental Disorders (DSM-III) in 1980. In this classification, autistic patients were grouped into three distinct classes: childhood autism, Asperger's syndrome and atypical autism.

Currently, both the International Classification of Diseases (ICD; 11th edition published in 2018 by the World Health Organization) and the DSM-V, no longer delimit the differences between different autism groups, but instead include them in a continuum of the autistic spectrum, which reflects more the broad range of variability in this disease.

1.1.2 Epidemiology and social impact

Although ASDs were long considered as very rare neurodevelopmental disorders (one birth in 2000), they are now more and more patients identified and diagnosed with ASDs. This increase is likely due to a better awareness of these disorders in the general population in conjunction with an improved medical system for sick children. It is also worth mentioning that nearly one in four patients diagnosed today with ASDs based in the current inclusion criteria would not have been recognized in 1993[10], which is a consequence of the refinement and improvement of diagnosis efficiency.

In France, the French High Authority for Health (HAH) estimate that 1% of children suffer from ASDs with about 7 500 new births per year in France, totaling about 700 000 people living with ASDs including 100 000 below the age of 20 years old[11]. Patients require help at multiple levels, to alleviate their symptoms but also to be better integrated into the society.

The cost of care can vary between 4 000 € and 50 000 € per year and per patient, depending on the age, the level of dependence and the severity of the pathology[11]. Poor care by state services often leads to cessation of the professional support of one of the parents, to exclusion from social life and to severe distress within the family sphere, such as divorces or suicides. Poor care is typically not referenced but should be taken into account when assessing the social impact and burden of ASDs.

ASDs affect patients without any ethnic, geographical or social distinction. However, a strong sexual dimorphism has been convincingly demonstrated with a ratio of four men affected for one woman. This point will be discussed in more detail in Section 1.1.4.

1.1.3 Spectrum, diagnosis and comorbidities

The cardinal symptoms of ASDs are characterized by impaired verbal and non-verbal communication, as well as the presence of restricted interests and stereotyped behaviors[12]. They are characterized by difficulties to perceive social codes commonly used during communication, such as eye contact, body language, appropriate laughter or formulation (**Figure 2**). They also present a lack of interest towards novelty or decreased creativity, which manifest through unusual and monotonous games (stacking, aligning and ranging toys rather than playing with) or lack of flexibility and adaptability (sharp schedules, crisis due to unexpected or unplanned events,…). They are keen on repetition either verbal (repeating words or complete sentences) but also in movements like hand flipping or circling. ASD-patients reject social touch most of the time, which may be due to hypersensitivity to touch, and manifest deficits in earing, vision and sometimes synesthesia.

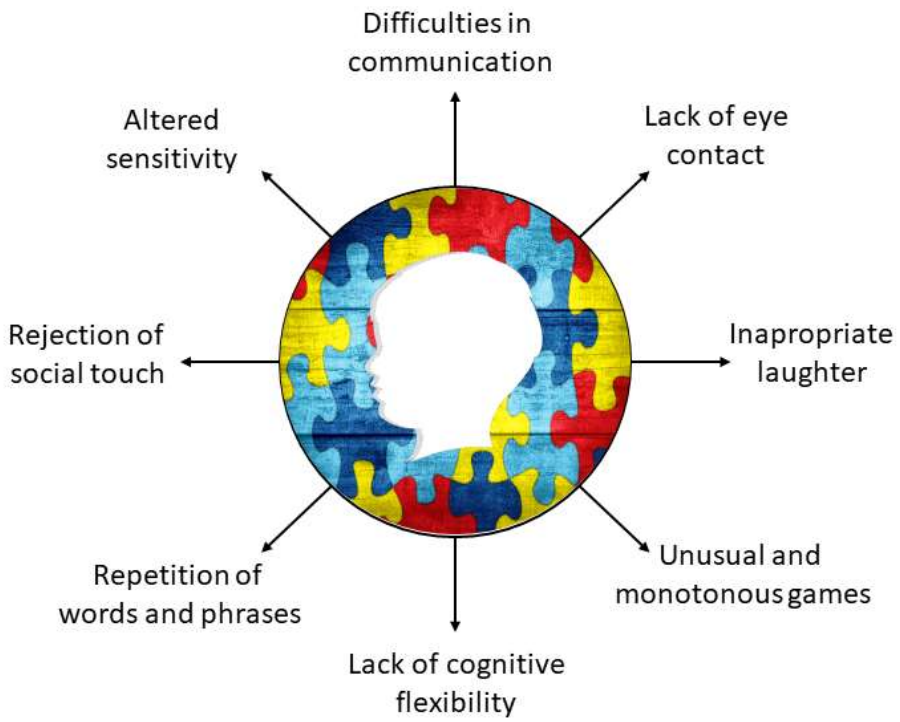
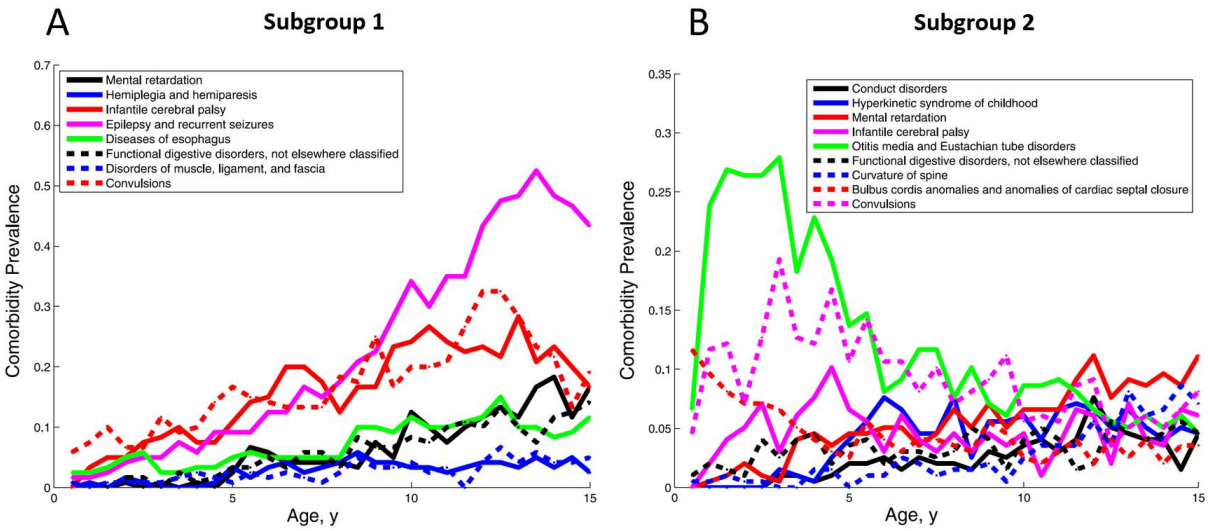


Figure 2 Manifestations of the cardinal symptoms in ASDs.

Diagnosis is typically made around three years old, even if some early signs can already be detectable around 12 months. The first signs generally include an absence of follow-up of the adult's gaze, as well as an absence of attempts to communicate with those around (verbal or non-verbal), or even an indifference towards interaction. At the behavioral level, most children with ASDs have a delay in speech learning. Sleep disorders are also reported in ASD patients such as insomnia, circadian rhythm shift and hypersomnia.

Although clinical diagnosis is primarily based on these cardinal symptoms, ASDs are in most cases accompanied by other associated symptoms[13]. In fact, it is quite challenging to estimate ASD comorbidity rates, because ASDs are clinically highly heterogeneous. One way to deal with this problem is to divide ASDs into several groups of autism and subsequently estimate comorbidity rates within each of these subgroups. To my knowledge, only one study has explored this avenue, both quantitatively and longitudinally from birth to 15 years of age[13]. Indeed, Doshi-Velez and colleagues first classified ASD patients into three subgroups, based on phenotypic convergence, before calculating the prevalence of the comorbidities. Interestingly, they showed distinctive pattern of comorbidities. For example, in the first group, at age 15

years, more than 50% of ASD patients showed epilepsy and about 20% of these same patients exhibited intellectual disability, a condition defined by an intellectual quotient (IQ) of less than 70 (**Figure 3A**). In the second group, otitis (30%) and convulsions (20%) were the two most frequent comorbidities in the five first years of life (**Figure 3B)**. In the third group, at 10 years of age, 35% of ASDs patients showed hyperkinetic syndrome (also known as attention-deficit hyperactivity disorder) and 15% had anxiety disorders but there was no indication of epilepsy in this group (**Figure 3C**).



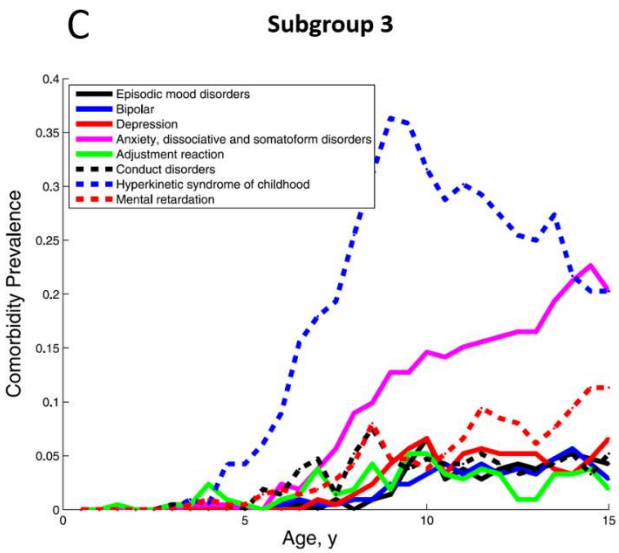


Figure 3 Prevalence of ASD-related comorbidities.

These comorbidities were calculated based on three subgroups (A-C). Adapted from Doshi-Velez *et al*. 2014.

This study is a way to conceptualize the diversity of the clinical picture of the patients and the caution needed when generalizing different aspects of ASDs.

The severity, as well as the presence or absence of these accompanying symptoms (commonly referred to syndromic versus non-syndromic forms of ASDs), can be extremely variable from one individual to another[14]. Occurrences of syndromic forms represent about 10 to 15% of ASDs cases and include Rett or Asperger's syndromes[15].

1.1.4 Sexual dimorphism

There are more men affected by ASDs than women, with a ratio of 4:1. Since women can be affected by this disease and are not completely ruled out from patient population, it is of great interest to understand which are the mechanisms, either diagnostic or physiological, which are at the origin of this gender difference.

The first clue relates to the way of making the diagnosis, which would be more inclined to identify cognitive deficits in men than in women. An increase in the number of women diagnosed in recent years argues in this favor and shows that education with health workers as well as the refinement of diagnostic methods tend towards a better perception of the different cases of autism within the whole population. This should prevent missed diagnosis for women autistic-patients.

This diagnostic bias is supported by a theory presented by Professor Simon Baron-Cohen of the University of Cambridge (UK), which depicts autism as an extreme state of the empathization-systematization continuum (ESESC), and enables to characterize, at the population level, social behaviors related to each gender. This theory bounds the autistic state to that of a hyper-masculinized brain (Extreme Male Brain theory), and consider that patients with ASDs present an increase in the behavioral characteristics of systematization (obsession of the systems, stereotyped behaviors, restricted interests) and a decrease in those related to empathy (communication, socialization and theory of mind)[16]. To resume, they observed that when plotting the scores obtained during empathization test (EQ score) versus systematization test (SQ score), ASD-patients were more likely to have SQ score > EQ score, even more than what was observed in men, with a greater proportion in the extreme-systematization group[17]. This theory was confirmed recently on a panel of nearly 700,000 participants, including 36,000 patients with autism spectrum disorders[18]. Indeed, Greenberg and colleagues observed that the cumulative frequency curves from autistic males and females are shifted towards higher systematization, when compared with respective controls (**Figure 4**). However, we can

observe that the score of autistic females is not significantly different from control males and may not cross the disease threshold, historically based upon male results, which may explain the difficulties of the ASD-diagnostic in women.

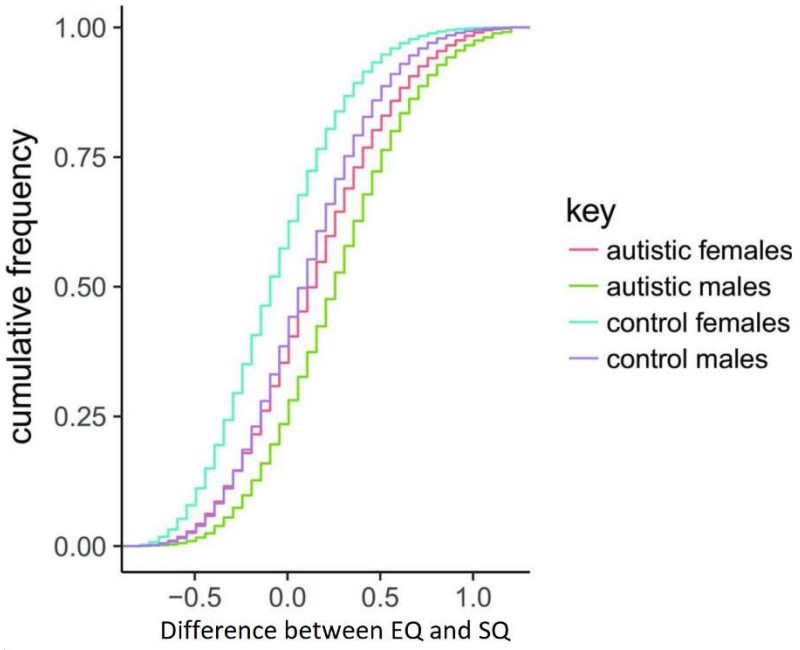


Figure 4 Empathization-systematization score in autism.

Cumulative frequency of the difference between EQ and SQ scores obtained by male and female from ASD- and control population. Score over 0 indicate better scoring in systematization. Adapted from Greenberg *et al*. 2018.

The second clue relates to physiological differences in ASD patients that may be triggered or modulated by sexual hormonal pathways, either by deleterious effects through testosterone[19], or by protective ones from estrogens[20]. Apart from a recent study on the impact of sexual hormones on human neural stem cell, the underlying sex specific molecular pathways remain largely unknown[21].

1.1.5 Neuroanatomy of autism

Many studies have focused on the identification of neuroanatomical landmarks as endophenotypes for ASDs, to understand the underlying neurobiological mechanisms or for diagnostic purposes. A study conducted on a cohort of 67 families declared that an estimation of 81% of autistic children present abnormal head circumference (microcephaly or macrocephaly), in addition to the core symptoms[22]. However, in view of the extensive clinical heterogeneity of ASDs combined with the complex anatomy of

the brain, this number should be considered with caution and larger cohorts would be required to make the link between neuroanatomy and autism, in an unbiased manner. We should therefore keep in mind that studies of brain anatomy in autistic patients, either by post-mortem studies or by magnetic resonance imaging (MRI), are relative to a specific cohort of autistic individuals, thus complicating this link. Besides, studies of brain anatomy in autistic patients are usually relative to one age group, with longitudinal studies lagging behind. It is therefore difficult to discriminate whether any neuroanatomical changes reported in the literature are dynamic over time or whether it could be due to heterogeneity and the lack of standardized procedures between studies.

Having said that, there is therefore growing evidence in the literature for the implication of the limbic and commissural systems in autism. Here I will cover the reported implication of the anterior cingulate cortex and the corpus callosum, as brain structure playing relevant role in ASD, as well as functional connectivity and synaptic plasticity, which may be key processes worth exploring.

1.1.5.1 Anterior cingulate cortex

The anterior cingulate cortex (ACC), corresponding to Broca's area 24a and 24b, is a key structure of the limbic lobe[23] and has been associated for a long time with the management of emotions, especially in the Jakob-Papez circuit[24]. Since then, more and more publications confirmed its role as a hub for: i) the integration of attention and emotions such as conflict-monitoring[25], ii) goal-directed behaviors[26] and, iii) reward input into the hippocampal memory system[27]. These are all key cognitive mechanisms often impaired in autistic patients.

MRI studies conducted on several ASD-patients cohorts frequently reported alteration in the ACC, confirming its importance in ASDs. This have been characterized by a decrease gyrification of the ACC (**Figure 5**) in a recent MRI study conducted in a cohort of autistic patients aged between 40-61 years[28]. This difference between ASD-patients and neurotypical controls was not observed during child- or teenage-hood[29], suggesting that the decreased gyrification started only after adolescence.

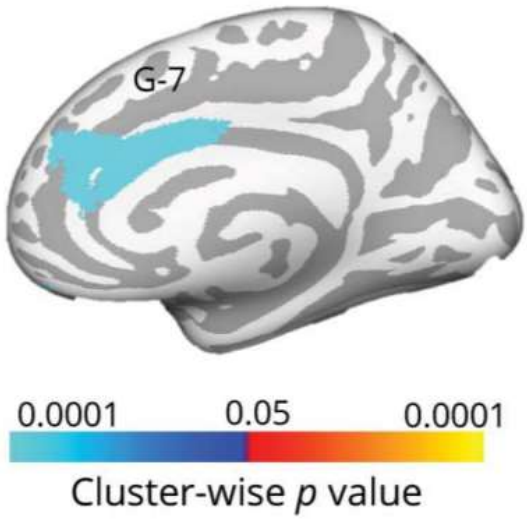


Figure 5 Abnormal gyrification in the anterior cingulate cortex of ASD patients. Adapted from Kohli *et al*. 2019b

A meta-analysis regrouping more than 50 MRI publication data, also pointed out the presence of a left-right difference in ACC thickness in ASD patients[30]. Interestingly, this asymmetry seemed to be influenced by sex, since only males were affected, thus reinforcing the interest towards this structure and the different processes it may be implicated in.

1.1.5.2 Corpus callosum

The corpus callosum (cc) is the largest commissure whose function is to connect the two cerebral hemispheres in placental mammals. Composed in human of 200-250 million axonal projections of cortical neurons that have crossed the midline to form homotopic and heterotopic connections in the contralateral hemisphere, the cc has a central function in integrating motor and sensory information from the two sides of the body and plays an important role in higher cognitive functions including social interaction, language abilities, processing speed and problem-solving abilities[31,32]. Previous studies have shown that ASD patients present a thick corpus callosum in the first postnatal months (**Figure 6**)[33]. This early postnatal increase in the size of the corpus callosum seems to come back to normal, as the children are aging.

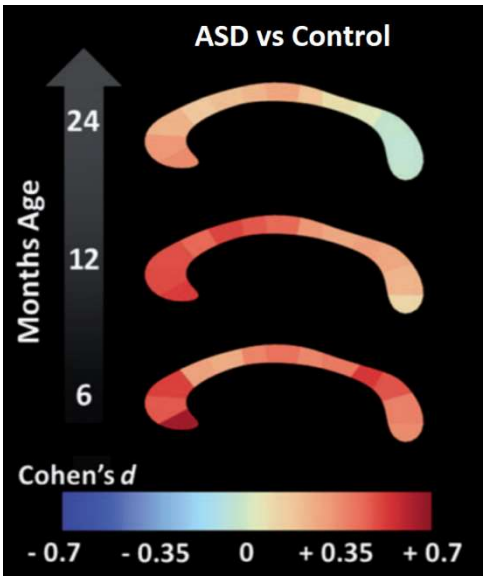


Figure 6 Changes in the size of the corpus callosum in young ASD patients.

The corpus callosum is represented in the sagittal plane with the rostral part on the left side and the caudal part on the right side. Cohen's d represents the longitudinal effect size. Adapted from Wolff *et al*. 2015.

Studies conducted on adult ASD-patients consensually agreed on an either partial or total reduction of the size of the corpus callosum when compared to controls[34,35]. For example, Alexander and colleagues used diffusion tensor imaging (DTI), a technique for mapping the diffusion properties of water within tissues, making it extremely sensitive to subtle differences in the architecture of the corpus callosum at the microstructural level. They found smaller total and regional corpus callosum volumes in a cohort of 43 high-functioning ASD patients (age ranging between 7 and 33 years)[34].

This is consistent with a higher incidence of ASD cases in cohorts of patients presenting corpus callosum agenesis[32]. However, a meta-analysis conducted using the Autism Brain Imaging Data Exchange (ABIDE) Project did not show any statistically significant difference and failed to recapitulate these observations using a large cohort of 694 adult ASD patients[36]. The authors pointed out to the increase of statistical power as a potential explanation for the difference observed between their analyses.

These different studies, amongst other in the literature, show evidences that the corpus callosum size is affected along the life of ASD patients, which may have consequence on the global cortical connections.

1.1.5.3 Functional connectivity

Rather than studying the size and the microstructure of the corpus callosum as a potential biomarker of autism, the analysis of interhemispheric functional connectivity have been proposed to better understand the underlying mechanisms in ASDs using functional Magnetic Resonance Imaging (fMRI)[37].

At young age, Supekar *et al.* showed an increased of the whole-brain connectivity in ASD-children (**Figure 7**)[38]. They observed that all of the functional connectivity between associative, limbic, paralimbic, primary sensory and the subcortex were increased compared to control group, and that it correlates with the severity of the social capacities of the children (the more the activity, the more severe the symptoms).

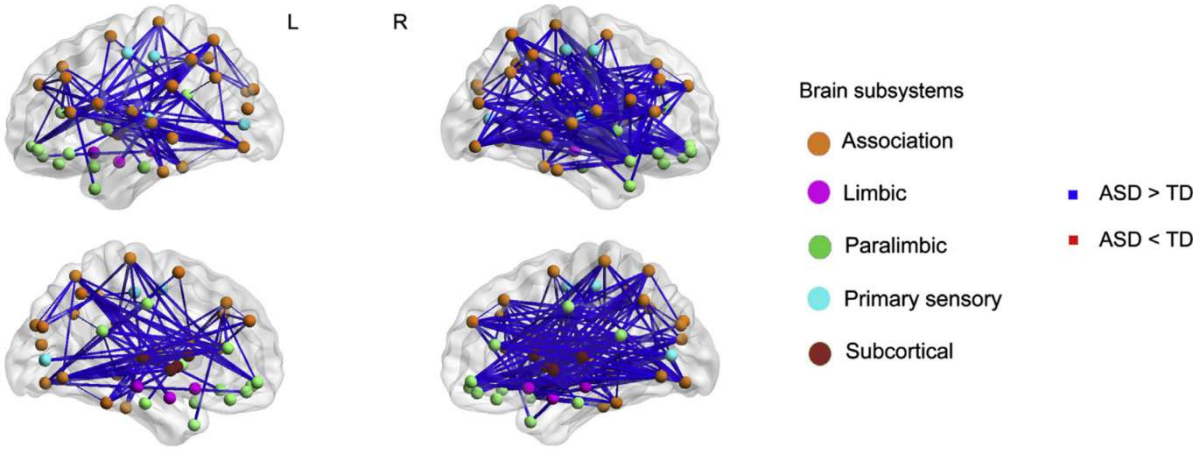


Figure 7 Whole-brain hyper-connectivity in children with ASD.

This figure shows 588 pairs of anatomical regions in children with ASD compared to IQ-matched typically developing children (referred to TD).L means left, and R means right hemisphere. Adapted from Supekar *et al*. 2013.

In adult, Courchesne and Pierce also pointed out abnormal connections in the frontal lobe of ASD-patients. Indeed, they analyzed connectivity at macro- and microscopic level and found out that the frontal lobe in ASD-patients was characterized by excessive connectivity within the frontal lobe, which led to be disorganized and poorly

selective, whereas communication with other brain structures was reduced and failed to be synchronized and efficient[39]. This pattern with increased local connections and decreased long-range ones, lead to frontal structures that are locally very active, but fail to receive and respond to other brain systems.

However, in a similar way to brain structural studies, there is a great importance to be more standardized and unbiased in order to depict satisfying classification of connective defects present in ASDs.

1.1.5.4 Synaptic plasticity

Autism is associated with impairments of communication at multiple levels, from brain regional to sub-neuronal levels. The synapse is the key sub-neuronal structure that enables communication from one neuron to another, transfer and integration of signals. The canonical chemical synapse contains pre- and post-synaptic compartments. At the axonal tip of the neuron, the pre-synapse releases neurotransmitters under the impulse of electrical stimuli, while the post-synapse, through its receptors, such as glutamate receptors, senses stimuli and activates downstream transduction cascades (**Figure 8**)[40]. Synapses can be divided as excitatory or inhibitory, according to their neurotransmitters/receptors.

Synaptic plasticity is defined as the capability of synapses to strengthen (or weaken) over time, in response to increases or decreases in their activity. These changes result from the alteration of the number of neurotransmitters/receptors at the synapse, the size of the synaptic cleft as well as other complex tightly regulated pathways.

ASDs were first described as synaptopathies, since the first genetic causes of autism came from the identification of major player in synaptic plasticity such as neurexins and neuroligins (**Figure 8**), which are key trans-synaptic molecules for both cell adhesion[41] and scaffold function[42]. Modifications of the fragile equilibrium which govern efficient plasticity may be leading to severe abnormalities alike to trigger the unset of ASDs[43].

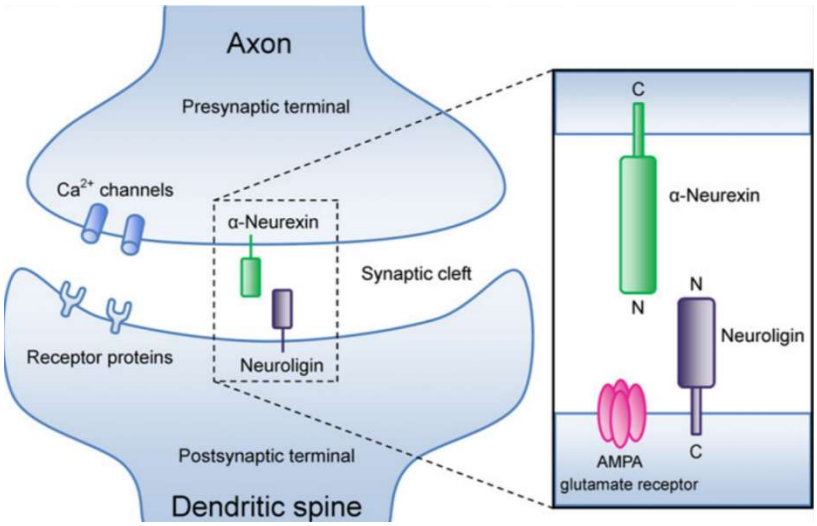


Figure 8 Structure of the pre- and post-synaptic compartments.

C and N letters indicate the C-terminus and N-terminus of the proteins. Adapted from Reichelt *et al*. 2012.

1.1.6 Genetic etiology

ASDs have an extremely complex etiology involving either environmental (such as *in utero* exposure to valproic acid or infections to Rubella) or genetic factors, or can stem out from a combination of both. The involvement of genetics in ASDs is now fully accepted within the scientific community. This has been evidenced by studies done on or family or twins in which the heritability rates were convincingly pointed out[44,45]. In studies comparing concordance of ASDs in twins, 40-90% of monozygotic twins (identical DNA) were concordant for diagnosis of ASDs compared to 0-30% of dizygotic twins, whereas in family studies, siblings of individuals with ASDs had a higher likelihood of having ASDs (up to 26% recurrence risk for siblings)[14].

Among ASD-patients, 25-35% present identified genetic causes (reviewed in [45]. This number is likely an underestimation, since about half of undiagnosed cases are predicted to have a genetic basis. When the underlying genetic architecture is extremely complex or when the variants are ultra-rare, it can be challenging to pinpoint the underlying causes.

It has become clear that the genetics of ASDs is driven by monogenic forms, explained by point mutations in one single gene, or by polygenic forms which are explained by an accumulation of common genetic variants[46]. Copy number variants (or CNVs;

deletion or duplication of large genomic regions leading to changes in the number of copies of the genetic elements encoded within those regions) are also playing an important role in the occurrence of ASDs with different levels of penetrance. These genetic models have been reviewed at multiple occasions[43,47,48]. **Figure 9** summarizes these models. There is a negative correlation between allele frequency and effect size of ASD-risk variants, indicating a negative selection pressure on variants leading to strong impact.

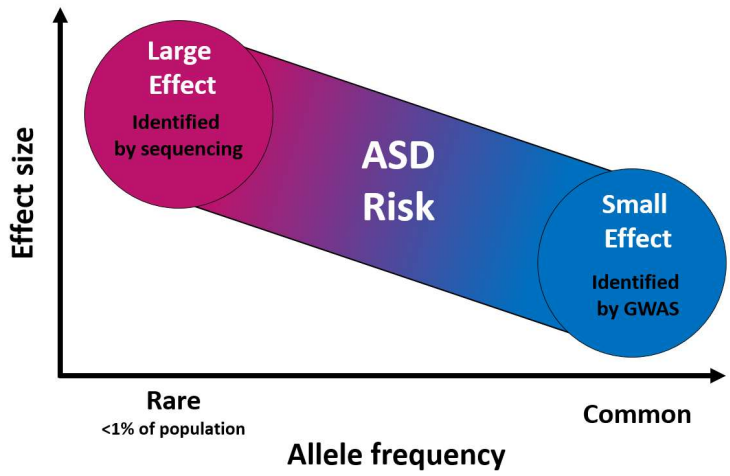


Figure 9 Models of ASD-risk variants.

Representation of the inverse relationship of allele frequency (x-axis) and effect size for ASD-risk variants. Rare mutations have larger effects than common variations.

1.1.6.1 Monogenic forms

For a long time, the wide range of ASD symptoms was an obstacle to consider that autism could be due to genetic mutations. However, the discovery of patients presenting mutations in the gene *SHANK3* made it unavoidable. Since then, and with the advent of exome sequencing, more and more point mutations have been associated with monogenic forms of ASDs. To date, more than hundreds of genes have been identified as causative for the onset of ASDs (reviewed in Betancur 2011)[49].

1.1.6.2 Polygenic risk

Several recent studies at the genomic scale have revealed that ASDs can also be caused by an accumulation of several variants each of small effect size. These

variants, also known as polymorphisms, are not sufficient *per se* to trigger ASDs, but each of them have a risk factor that together increase the susceptibility of developing autism. To tackle this problem, genome-wide association studies (GWAS) have been utilized in cohorts of thousands of patients and their relatives. GWAS studies pointed out to groups of genes, mostly related to neuronal function and corticogenesis, that may be at risk[46,50].

1.1.6.3 Copy Number Variants

In about 5% to 10% of ASDs cases, genomic rearrangements were implicated[51] as opposed to 1-2% of presence of CNVs in the general population[52]. This type of genomic remodeling can lead to either the deletion or the duplication of one to several genes, according to the length of the rearrangement. Over 50 genomic loci have already been associated with arising of ASDs symptoms (reviewed in Betancur, 2011[49]). In some cases, studies reveal that both the deletion and the duplication of the same region (called reciprocal CNVs) can lead to ASDs. For instance, CNVs at the 15q11-q13 locus are associated to autistic traits when duplicated (Dup15q11-q13 syndrome) or when deleted (Prader-Willi syndrome)[53].

1.1.6.4 Gene networks

Genes causing autism, through either monogenic forms, polygenic risk or CNVs, are interconnected within protein-protein interaction networks (**Figure 10**). Iakoucheva and collaborators have recently proposed different mode of action on gene regulatory networks (reviewed in [54]).

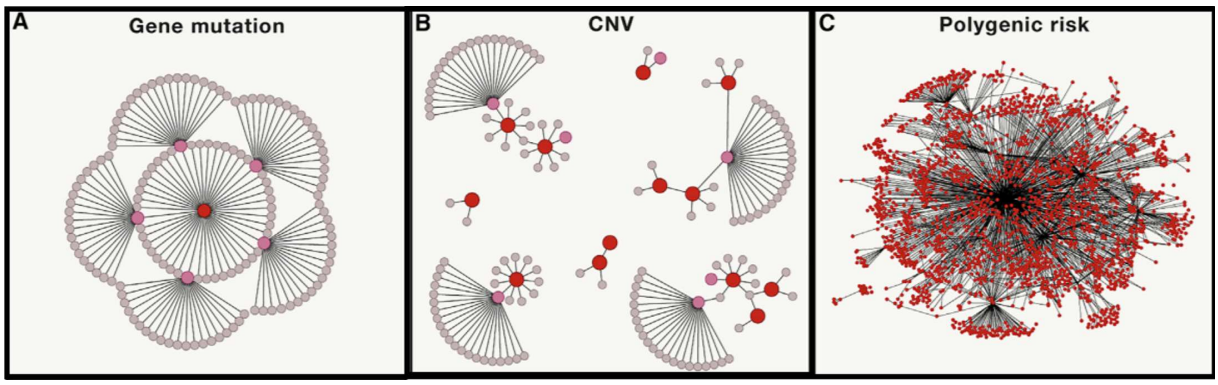


Figure 10 Models of gene network action in ASDs.

The impact of a single gene mutation in a transcription factor can easily spread out through a gene regulatory network (A). CNVs can alter several genes, each

contributing to a network (B). Polygenic risk is frequently distributed across many gene regulatory networks(C). Adapted from Iakoucheva *et al*. 2019.

The impact of a single gene mutation encoding for example a transcription factor can be sufficient to alter an entire gene network (**Figure 10A**). For CNVs, since multiple genes can be affected, the global effect on gene networks can be even broadly distributed (**Figure 10B**), whereas in polygenic forms, variations can act on numerous networks (**Figure 10C**). Genome-wide predictions studies have however identified only a small number of key networks. For instance, Krishnan and collaborators developed a machine-learning algorithm to identify brain-specific functional modules according to Gene Ontology (GO) terms, and found nine main clusters (**Figure 11**)[55], providing a landscape of cellular functions potentially dysregulated by ASD-associated mutations.

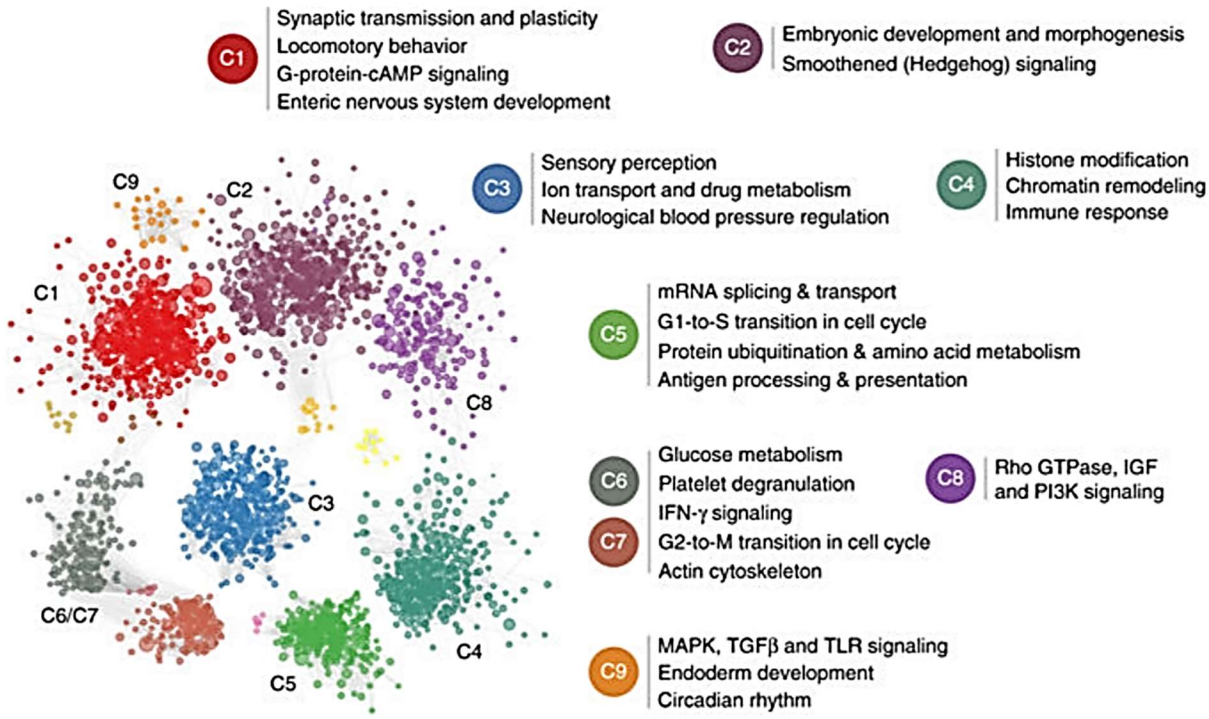


Figure 11 Main functional networks in autism.

Nine networks (C1 to C9), containing ten or more genes, were found through the analysis of 2 500 ASD-associated genes. Representative processes and pathways enriched within each network are presented alongside the cluster label. C6 and C7 were merged because they shared a number of strong links. Adapted from Krishnan *et al*. 2016.

These functional clusters regroups genes associated with synaptic transmission and plasticity (see cluster C1 in **Figure 11**) and chromatin remodeling (see cluster C4 in **Figure 11**). This is in line with previously reported networks[56]. Cluster C9 (**Figure 11**) highlighted signaling pathways, including mTOR and MAPKinases, whose convergence was also recently reviewed (**Figure 12**)[54].

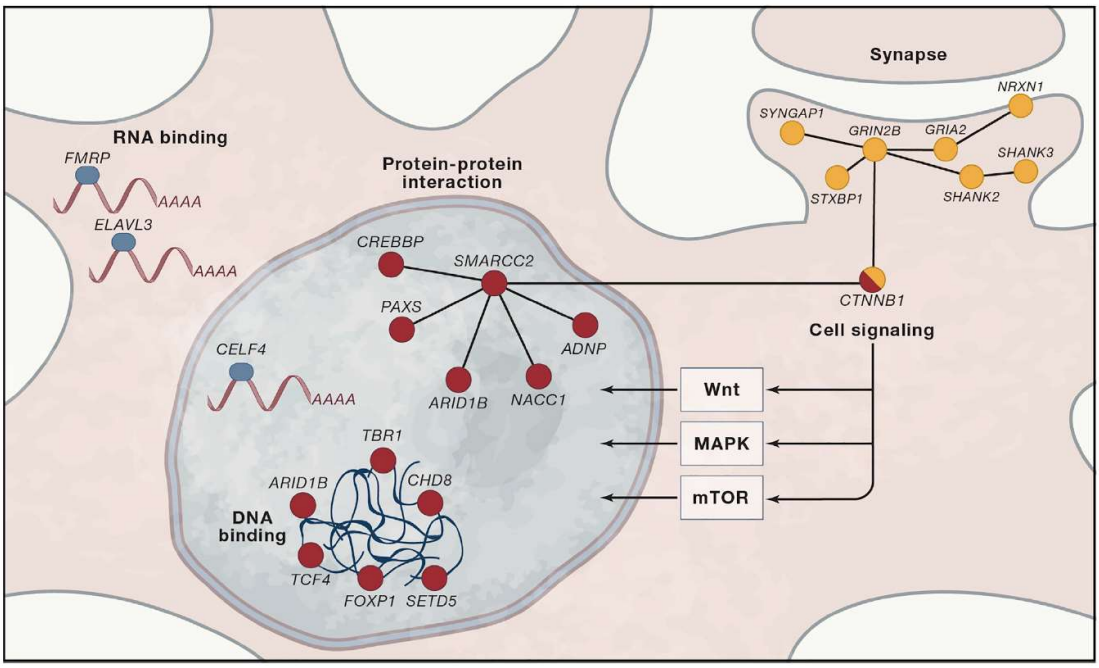


Figure 12 Key convergence in cellular functions among ASD genes.

ASD genes are highlighted as biological processes within a single neuron (cytoplasm in pink and nucleus in blue). Convergence is evident at multiple levels of interaction including DNA binding, RNA binding, and protein-protein interactions. Rare gene mutations in ASD also converge upon specific signaling pathways involved in the regulation of cell growth including mTOR, MAPK, and Wnt signaling. Adapted from Iakoucheva *et al*. 2019.

To summarize on this first part of my introduction, considering ASDs' clinical and genetic heterogeneity, it is not surprising that common endo-phenotypes have thus far been hard to identified and that no specific brain region or neuronal cell type been recognized in understanding the biology of ASDs. However, over the past 10 years, more and more studies recognized the great contribution of CNVs in autism biology. I will now focus on one particular CNV region associated to an autism-related syndrome: the 16p11.2 microdeletion syndrome.

1.2 Autism-associated 16p11.2 microdeletion syndrome

1.2.1 Identification of the locus

The 16p11.2 locus was identified as a risk factor for ASDs in 2008 in two independent GWAS studies which revealed the importance of rare CNVs (deletions and duplications) at this locus[57,58]. In the first GWAS study, Marshall and collaborators analyzed 427 ASD cases and their relatives by using SNP microarrays to identify the underlying causative CNVs[57]. They found five patients presenting either a deletion or a duplication of this region. These five patients segregated in four families as shown in **Figure 13**. In one of the family (NA0133), the duplication event was inherited from a healthy mother, whereas deletion events were *de novo*.

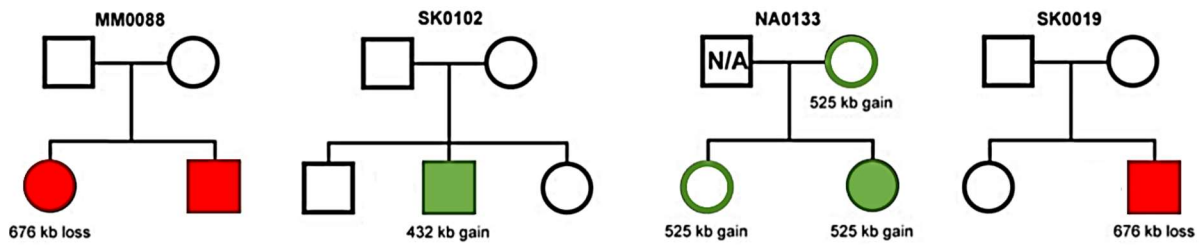


Figure 13 Pedigrees presenting CNVs at the 16p11.2 locus.

Males are denoted by squares and females by circles. The estimated size of each de novo or inherited event is shown below each family member. Open shapes are unaffected, filled have ASD diagnosis. The colour code corresponds either to a deletion (red) or to a duplication (green) of the locus. Adapted from Marshall *et al*.2008.

The second GWAS study, published at the same time by Weiss and co-workers, was based on 512 patients from the Autism Genetic Resource Exchange (AGRE) database and identified five children (four boys and one girl) carrying a *de novo* deletion of the 16p11.2 locus. Three other families showed at least one child affected by the reciprocal duplication amongst which two were inherited[58].

Together these two pioneer studies discovered CNVs at the 16p11.2 locus as the most common genetic cause of ASD-risk factor, accounting for about 1% of all ASD cases. Since then, studies often of higher sample size, confirmed the association between the 16p11.2 locus and autism, bringing the proportion of 16p11.2 cases to about 0.5% of global autistic cases[59], also reviewed in[60].

1.2.2 Genomic architecture

The genomic region of interest concerns the proximal part of the 16p11.2 locus located on chromosome 16 between coordinates 29.60 and 30.19 Mb (based on Ensembl Build 37). The 593-Kb region encompasses 27 protein coding genes: *SPN, QPRT, C16ORF54, ZG16, KIF22, MAZ, PRRT2, PAGR1, MVP, CDIPT, SEZ6L2, ASPHD1, KCTD13, TMEM219, TAOK2, HIRIP3, INO80E, DOC2A, C16ORF92, FAM57B, ALDOA, PPP4C, TBX6, YPEL3, GDPD3, MAPK3* and *CORO1A* (**Figure 14**). This region is flanked by two Low Copy Repeat (LCR), namely BP4 and BP5 (BP: Breakpoint), which are 147-kb in length and share more than 99% of sequence homology (**Figure 14**). This similarity creates a high predisposition to the formation of structural variation through Non-Allelic Homologous Recombination (NAHR) process. This mechanism occurs during recombination and is at the origin of both the deletion and the duplication of the locus. This flanking regions also contain three additional genes (*BOLA2*, *SLX1B* and *SULT1A*) which could also be involved in 16p11.2-associated symptoms.

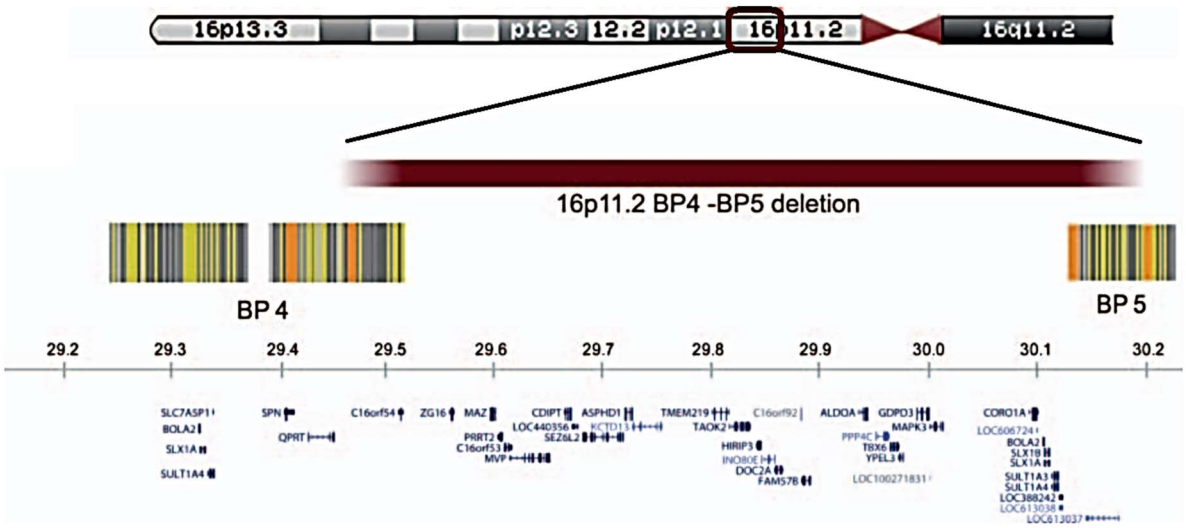


Figure 14 Localization of the 16p11.2 BP4-BP5 locus in the human genome.

BP4 and BP5 represent the Low Copy Repeat (LCR) regions flanking the locus of interest, and are at the origin of the CNVs. At the bottom part of the scheme are represented the genes contained at the locus and their genome sequence coordinates. Adapted from Zufferey *et al*. 2012.

Considering the high concentration of gene in this region, it is fair to think that the deletions and the duplications stemming out from the 16p11.2 locus may lead to a broad set of clinical manifestations in patients.

1.2.3 Symptoms of 16p11.2 microdeletion carriers

When comparing the symptoms of the patients affected either by the deletion or by the duplication, it was evidenced that most were mirrored symptoms. The deletion is associated with macrocephaly, autism and obesity, while the duplication is linked to microcephaly, schizophrenia and being underweight[61]. It is noteworthy to mention that these phenotypes are not completely penetrant with about 50% of CNV carriers not developing any phenotypes[62]. However, CNV carriers developing phenotypes show a wide array of phenotypes, ranging from weight issues to neurodevelopmental disorders.

1.2.3.1 Body weight

In the 16p11.2 breakpoint 4 to 5 (BP4-BP5) deletion carriers, one of the most common symptoms is the increase of body weight. This ranks the 16p11.2 deletion as the second most frequent genetic cause of obesity after the point mutations reported in the appetite-regulating melanocortin-4 receptor MC4R[63]. Zufferey and colleagues performed detailed evaluations on 72 deletion carriers and 68 intra-familial non-carrier controls and found that, despite a birth weight below average, around 70% of the adult patients meet criteria for obesity[59]. The body mass index (BMI) was already higher by the age of 3.5-year-old increasing continuously through age (**Figure 15**). Among the obese cases, 45% of them were classified as being in morbid obesity (**Figure 15**). Hyperphagia was recorded in most cases, suggesting that the elevation of BMI may be due to reduced satiety. This would be more in favor of behavioral and neurological issues than metabolic impairments[64].

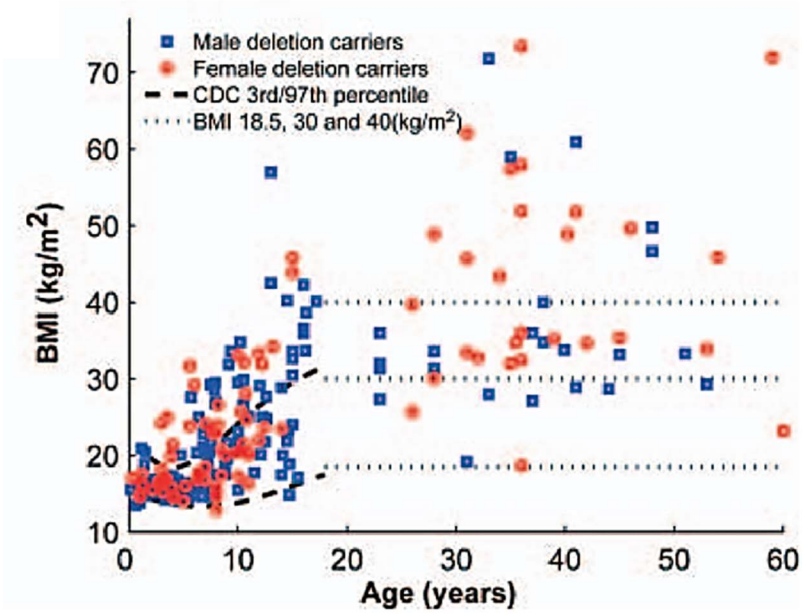


Figure 15 Distribution of body mass index in 16p11.2 deletion carriers.

Body mass index (BMI) progressively increased throughout life. 70% of adult carriers are obese (BMI>30kg/m², middle dotted line) with 45% of them in morbid obesity (BMI>40 kg/m², upper dotted line). Circles represent females and squares males. From Zufferey *et al*. 2012.

1.2.3.2 Neuropsychiatric impairments

Besides autism, other psychiatric diseases have been reported in children with 16p11.2 deletion. In a recent study, Niarchou and colleagues evaluated the prevalence of psychiatric disorders in 217 deletion carriers and 77 deletion family controls (**Figure 16**)[65]. Overall, 50% of deletion carriers were diagnosed with one or several neuropsychiatric disorders including autism, intellectual disability, attention deficit hyperactivity disorder, anxiety disorders, psychotic symptoms and oppositional defiant disorder/conduct disorder. Male-carriers present more risk than female-carrier to be associated with ASD or ID. The authors also found that the odds of developing psychiatric disorders were nine times higher among individuals bearing the 16p11.2 deletion than controls. However, no significant differences on the occurrence of anxiety disorders, psychotic symptoms (*i.e.* schizophrenia) or oppositional defiant disorder/conduct disorder (ODD/CD) were identified when compared to family controls (**Figure 16**). Intellectual disability (ID) is known to be highly prevalent in ASDs[49]. Niarchou and colleagues reported a prevalence of 30% of ID in deletion careers (**Figure 16**).

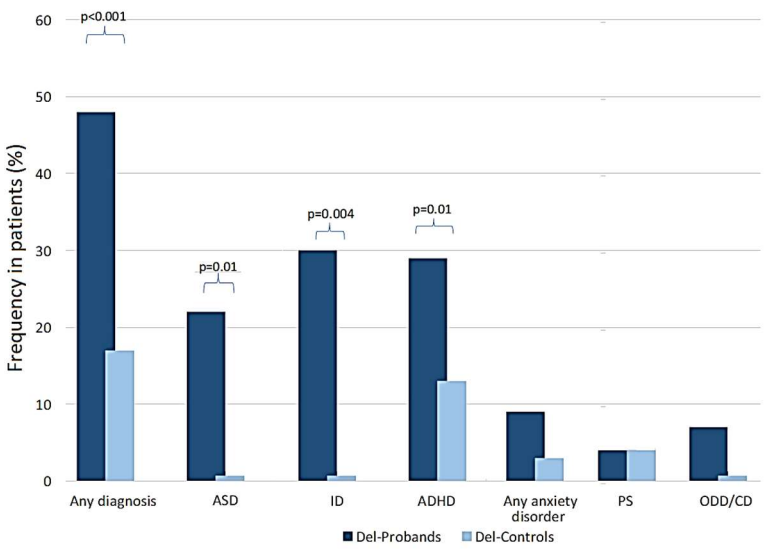


Figure 16 Neuropsychiatric disorders in children with 16p11.2 deletion.

ADHD, attention deficit hyperactivity disorder; ASD, autism spectrum disorder; ID, intellectual disability; ODD/CD, oppositional defiant disorder/conduct disorder; PS, psychotic symptoms. Adapted from Niarchou *et al*. 2019.

A similar ratio was also found in a second study where 20% of the 16p11.2 deletion carriers collected by Zufferey and colleagues met the criteria for ID[59]. The deletion carriers exhibited a reduction of their IQ score by about 32 points when compared to family non-carrier controls (**Figure 17**).

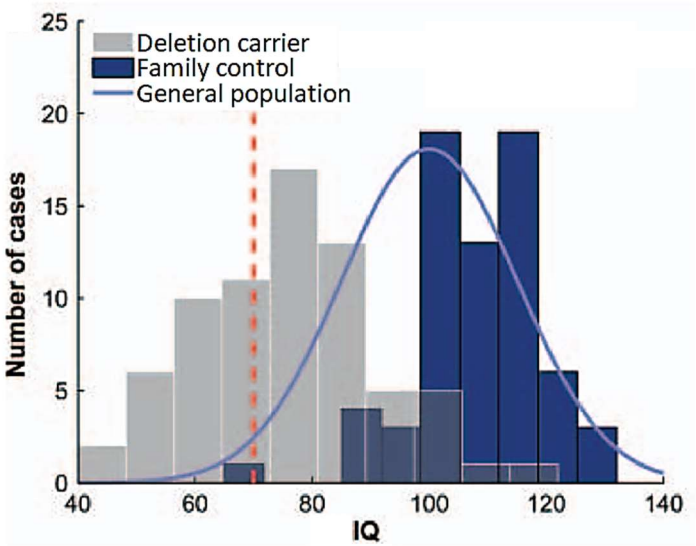


Figure 17 Distribution of IQ scores in deletion carriers.

Distribution of *IQ scores in 16p11.2 BP4-BP5 deletion carriers (grey bars),* intra-familial non-carrier relatives (control, blue bars) and general population (blue bell curve). The red dashed vertical line represents the IQ threshold (70) below which the diagnosis *of intellectual disability (ID*) is typically made. Adapted from Zufferey *et al. 2012.*

1.2.3.3 Neuroanatomical defects

Macrocephaly, defined as an increase in head circumference (HC) z-score higher than two, is one of the most frequently observed neuroanatomical feature in the 16p11.2 deletion carriers. Zufferey and colleagues reported 29 among 170 (17%) of the BP4-BP5 deletion carriers with an overall increase in HC. In a similar way to BMI, head circumference in 16p11.2 deletion carriers is lower than average at birth but underwent a fast progression during the first two years of life (**Figure 18**)[59].

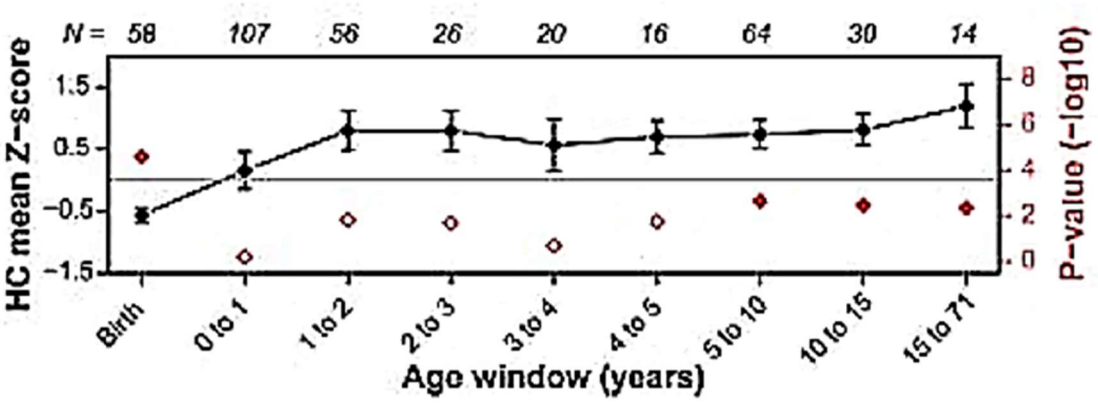


Figure 18 Head circumference in 16p11.2 deletion carriers.

Head circumference (HC) mean z-scores and corresponding p-values (in red) for each developmental age window. The number of cases is indicated for each age category at the top of the figure. Adapted from Zufferey *et al*.2012.

Since this first report of brain size defects in 16p11.2 deletion carriers[59], four additional studies have been published using structural MRI to gain a better understanding of the specific brain structures altered in 16p11.2 deletion carriers.

In the first study, Qureshi and collaborators conducted structural MRI measurements on 29 deletion carriers (from 8 to 17 years old) and 25 age-matched controls. MRI data was collected according to the Simons VIP neuroimaging protocol database. They showed that children presenting the deletion have an increased intracranial volume (**Figure 19A**) as well as an increased brain volume (**Figure 19B**). Both the size of gray (**Figure 19C**) and white (**Figure 19D**) matters were also enlarged[66].

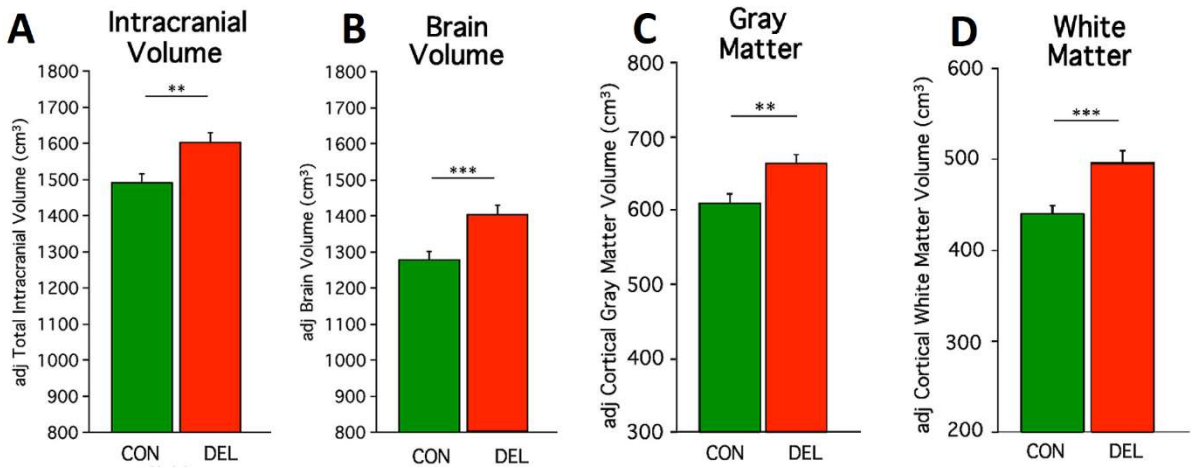


Figure 19 Brain MRI measurements of 16p11.2 deletion carriers.

Adjusted measures of intracranial volume (A), brain volume (B), gray matter (C) and white matter (D) structures. Adapted from Qureshi *et al*. 2014.

The second structural MRI study was conducted in 2018 using 79 deletion carriers with an age range of 1-48 years. This study showed a thicker corpus callosum as well as cerebellar ectopia and dens and craniocervical abnormalities in deletion carriers (**Figure 20**)[67].

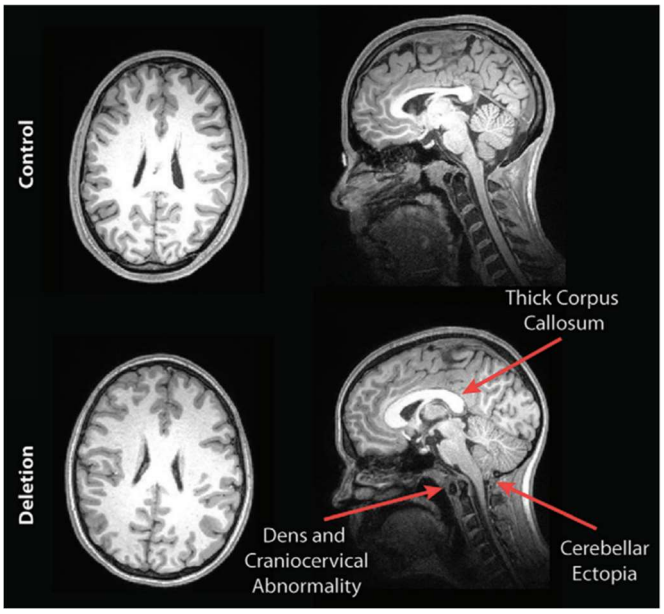


Figure 20 MRI pictures of 16p11.2 deletion carriers.
Adapted from Owen *et al*. 2018.

The third study used voxel- and surface-based brain morphometric methods to analyze structural MRI data collected at seven sites from 78 individuals with a deletion and 212

individuals without a CNV[68]. The authors observed an increase in the size of the insula, the calcarine cortex and the transverse temporal gyri while the superior and middle temporal gyri were both decreased in size.

Finally, the fourth voxel-based study of 56 deletion carriers was more precise and specific, showing regional increase of the insula, the posterior orbital gyri as well as the anterior cingulate gyrus and frontal and occipital white matter (**Figure 21**)[69]. However, this study also showed that some structures were decreased in size such as the cerebellum, the middle cingulate gyrus, the motor cortex and the putamen.

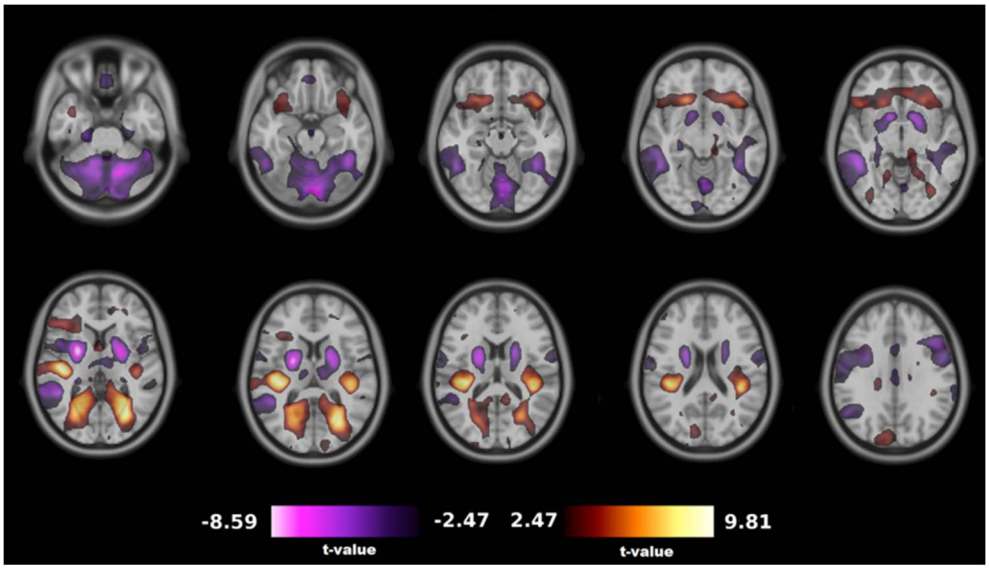


Figure 21 Voxel-based differences of 16p11.2 deletion carriers.

Projection of voxel-based volume differences between 16p11.2 deletion carriers and controls on horizontal anatomical MRI pictures. Regions highlighted in purple are decreased in size while red-yellow ones are increased in size. Only regions with a false discovery rate (FDR) correction ($q < 0.05$) are presented. Negative t-values represent the contrasts DEL < CTRL, positive t-values represent the contrasts DEL > CTRL. Adapted from Cárdenas-de-la-Parra *et al*. 2019.

Together, these four independent studies reinforced the enlarged global brain morphometry in the 16p11.2 deletion carriers, which was concomitant with enlarged gray and white matter structures[66–69]. However, it also pinned out differences between these studies. The lack of reproducibility could stem out from the lack of standardization in the equipment itself, in the resolution of the acquisition protocols, in the quantification step or homogeneity of patients cohorts.

1.2.4 Molecular consequences

Another approach used to understand the underlying pathways is to analyses gene expression differences between the 16p11.2 deletion carriers and the controls. To do so, most studies reported the impact of the 16p11.2 deletion on the transcriptome using lymphoblastoid cell lines (LCLs) derived from patients since brain samples are not very often available. For example, Blumenthal and colleagues measured the level of *cis*-transcripts from the genes within the locus and found a uniform expression pattern of 16p11.2 genes (**Figure 22**)[70]. This indicates that there is little or no dosage compensation involved in the regulation of expression of genes within the 16p11.2 locus, suggesting that the core symptoms are likely due to haplo-insufficiency of one or more of the 27 coding genes within the 16p11.2 region.

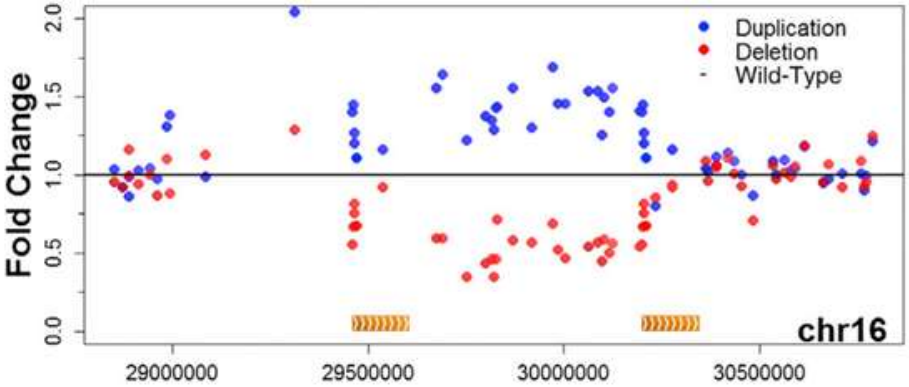


Figure 22 Expression profile of genes within the 16p11.2 proximal region.

The X-axis represents the genomic coordinate position and the Y-axis the relative fold change in expression across the 16p11.2 CNV in human lymphoblastoid cell lines (LCLs). The level of expression is shown for either duplication (in blue) or deletion (in red) compared to control family members (black line). Breakpoints 4 and 5 are represented in orange. Adapted from Blumenthal *et al*. 2014.

The authors also evaluated *trans* effects due to structural impact on the chromatin organization and found experimental evidence for interactions between 16p11.2 and each of 6p22, 12q24, and 19q13 loci[70]. Another study that explored the *trans* effects of the 16p11.2 BP4-BP5 deletion found 70 differentially expressed genes outside of the 16p11.2 region[71]. GO enrichment analysis of these 70 differentially expressed genes revealed an excess of genes implicated in cellular pathways, including neural-

related ontologies such as the synapse, the post-synaptic membrane and long-term potentiation (**Figure 23**).

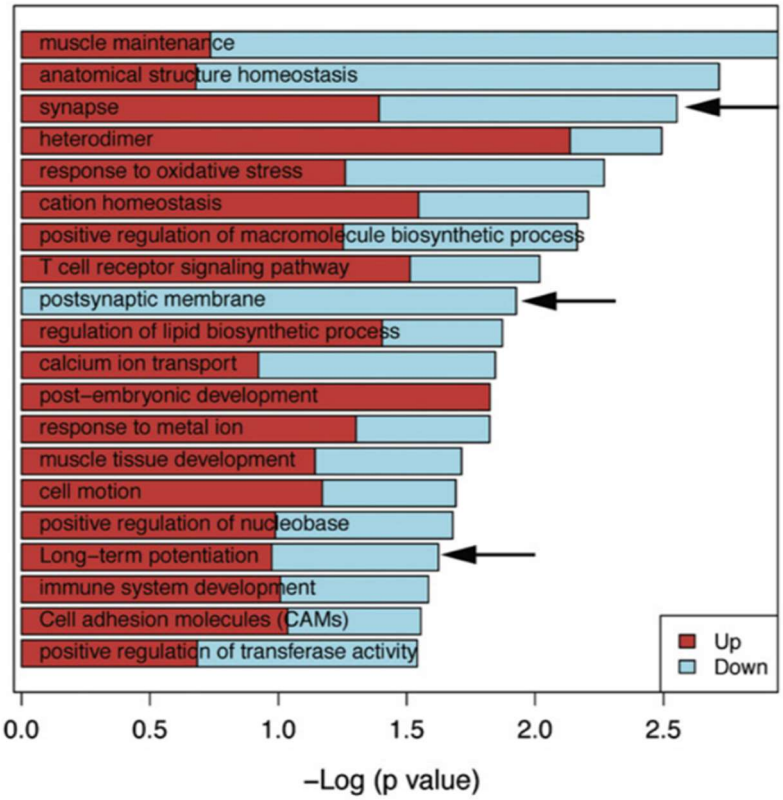


Figure 23 Enrichment of differentially expressed genes in deletion carriers.

The –log10 of the uncorrected p-value is shown in the x-axis. The Y-axis represents the DAVID GO term enrichment tested in the 70 differentially expressed genes using lymphoblastoid cell lines (LCLs) derived from the 16p11.2 deletion patients. Adapted from Luo *et al*. 2012.

One of the main limitations of this approach is the lack of relevance of LCLs in the field of neurodevelopmental disorders as there may be some essential genes missed because they are simply not expressed in lymphoblastoid cell lines but are in the brain. However, it can offer some clues to uncover the underlying cellular pathways affected. One alternative approach to identify the direct consequences of the 16p11.2 deletion on the development of the brain is the use of animal models.

1.2.5 Mouse models of the 16p11.2 microdeletion syndrome

The laboratory mouse offers a number of powerful tools to study human genetic disorders. Two of these are the easy access to brain sample as well as the ability to modify mouse genome with genetic engineering techniques. Soon after the discovery

of the 16p11.2 CNVs and their implication in neurodevelopmental disorders, several mouse models were developed in order to model the mutation, characterize the phenotypes and understand the disease mechanisms.

This was facilitated by the good conservation of this locus through evolution. Indeed, the mouse genome present a syntenic region corresponding to the 16p11.2 human one, with the same gene content and order, and located on the band F3 of chromosome 7 (**Figure 24**). Already three different mouse models were engineered and fully characterized, each of them encompassing the common 27 genes starting from *Coro1a* and ending with *Spn*[72] with additional genes enlarging the deleted interval to *Slx1b-Sept1*[73] or *Sult1a-Spn*[74]. The boundaries for each of the mouse models are represented in **Figure 24**.

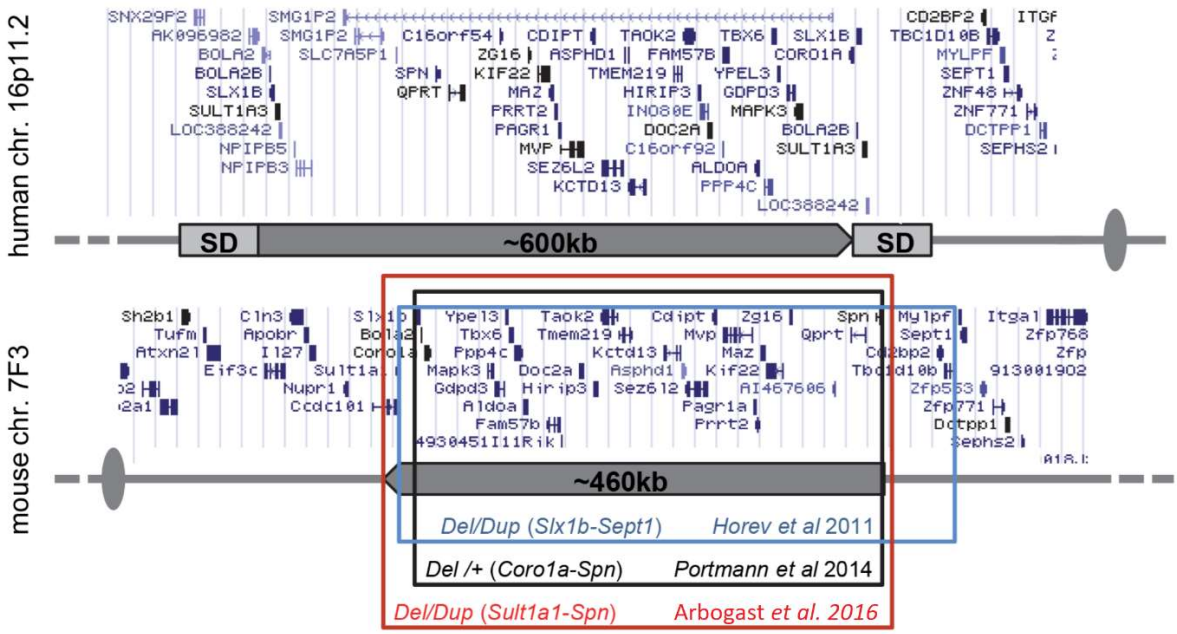


Figure 24 Mouse models for the 16p11.2 rearrangements.

Top panel: human 16p11.2 region and proximal segmental duplications prone to generate BP4-BP5 copy number variations (CNVs) by a non-allelic homologous recombination (NAHR) mechanism. Bottom panel: 16p11.2 syntenic region on mouse chromosome 7F3. Colored boxes indicate the genetic intervals studied in the respective mouse model studies. All genomic positions are given according to UCSC mouse genome browser GRCm38/mm10. Adapted from Arbogast *et al*. 2016.

1.2.5.1 *Slx1b-Sept1* deletion model

Horev and collaborators published in 2011 the first description of mouse models presenting 16p11.2 CNVs, encompassing genes from *Slx1b* to *Sept1*[73]. In their description of the 16p11.2 *Slx1b-Sept1* deletion mice (generated on 129sv background), the authors first reported an increase in early perinatal mortality and a decreased number of pups bearing the deletion at weaning.

No muscular, hearing nor social impairments were observed in this strain when compared to littermate controls, but the mice presented higher locomotor activity during both light and dark periods. More specifically, this was accompanied by an increased vertical activity (*i.e.* rearing and climbing) and reduced events of grooming and resting. Furthermore, the authors noticed stereotypic and repeated behaviors in the mutant mice when climbing in the ceiling. While wild-type mice presented adaptive patterns to climb on the cage ceiling, mice deleted for the 16p11.2 *Slx1b-Sept1* locus only repeated one kind of skill to go up and down, and often were blocked at the ceiling with apparent difficulties to return down.

In order to investigate the neuroanatomical architecture which may be at the origin of these behavioral deficits, MRI studies were performed. These analyses showed a mild increase in the volume of several subcortical brain structures including the superior colliculus, the fornix, the mammillothalamic tract, the midbrain, the periaqueductal gray and the hypothalamus (**Figure 25**). No gender differences were observed when comparing male and female deleted mice.

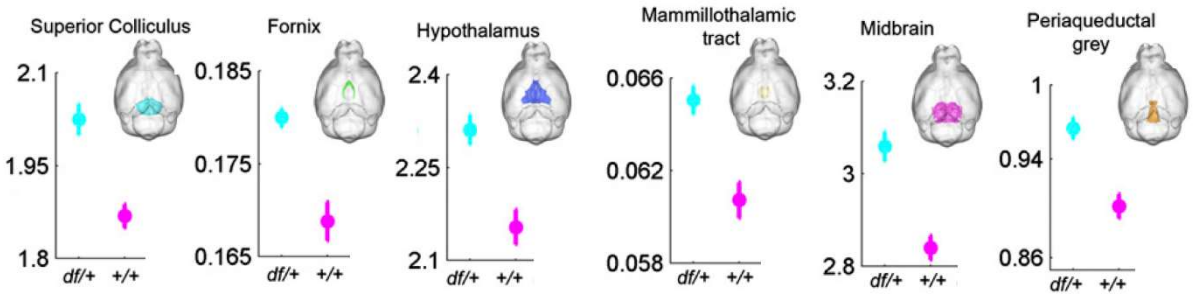


Figure 25 Brain structure in mice deleted for 16p11.2 *Slx1b-Sept1*.

Relative volume (in percentage of total brain volume) of six significantly increased brain structures determined by MRI. Df/+ correspond to the deletion of the region. Adapted from Horev *et al*. 2009.

Later on, Pucilowska and colleagues used the same 16p11.2 *Slx1b-Sept1* deletion mice and showed anxiety-like behaviors (assessed by shorter time in exploration of the open arm of the elevated-plus maze), as well as recognition memory deficits (by using the novel object recognition or NOR test)[75]. This study also reported cortical cytoarchitecture abnormalities and altered neurogenesis, especially a decreased number of upper-layer and increased number of deeper-layer neurons.

1.2.5.2 *Coro1a-Spn* deletion model

In 2014, the second mouse model of the 16p11.2 deletion was created by the insertion of LoxP sites at the determined loci in mouse embryonic stem cells (mESC) spanning the 16p chromosome from *Coro1a* to *Spn*[72]. The genetic background was mixed on C57BL6/N and CD1 mouse strains.

They observed that the offspring was born at Mendelian ratios, suggesting no prenatal mortality due to the mutation. However, some of them died within the first postnatal weeks, which could be partially restored by adapted regimen and eased access to food. The mice were smaller and leaner than their controls in the first days of life and onwards, despite no differences in food intake. Vision and olfaction functions were not impaired, but auditory capacities were drastically reduced.

Next, the authors performed a battery of tests in order to test for motor and behavioral abnormalities. The mutant mice exhibited motor deficits, such as tremor and decreased gait fluidity but without lack of coordination ability, and were hyperactive notable with an increased in hanging and circling behaviors. The mutant mice spent less time for grooming and resting. They were less keen to novelty which was assessed by a reduced exploration in the open field test, and performed less in memory tests such as in the NOR test. However, these mice did not show any pattern of anxiety-like behavior nor sociability impairments.

At a neuroanatomical level, Portmann and colleagues only observed a slight but significant reduction of brain weight at P7. They also conducted MRI studies in order to look at region-specific brain volume abnormalities. They observed thinning of most cortical regions, as well as volume increase of the basal ganglia (**Figure 26**).

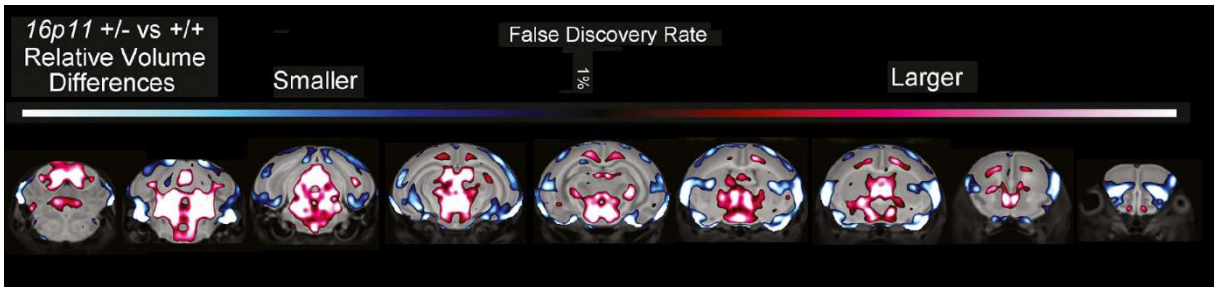


Figure 26 Neuroanatomical abnormalities in 16p11.2 Coro1a-Spn deletion model.

Coronal flythrough highlighting significant differences in the relative volume of P7 16p11+/ mouse and control (red: larger, blue: smaller); only highly significant areas are shown (q < 0.01). Adapted from Portmann *et al*. 2014.

Finally, the authors conducted functional studies to understand the molecular causes that could be at the origin of the deficits in these regions. They found an imbalance between dopaminergic neurons in the striatum and dopamine-sensitive cortical neurons (increased and decreased, respectively). This was leading to electrophysiological abnormalities, suggesting impaired basal ganglia circuitry function in these 16p11.2 *Coro1a-Spn* deletion mice.

Another group used the 16p11.2 *Coro1a-Spn* deletion as a model and reported impairments in pain-sensitivity, deafness and reduction of ultrasonic vocalization during social interaction, as well as stereotypies in these mice[76].

1.2.5.3 *Sult1a-Spn* deletion model

In 2016, the third deletion mouse model was engineered by Arbogast and colleagues at the Mouse Clinical Institute[74]. This model corresponded the best to the genomic rearrangements present in patients, spanning genes from *Sult1a* to *Spn*. The influence of the genetic background was also studied on pure inbred (C57BL/6N) and hybrid (F1 C57BL/6NxC3B) mice.

Perinatal mortality was observed in inbred 16p11.2 *Coro1a-Spn* deletion mice whereas no lethality was observed on a mixed background. Both inbred and mixed background mouse models showed a decreased in body weight and adipogenesis both on standard and high-fat diet. Mice bearing the 16p11.2 *Sult1a-Spn* deletion did not show any hearing defaults. Arbogast and colleagues were the first to report social interaction phenotypes in the 16p11.2 deletion models. This phenotype was however specific to

the B6N/C3B hybrid background. Analysis of skull morphology also allowed detecting smaller skull but only in female mice.

1.2.5.4 Comparative study between mouse models

To sum up, **Table 1** shows a map of phenotypes observed in mice carrying deletions of *Slx1b-Sept1*, *Coro1a-Spn* and *Sult1a1-Spn* genetic intervals on hybrid and inbred genetic backgrounds.

We can observe that the different models do show some similarities, notably for increased of post-natal death, decreased body weight and length (despite no difference in food intake), impaired memory, absence of coordination problems, modifications of the general activity (decreased of rooming and resting in favor of exploratory behaviors as rearing) and the presence of stereotypic behaviors. However, there are also discrepancies between the models concerning the exploration of novel environment and anxiety, as well as for motor strength. Although the main mouse models did not show obvious sociability impairments, with classical tests, Yang *et al.* observed increased in ultrasonic vocalization in *Coro1a-Spn* model. This points out that the presence of social phenotypes in 16p11.2 models are difficult to address and may need more advanced protocols. The hybrid C57BL/6NxC3B *Sult1a-Spn* model was the only to mark a significant difference during the three-chamber social test, thus enforcing the impact that genetic background could play.

Of all, neuroanatomy is the parameter with the most uncertainty. Indeed, despite the fact that Arbogast *et al.* found smaller skull in female, Horev *et al.* were describing increased global brain size in males, whereas Portman *et al.* depicted a more complex picture with reduced cortical regions and bigger basal ganglia. This is either pointing out differences induced by the shift of the genetic frame in each model and importance of flanking gene for this phenotype, or to the fact that finer tuning in the assessment of neuroanatomy would be more accurate to describe the mild effect in 16p11.2 deletion mutants.

		Interval	*Slx1b-Sept1* del		*Coro1a-Spn* del		*Sult1a-Spn del*	
		Publication	Horev *et al. 2011*	Pucilowska *et al.2015*	Portmann *et al.* 2014	Yang *et al.* 2015	Arbogast *et al.* 2016	
		Genetic background	129sv	129sv	C57BL6/N / CD1		C57BL/6N	C57BL/6NxC3B
	Subcategory	Technique						
General health	Olfaction	olfactory habituation/dishabituation test			no difference			no difference
	Vision	forepaw reaching test			no difference			
	Audition	startle response			decreased	decreased		
	pain-sensitivity					Increased		
	Response to sedation	open field with risperidone			decreased			
	Prenatal lethality	Mendelian ratio of births	no difference		no difference		no difference	no difference
	Post natal death		increased		increased		increased	no difference
Behaviour	Dark-activity	home-cage recordings	increased		increased		increased	increased
	Light activity	open field	increased	no difference	decreased		increased in center	no difference
	Circadian activity						no difference	no difference
	Novel environment exploration		increased		decreased			
	Sociability	3 chamber test	no difference		no difference		no difference	decreased
	Vocalisation during social interation					decreased		
	Novelty social seeking	social novelty test			no difference			
	Memory	Novel object recognition		decreased	decreased		decreased	decreased
	Habituation	Inter Trial interval			decreased			
	Hanging		increased		increased			
	Rearing		increased				increased	increased
	Jumping		increased				increased	increased
	Grooming		decreased		decreased			
	Resting		decreased		decreased			
	Circling				increased			
	Anxiety	elevated plus maze		increased				
	Anxiety	open field		increased	no difference			
	Stereotypic behavior		increased		increased	increased	increased	increased
Motor	Coordination	rotarod			no difference		no difference	
	Fluid gait				decreased			
	Fluid gait	notch bar					decreased	
	Tremor				increased			
	Grip strength	grip strength test	no difference				increased	
	Muscles anatomy						no difference	
Metabolism	Body weight adult	Weight		decreased	decreased from P1		decreased	decreased
	Birth weight	weight					decreased	
	Food intake				no difference		no difference	
	Body length				decreased		decreased	decreased
	Fat				decreased		decreased	no difference
Neuroanatomy	Brain weight				no difference (except P7)			
	Global	MRI	increased		decreased cortices increased BG (P7)			
	Basal Ganglia	MRI			decreased absolute volume increased relative volume for NA and GP (P7)			
	Lateral hypo	MRI	increased					
	Skull size	computed tomography					decreased	
Cellular	PPR	electrophysiology			decreased in striatum		no difference in HP	
	mEPSC frequncy				Increased			
	mEPSC amplitude				No difference			
	Spine density				No difference			
	sEPSC frequency				increased			
	sEPSC amplitude				no difference			
Molecular	Leptin						decreased	
	Adiponectin						decreased	
	Midbrain markerss						decreased	
	mTOR						decreased	

Table 1 Comparative analysis between 16p11.2 microdeletion mouse models. Grey indicates parameters not available.

1.2.6 Similarities between 16p11.2 mouse models and human microdeletion carriers

Overall, the mouse models in the previous section recapitulated most of the phenotypes observed in 16p11.2 deletion carriers in human. Mice hyperactivity and presence of stereotypic behaviors can be associated respectively with ADHD and ASDs reported in 16p11.2 deletion carriers. As in human, mouse models do not show muscular impairments, and defects on motor tests may just reflect reduction of global activity in mutant mice.

However, although mouse models present robust metabolic impairments, the direction of the effect is opposite to what is seen in patients. In the same perspective, the decreased brain volume observed in 16p11.2 deletion mouse models also show directionality differences with phenotypes in patients, with most of them displaying macrocephaly. These discrepancies may be explained by the evolutionary gap present between the two organisms and by differences in species biology.

According to the complex networks that could be affected by the deletion of a CNV of the size of the 16p11.2 one, it is of great interest to look down at the genic level and start to understand how the genic content may affect pathways that are leading to the phenotypes observed in patients and mouse models. The 16p11.2 locus contains 30 genes, two of which are Mitogen-activated protein kinase 3 (MAPK3) and Major Vault Protein (MVP). The next sections of this introduction will focus on these genes and implication in physiological process, which would give inputs for the understanding of the core of my thesis work.

1.3 Mitogen-activated protein kinase 3 (MAPK3)

1.3.1 Mitogen-activated protein kinases (MAPKs) and origins

Conserved throughout eukaryotes, mitogen-activated protein kinases are cytoplasmic enzymes specifically involved in the phosphorylation of the amino acids serine and threonine[77]. MAPKs are implicated in the regulation of cellular responses to external stimuli from the surface to the interior of the cell, such as mitogens and growth factors, and the regulation of cellular functions including growth, survival and apoptosis.

Most MAPKs have common features including their activation typically dependent on two phosphorylation events. However, there are also some MAPKs that do not have dual phosphorylation sites, referred to "atypical" MAPKs (**Figure 27**).

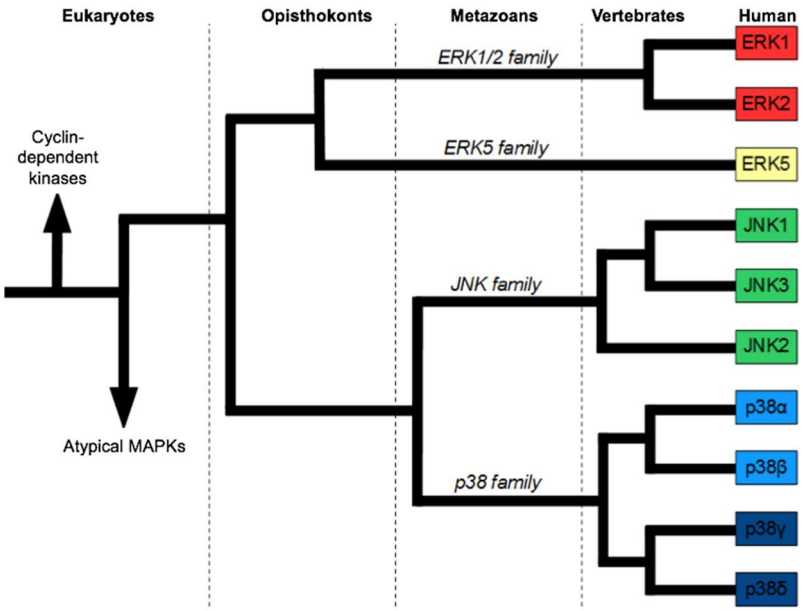


Figure 27 The origins of human mitogen-activated protein kinases.
Adapted from Li *et al*. 2011.

MAPKs includes the extracellular signal-regulated kinase (ERK) family, the c-Jun N-terminal kinase family (JNK) and the p38 kinase family (**Figure 27**). In mammals, the first mitogen-activated protein kinase to be discovered was MAPK3 (or ERK1), a kinase that is very close to MAPK1 (or ERK2), making MAPK1 and MAPK3 the best characterized mitogen-activated protein kinases amongst a total of 10 in the human genome (**Figure 27**).

MAPK3 consists of 379 amino acids and is 84% identical to MAPK1. Both isoforms appeared about 400 million years ago in the course of whole genome duplication in early vertebrates (reviewed in Roskoski 2012)[78].

1.3.2 Roles of MAPK1 and MAPK3

Despite an abundant literature on the mitogen-activated protein kinases MAPK1 and MAPK3, there are controversies as to whether they have unique physiological functions or whether they are redundant to reach a threshold of global MAPK activity. One way to discriminate the *in vivo* isoform-specific MAPK1 and MAPK3 functions is to study *Mapk1*- and *Mapk3*-knockout mouse models.

1.3.2.1 *Mapk1*-deficient mice

Whole-body *Mapk1*-deficient mice are early embryonic lethal[79] while neuronal-specific inactivation of *Mapk1* showed multiple phenotypes related to social behaviors, including elevated aggressive behaviors, deficits in maternal nurturing, poor nest-building, and lower levels of social familiarity and social interaction[80]. Neuronal-specific inactivation of *Mapk1* also showed a reduction in cortical thickness, impaired proliferation of neural progenitors and increased number of astrocytes[81]. Homozygous conditional inactivation of *Mapk1* in the developing neural crest resulted in lethality with embryos exhibiting a spectrum of major craniofacial defects including a shortened maxilla, mandibular hypoplasia, and cleft palate[82].

1.3.2.2 *Mapk3*-deficient mice

Whole-body *Mapk3*-deficient mice are viable and they showed subtle but enhanced phenotypes such as enhanced synaptic plasticity and memory improvements[83,84]. No neuroanatomical defects were reported in these mice[85]. Other studies suggest better insulin sensitivity and resistance to high-fat diet induced obesity of *Mapk3* knockout mice[86], increased locomotor activity[83,85] and better performance in active and passive avoidance tasks[83]. One study reported reduced level of thymocyte maturation in *Mapk3* knockout mice, suggesting its pleiotropic role in the immune system[87].

Interestingly, an increased level of MAPK1 activity has been shown in *Mapk3* adult mutants suggesting that MAPK1 may compensate MAPK3 deficiency or that MAPK3 is a repressor of MAPK1 activity[83,84]. In a similar way, there has been a report indicating that phosphorylated MAPK3 levels were fourfold higher in conditional *Mapk1*-deficient

mice during embryogenesis at E16.5 although the expression level of MAPK3 was not elevated[81].

1.3.2.3 Double *Mapk1::Mapk3*-deficient mice

Another way to address the same question of whether MAPK1 and MAPK3 are functional homologs or whether they have distinct roles, is to study double knock-out mice deficient for MAPK1 specifically in neurons (since whole-body *Mapk1*-deficient mice are lethal) with ubiquitous homozygous deletion of *Mapk3*. Satoh and colleagues have compared the phenotypes of these mice with those of single knock-out mice and found that the ventricular zone and the corpus callosum were thinner in the double knock-outs while deficiency specifically in *Mapk1* caused milder phenotypes, suggesting that both MAPK1 and MAPK3 are required to ensure proper brain development[80].

To summarize, while MAPK1 and MAPK3 share 84% amino acid identity and have a very similar substrate profile, there are strikingly different phenotype between *Mapk1*- and *Mapk3*-deficient mice. It is thus possible that these kinases both compensate for each other as well as having specific roles. This is an ongoing debate in the field (reviewed in Busca *et al.* 2016)[88]. Interestingly, the N-terminal domain of MAPK3 have been shown to be responsible for the functional differences with MAPK1 [88].

1.3.3 ERK1/2 signaling pathway

The extracellular signal-regulated protein kinases ERK1 and ERK 2, encoded by *MAPK3* and *MAPK1* respectively, are hydrophilic non-receptor proteins that are involved in the Ras-Raf-MEK-ERK signal transduction cascade, also known as the ERK signaling pathway (**Figure 28**). ERK1/2 signaling pathway represents one of the principal signaling cascades mediating the transmission of signals from cell surface receptors to cytoplasmic and nuclear effectors[89].

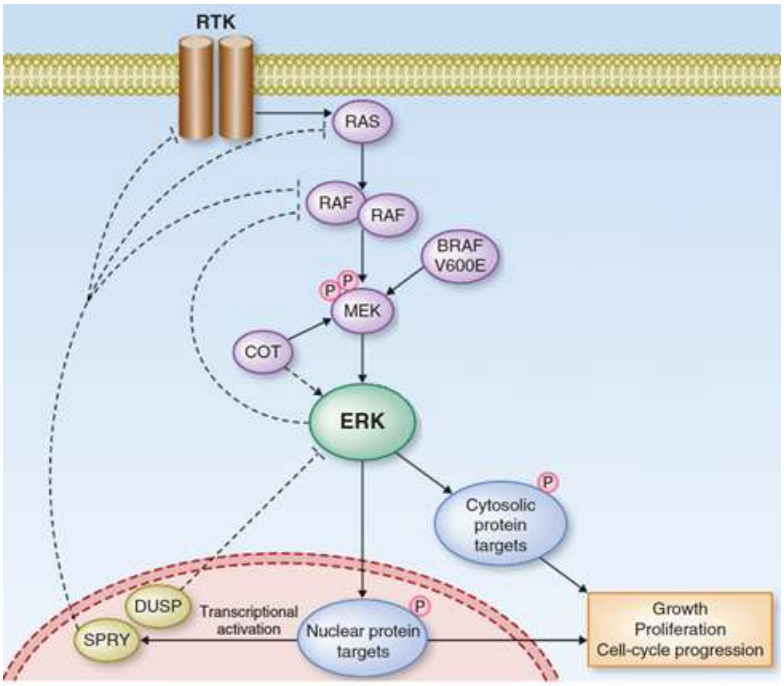


Figure 28 Representation of the Ras-Raf-MEK-ERK signal transduction cascade.

A detailed description of the pathway is described below in the main text. RTK, receptor tyrosine kinase; RAS, rat sarcoma; RAF, rapidly accelerated fibrosarcoma; MEK, MAPK/ERK Kinase; ERK, extracellular signal-regulated protein kinase; DUSP, dual-specificity phosphatase; SPRY, sprouty. Adapted from Nissan *et al.* 2013.

The main activators of the ERK signaling pathway are extracellular growth factors that activate receptors in the cell surface of the cell. These receptors are typically classified as receptor tyrosine kinases (RTKs). The first RTKs to be discovered were epidermal growth factor (EGF) and nerve growth factor (NGF) in the 1960s, but the classification of receptor tyrosine kinases was not developed until the 1970s. Later on, when the human genome was fully sequenced in 2001[90], 90 unique tyrosine kinase genes were annotated.

The activation of these receptors, in turns, activates RAS through guanosine triphosphate (GTP), notably H-RAS, K-RAS or N-RAS. RAS protein members belong to a class of protein called small GTPase. RAS genes were initially discovered from studies of cancer-causing viruses in rats during the 1960's, hence the name RAS for **ra**t **s**arcoma.

Activated RAS-GTP then leads to the activation of the protein kinase activity of RAF kinase family by an intricate multistage process that involves homodimer and heterodimer formation. The first RAF gene (V-RAF), identified in 1983, was isolated

from mouse retrovirus and was demonstrated to transform mouse fibroblasts to cancerous cell lines, hence its name **v**irus-induced **r**apidly **a**ccelerated **f**ibrosarcoma. The RAF kinases have restricted substrate specificity and catalyze the phosphorylation and activation of two mitogen-activated protein kinase kinases.

These enzymes described for the first time in 1995[91], are commonly known as MEK1 and MEK2 but they also have other names such as MAPKK1 and MAPKK2 with the acronym MEK stemming out from **M**APK/**E**RK **K**inase. MEK1 and MEK2 are dual-specificity protein kinases that mediate the phosphorylation of tyrosine and threonine in MAPK3 (ERK1) and MAPK1 (ERK2), respectively, which are their only known physiological substrates. This phosphorylation activates MAPK1/3, which are protein-serine/threonine kinases.

In the technical sense, RAF, MEK, and MAPK are all mitogen-activated kinases, but unlike the RAF kinases and MEK1/2, which have narrow substrate specificity, MAPK1 and MAPK3 have 284 documented downstream cytoplasmic and nuclear substrates[92]. This cascade participates in the regulation of a large variety of processes including cellular growth, cell cycle progression, cell survival, differentiation, metabolism, proliferation, and transcription.

Fine-tuning of the ERK signaling pathway is achieved through positive but also negative feedback loops. There are two types of negative feedback loops. Direct feedback loops implicate ERK-mediated phosphorylation of sites on RAF that directly inhibits its activation. Indirect feedback loops involve the transcription of SPRY (sprouty) and DUSP (dual-specificity phosphatase) genes upon the activation of ERK activity. SPRY downregulates RTK, RAS and RAF activation while DUSPs inactivate ERK through hydrolysis of phosphate. Taken together, it is therefore not surprising that any changes in the tightly regulated ERK signaling pathway lead to pathological diseases.

1.3.4 Relevance of ERK signaling in human diseases

Deregulation of ERK signaling can result from either germline gain-of-function or loss-of-function mutations. These mutations are linked to a group of rare human congenital disorders characterized by facial dimorphism, cardiac malformations, cutaneous and musculoskeletal abnormalities, and cognitive impairment (reviewed in[93,94]. ERK signaling pathway is also commonly deregulated in common human disorders

including cancers, brain ischemia, epilepsy[95–97] and has been linked to functional gene networks affected in ASDs[52].

1.3.4.1 Rasopathies and autism

The ERK signaling pathway is now recognized as a central hub involving many genes and CNVs associated with ASDs including the 16p11.2 microdeletion syndrome. The link between ERK signaling and autism was first identified in 1983[98]. In this study, Paul and colleagues reported a child with Noonan syndrome carrying mutations in components of the Ras-Raf-MEK-ERK signal transduction cascade, which presented autistic-like behaviors. Other syndromes having an effect on the ERK signaling pathway include Costello, Neurofibromatosis, DiGeorge, Fragile X, Rett and 16p11.2 microdeletion syndromes. These disorders have now collectively been referred to as "Rasopathies" owing to mutations in genes that encode components of the signaling pathway that are positioned either upstream or downstream of RAS. At the exception of DiGeorge and Rett syndromes, the level of ERK activity has been shown to be upregulated in human cells derived from patients diagnosed with these autistic syndromes (**Table 2**)[99].

Syndrome	Mutated Gene	Effect on MAPK Pathway
Noonan syndrome	PTPN11, RAF1, KRAS, NRAS	Upregulated Downregulated
Costello syndrome	HRAS	Upregulated
Neurofibromatosis	NF1	Upregulated
Cranio-facio-cutaneous syndrome	BRAF	Upregulated
DiGeorge syndrome	22q11.2del	Downregulated
Fragile X syndrome	FMR1	Upregulated
16p11.2 deletion	16p11.2del	Upregulated
Rett syndrome	MECP2	Downregulated

Table 2 Dysregulation of ERK activity in Rasopathies.
Adapted from Vithayathil *et al*. 2018.

1.3.4.2 Disease models for ERK pathway

Upregulation of ERK activity was also demonstrated in autism mouse models. For example, the 16p11.2 deletion mice exhibiting altered cortical neurogenesis that results in aberrant cortical circuitry associated with dysregulation of ERK signaling

pathway[75]. More specifically, the deletion mice show an increase of progenitor proliferation during early and mid-neurogenesis (E12.5 and E14.5) in the ventricular and subventricular zones, along with reduced number of Pax6+ radial glia and Tbr2+ intermediate progenitor cells. Enhanced progenitor proliferation and premature cell cycle exit are a consequence of altered levels of downstream ERK effectors cyclin D1 and p27 (*Kip1*) during mid-neurogenesis. Intriguingly, the deletion mouse model having only one copy of the Erk1 (*Mapk3*) gene, exhibit a paradoxical elevation of ERK phosphorylation in the cortex and hippocampus during mid-neurogenesis[75]. Another example includes the brain specific RAS-GAP, SynGAP, heterozygous knockout mouse model characterized by a deficient long-term potentiation in brain slices in the hippocampal CA1 region and an elevation of the baseline ERK activity[100]. Disease models and their corresponding features have been described in this review article[101].

1.3.4.3 Modulation of ERK signaling as a therapeutic target of autism

It is clear that the ERK biochemical pathway links together multiple genes implicated in autism spectrum disorders, which makes this pathway an interesting target for developing therapeutic interventions. Furthermore, the Rasopathies comprising syndromes with genetic mutations in components of the ERK pathway, have been associated with a significantly higher risk of autism[102]. In this study, 231 probands presenting various Rasopathies such as Noonan and Costello syndromes (see **Table 2** for other examples) were assessed for the core symptoms of ASDs and presented a 3-fold enrichment compared to the unaffected siblings[102]. Disease models of autism such as SynGAP and 16p11.2 microdeletion syndrome are almost all associated with higher levels of ERK activity.

For all of these reasons, the modulation of the ERK signaling pathway has emerged as therapeutic target to ameliorate autism symptoms and disease onset (reviewed in[103,104]). There are several routes whereby the ERK signaling pathway can be modulated. While RAS inhibitors has had limited success, for example in treating neurofibromatosis both in mice[105] and humans[106], ERK inhibitors that passes through the blood brain barrier has been recently tested in vivo in the 16p11.2 mice during gestation with great success[107]. Indeed, in this study, the authors reported a permanent rescue of the anatomical and behavioral phenotypes such as hippocampus dependent spatial and working memory[107]. Future work would be needed to determine the

temporal windows when the inhibition would be effective at the right developmental window as well as to determine the specific brain regions and different cell compartments where the inhibition would need to take place (reviewed in[108]).

1.4 The Major Vault Protein

The Major Vault Protein (MVP) is a protein encoded by the *MVP* gene. This protein auto-assembles into the largest cytoplasmic ribonucleoprotein (RNP) complex, named the "vault". Vault research was mainly conducted in the fields of structural biology and cancer biology. Although some associations were made in multi-drug resistance, nucleocytoplasmic transport, RNA transport and cellular growth, the physiological function of this complex remains to date unclear. In the present section, I will give a comprehensive review of the literature about the history of vault research, the structure and the putative function attributed to this marvelous, and yet still mysterious cellular organelle.

1.4.1 History of vault research

In 1986, the team of Leonard H Rome, a distinguished professor in biological chemistry, was working on isolation of clathrin-coated vesicles and discovered a new ribonucleoprotein particle by serendipity, not reported before[109]. Of a consequent size, largely superior to the size of the ribosome, this hollow barrel-shaped complex was intriguing and named "vault" in reference to its likeness with cathedrals ceilings. The team then had the genius idea to characterize this complex, and discovered that it was made mainly of a 104kDa protein named the Major Vault Protein (MVP), accounting for 70% of the total mass of the vault[109]. Since this major discovery, a series of studies were potentiated, led by the team of Leonard H Rome, but rapidly joined by different research groups worldwide. These studies are summarized in **Figure 29**.

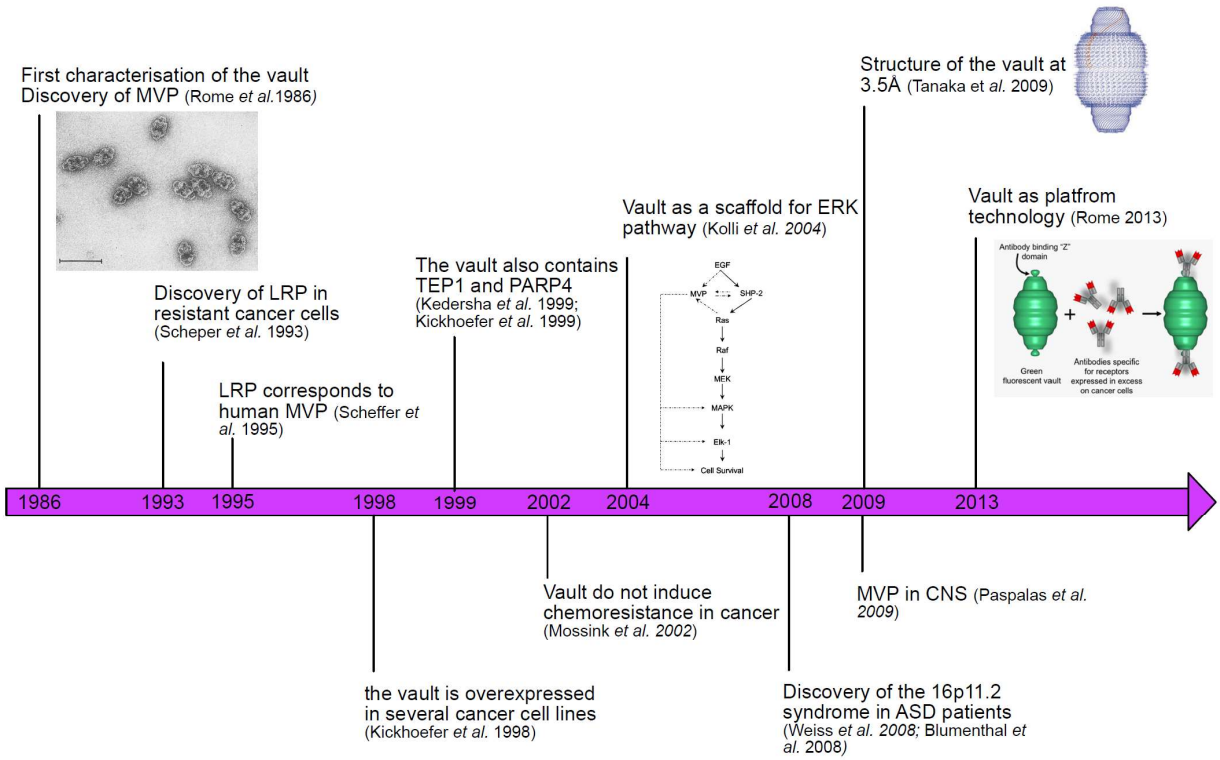


Figure 29 Timeline of major publications in vault biology.

In 1993, Scheper *et al.* observed that in some cancer cell lines (notably lung and breast cancer, fibrosarcoma and myeloma cell lines) which presented a resistance to chemotherapy, an unknown protein was overexpressed[110]. They named this 110kDa-protein LRP for Lung-Resistant Protein. Two years later, a publication signed by the same authors found out that LRP was actually corresponding to MVP[111]. This opened a new era in vault research investigating the possible functions of this complex in cancer resistance mechanisms.

Further characterization of this complex enabled to identify the three other components of the vault complex: the telomerase protein 1 (TEP1)[112], the Poly-ADP Polymerase 4 (vPARP)[113] and a small untranslated RNA[114].

In 2002, a research conducted by Mossink *et al.* on *Mvp* knockout mouse model, stated on the fact that, despite the previous encouraging putative involvement of MVP in chemoresistance, no tangible proofs were exposed to corroborate the direct effect of the vault in the sensitivity to diverse cytostatic[115]. The upregulations, observed both *in vitro* and *in vivo,* could be the result of the activation of other cellular mechanisms in which the vault would be involved. Since then, there has been numerous controversies

onto whether the level of vault could induce chemoresistance or not, with no consensus being reached to date.

In 2004, an elegant publication positioned the vault in MAPKinase pathway, by its interaction with SHP-2 in response to epidermal growth factor (EGF)[4]. Interestingly, both the MAPK3 and MVP genes are localized on the on the 16p11.2 autism-associated locus[58,70]. This opened new possibilities for the involvement of these two genes in cognitive diseases.

In 2009, the first implication of MVP/vault in the central nervous system was published by Paspalas *et al.* confirming the presence of MVP/vault in the cytoplasm of neurons and suggested several roles of the vault in this cell type[116]. During the same year, the structure of the vault was resolved at a resolution of 3.5Å, enabling to gain better knowledge on the assembly of its different constituents.

In more recent years, the vault was also considered as a platform for nanotechnologies. The idea is that the vault could be used as a container for cargo of interest such as fluorescent probes or chemotherapy, and be associated with antibodies, in order to direct the vault to specific cell types (ex: tumor cells)[117]. This idea led to the creation of a start-up directed by Leonard H Rome who first discovered the vault in 1986. The aim is to use the vault in cancer therapy. Indeed, the first clinical trial is ongoing and aims to cure lung cancer and lower the side effects of chemotherapy by targeting specifically the delivery of cytostatic near cancer cells.

1.4.2 Vault evolutionary biology

The vault is evolutionary highly conserved, suggesting that it has a crucial role in cellular pathways. Indeed, *MVP* has sequence homology in diverse species including mammals (*Rattus norvegicus,* in which it was discovered; *Mus musculus; Homo sapiens*), amphibians (*Rana catesbeiana* and *Xenopus laevis*), avians (*Gallus Gallus*), the sea urchin and the slime mold (*Dictyostelium discoideum)*[118]*.* Interestingly, instead of identifying classical sequence homology, a recent study used the power of *in silico* reconstruction to find structural features needed to form the vault in genome of several lower and higher eukaryotes. The authors realized that MVP/vault structure-homologs were present in the entire eukaryotic superkingdom, and consider that the last

eukaryotic common ancestor already possessed a protein presenting the same features as MVP/vault[119] (**Figure 30**).

Fungi and insects do not present MVP homologous sequence despite a clear MVP annotation in their common ancestor, which may reflect a loss of MVP/Vault function in these two groups.

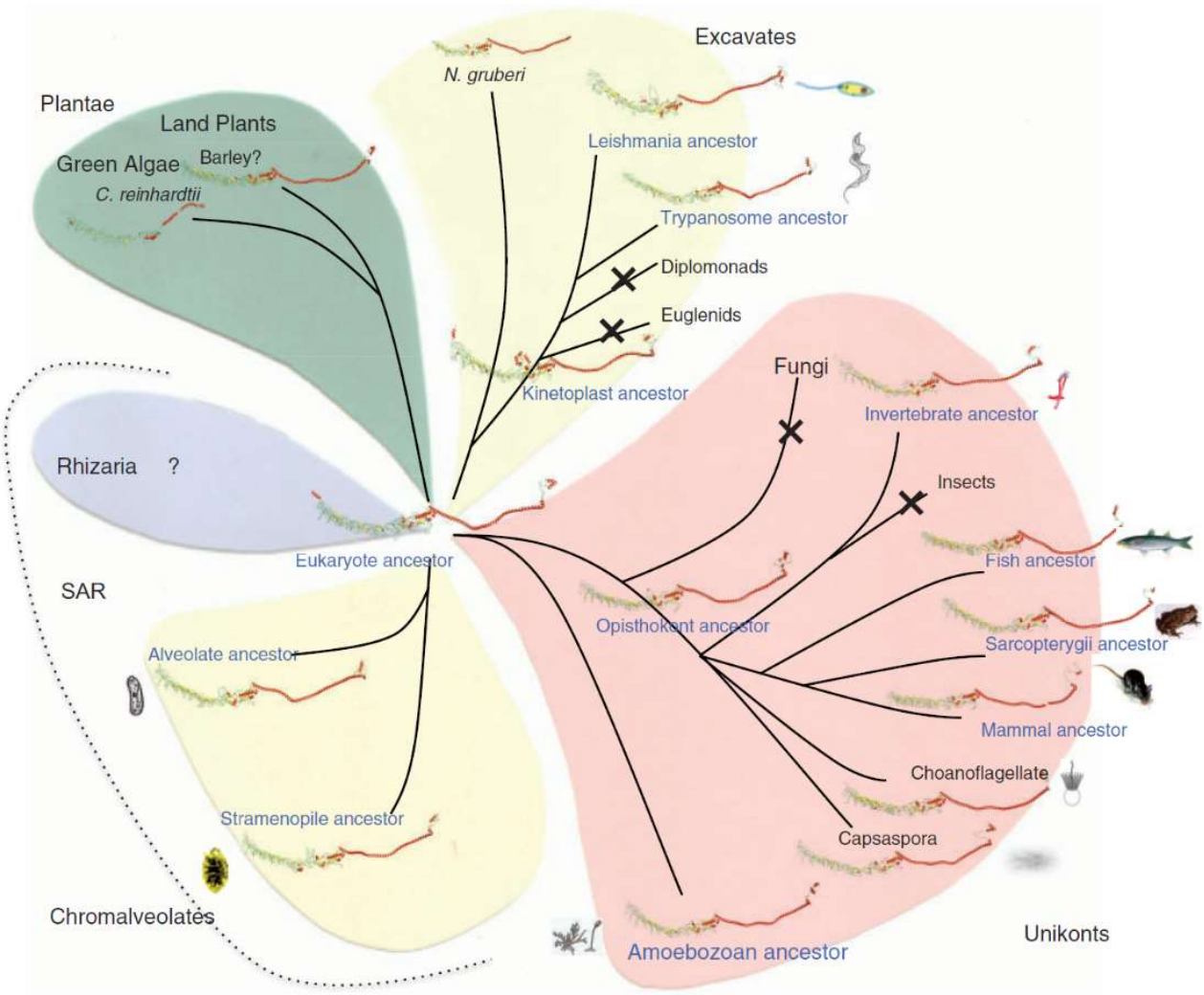


Figure 30 MVP/Vault evolutionary predictions.

Proteins presenting MVP-like structure are represented in a eukaryotic tree summarizing its presence in every embranchment. The putative MVP structure is represented for each taxonomic ancestor. Adapted from Daly *et al.* 2013.

1.4.3 MVP/Vault structure and composition

The vault is the biggest ribonucleoprotein particle ever describe in the cell, with a size of 13MDa that is approximately three time the size of a ribosome. Its shape forms a hollow barrel, formed by MVP, and encompasses two other proteins (vPARP and TEP1) and the untranslated vault RNA (vRNA).

1.4.3.1 General structure of the vault

The structure of the rat vault was extensively described over the last decades using more and more advanced technologies allowing resolving the specific features of this complex at ultra-high resolution. In 2009, Tanaka and collaborators succeeded to determine the x-ray structure of rat liver vault at 3.5 angstrom resolution (**Figure 31**)[120].

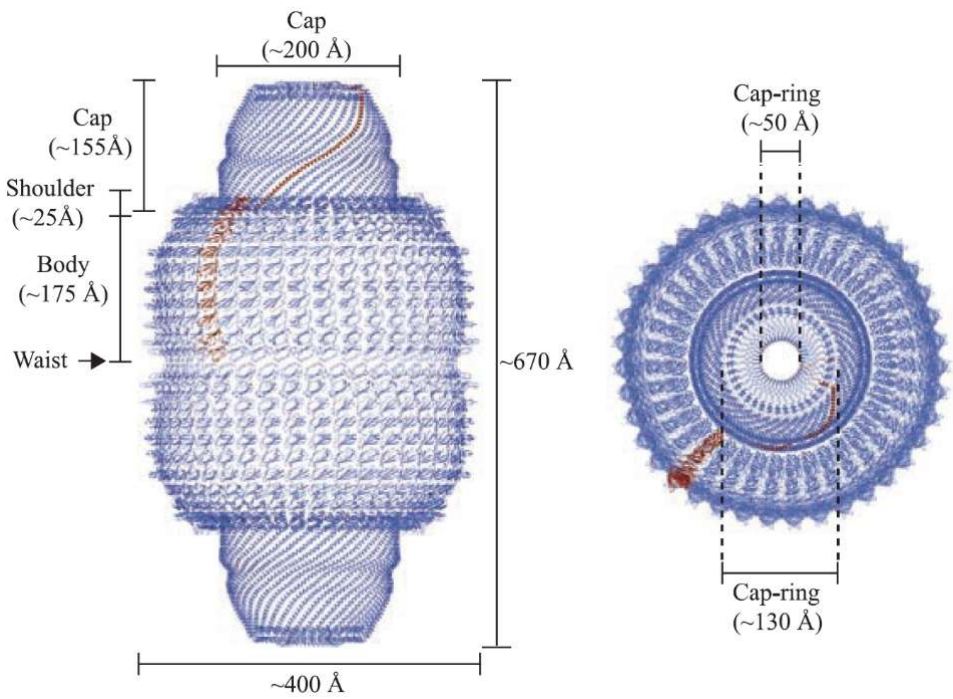


Figure 31 Overall structure of the vault.
Adapted from Tanaka *et al.* 2009.

The vault is composed of two identical cup-like structures joined at their open end (called the waist) to form a 39-fold symmetric barrel which measures ~670 Å in length and ~400 Å in maximum diameter. The barrel wall of only 15 to 25 Å in thickness encloses an internal cavity with the length of ~620 Å, and the maximum diameter of ~350 Å, large enough to enclose most objects found within the cell[120].

Each half vault is delineated into three regions. The body, with a length of 175Å, constitute the central wall surrounding the inner chamber. It is connected to the shoulder domain which is narrowing the diameter of the vault and lead to the cap region with a size of 155Åx200Å. At its extremity, the cap is forming a ring (outer diameter: 130Å; inner diameter: 50Å). The morphology and dimensions of the vault is similar in the different species described so far[118].

CryoEM revealed that vault could open and flatten into two half vaults resembling flower-like structures with eight petals symmetrically arranged around the central ring (**Figure 32**)[121].

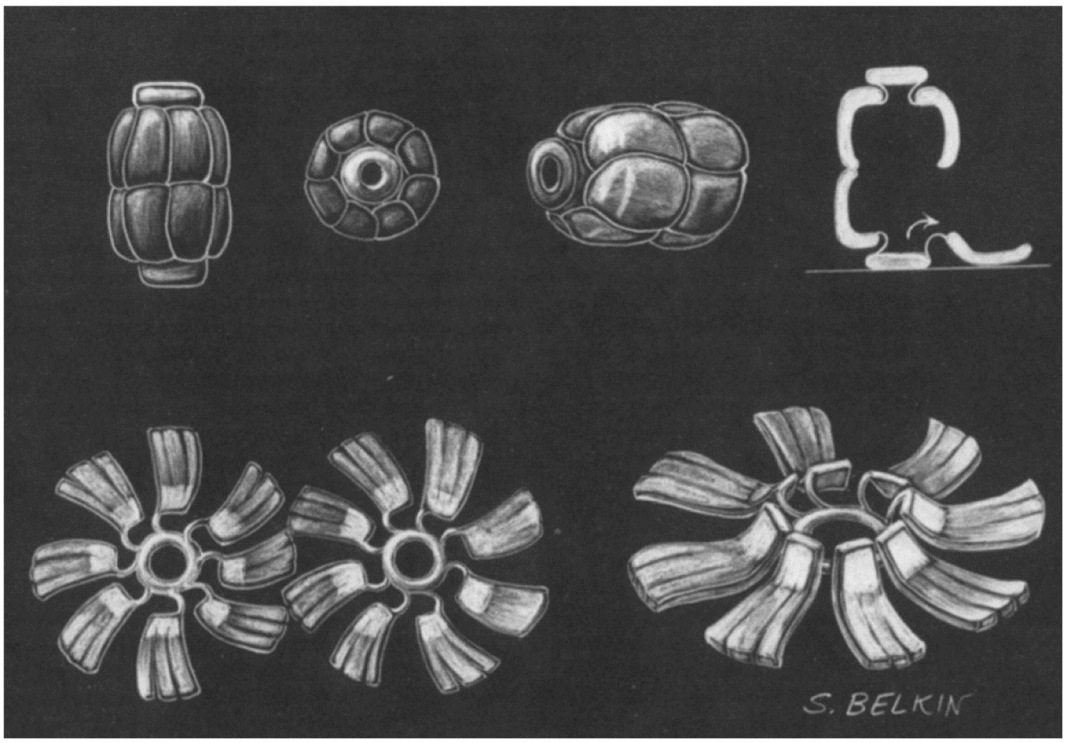


Figure 32 Model showing the opening of a vault into flower-like structures.
Adapted from Kedersha *et al.* 1991.

The authors also observed that modification of the pH triggered vault opening[121,122]. However, other treatments such as application of 10mM DTT, 2M urea, 1% Triton X-100, 1mM spermidine, 10mM ATP/GTP, heparin, NP-40, EDTA nor calcium chloride did not enhanced the flower-like conformation[123], thus suggesting a high stability of the vault. Tests on mechanical stability showed that the vault is highly adaptive to mechanical deformation and that its structure facilitate low energy conformational changes that reverse easily[124].

1.4.3.2 Description of the vault-shell element: MVP structure

It is important to highlight that the major vault protein is the only necessary and sufficient component to form the vault shell[125]. Therefore, its constitution is aimed to ensure the scaffold of the entire complex.

The rat MVP is a linear protein composed of 845 amino acids (**Figure 33**). The part of MVP monomer that is forming the body of the vault is composed of nine repeated globular domains, composed of α-helix and β-sheets, and going from the N-terminal to the amino acid 520. The shoulder domain, of 101 amino acids, enables to give the right bend to the structure and connects to the cap. The cap is formed by a 42-turn α-helix and end by the cap-ring, a hooked domain localized at the C-terminal of the protein[120].

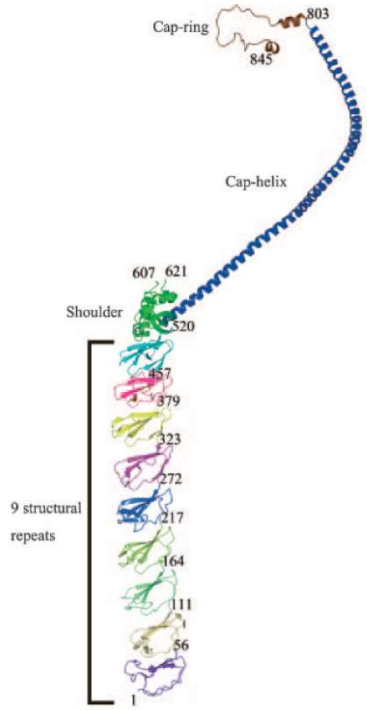


Figure 33 Stereoscopic ribbon drawing of the overall fold of an MVP monomer.

The MVP monomer is folded into nine structural repeat domains, a shoulder domain, a cap-helix domain, and a cap-ring domain. Adapted from Tanaka *et al*. 2009.

Two neighbor MVP monomers are bound together side by side through hydrophobic interactions mainly present in the cap helix domain, which constitute the most stable region of the vault. The two half-vaults are bound together at the waist by intermolecular antiparallel β sheet from two domain 1 in the N-term of opposite MVP monomers. These interactions, of lesser strength are the site of the opening of the shell[120].

1.4.3.3 Description of the other vault components: vPARP, TEP1 and vRNA

Beside MVP, the vault is also composed of three other minor elements: the vault-associated Poly ADP-ribose polymerase 4 (vPARP), the Telomerase protein 1 (TEP1) and an untranslated vault RNA (vRNA).

The vPARP is a 193kDa protein which was described for the first time because of its association with the vault[113]. It possesses an enzymatic activity and the ability to catalyze the polymerization of long poly-(ADP)-ribose chains, resulting in changes in interactions of the substrate with other proteins or DNA. Unlike other PARP family members, vPARP do not seem to be activated by DNA damage or involved directly in DNA repair mechanisms[113].

The second minor vault protein is TEP1 (240kDa), a protein also known to be a part of the telomerase complex[112]. TEP1 do not present any telomerase activity but contains a WD40 propeller-like structure, which may act as a platform. Within the vault it was shown that its N-terminal binds to vRNA and stabilizes it[112].

The last vault components are untranslated small RNAs, which represent only approximately 5% of the total vault mass. The vRNA are transcribed by RNA-polymerase III and can vary in nature and number in a species-specific manner. Indeed, while mice and rats present only one vRNA of 141bp, human present three distinct vRNA (*hvg* 1-3) of shorter size (88-98bp)[126]. Despite a low sequence homology between species, the secondary structure of vRNA is highly conserved.

The three minor vault components are integrated inside the core MVP-homopolymer. Difference in density map observed in vault treated with RNase indicate that the vRNA is localized at the extremity of the vault. It consists of a 15 x 15 x 30Å^3 circular column and an 82 x 82 x 20Å^3 circular disc (colored in pink in **Figure 34**), which interact with TEP1, also located in the cap (blue in **Figure 34**). TEP1 shape consists of a 100 x 100 x 80Å^3 circular column and a 45 x 45 x 30Å^3 cylinder, connecting with both the cap-ring and vRNA. Each half-vault is also hosting four copies of vPARP which interact with domains 4 of MVP, hence their localization in the inner side of the barrel (orange in **Figure 34**)[127].

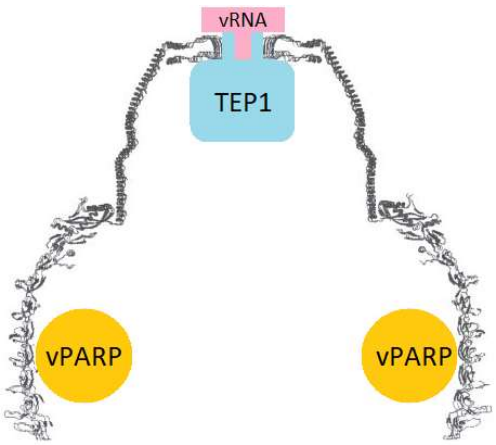


Figure 34 A vertically sliced section of a half-vault structure.

The capsid structure consisting of MVP molecules is shown in a ribbon representation. Yellow circles denote VPARP molecules on the inside of the R4 repeat domain; TEP1 (blue) in the cap region of v vault slot together with the vRNAs (pink). Adapted from Tanaka *et al*. 2012.

1.4.4 MVP/Vault life cycle

1.4.4.1 MVP transcription

MVP gene promoter was described in 2000 in SW-1573 lung cancer cell line[128]. The amplification and analysis of the 5' upstream *MVP* genomic region revealed a promoter-active region with the presence of transcription factor binding sites. One of them consists of a Y-box, which can be activated by binding of YB-1 or NF-Y, two transcription factors that are known to be transferred in the nucleus after exposition of cytostatic agents[129]. MVP promoter region also presents Sp1, p53 and signal transducer and activator of transcription (STAT) binding sites, also known to be involved in regulation of multidrug resistance (ex: MDR1 gene), cell cycle and apoptosis[128]. This promoter region, as well as other regulatory elements, enables the docking of RNA-Polymerase II to the genome and the transcription of the pre-mRNA, which contains 15 exons.

1.4.4.2 MVP translation and vault-shell assembly

After splicing, the mRNA is up-taken from the nucleus to the cytoplasm where it is translated. This process is done by multiple ribosomes that combine on a single mRNA (**Figure 35A and B**) to form a polyribosome and translate simultaneously several

copies of MVP (**Figure 35C**)[130]. An intriguing fact is that MVP monomers do not separate from the polyribosome once completely translated. Instead, two MVP copies from consecutive ribosomes associate together while being translated (**Figure 35D**). The dimer then associates with subsequent dimers, and so on, until the translation of a set of 78 copies of MVP is completed (**Figure 35E**). Of note, the nine repeated domains at the N-terminal take their conformation in an independent manner, and then the shoulder give the needed curve in order to make two MVP monomer to dimerise, and then the translation of the cap helix give the proper tension to allow the formation of the vault shell. MVP translation thus occurs in a rigorously tight manner preventing nascent MVP from degradation which confirms previous observation that MVP do not exist as monomers at any given time within the cell[130].

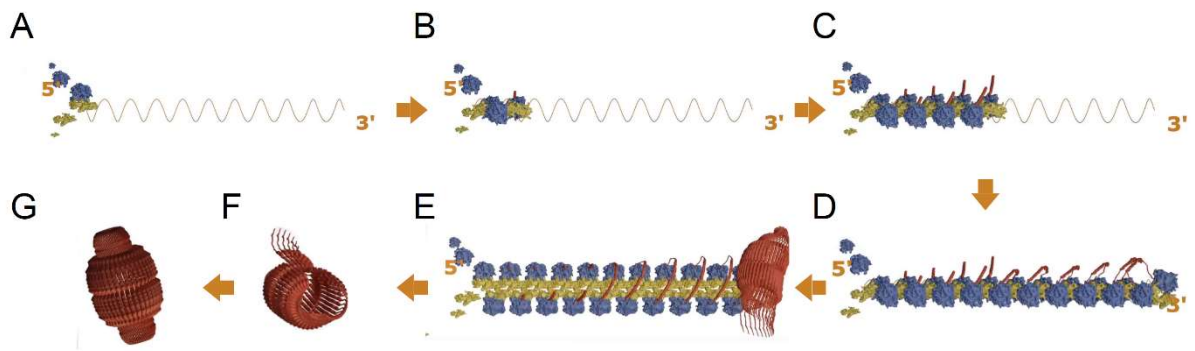


Figure 35 Polyribosomes contributes to vault assembly from *Mvp* mRNA.

The two subunits of a ribosome (blue and yellow particles) assemble on MVP mRNA (single red curve) (**A**) and start translation while other ribosomes joined the polyribosome (**B**). Successive ribosomes co-translate MVP as they progress along MVP mRNA (**C**). Successive MVP monomers form dimer, joined by their N-terminal, and then associates with the next dimer and so on (**D**), thus resulting in progressive vault assembly and integration of each MVP monomer while translation occurs (**E**). When the 78 copies are translated, the newly formed vault pinch-off from the polyribosome (**F**) and complete its assembly to form the vault (**G**). Adapted from an explicative video on http://vaultpharma.com.

Vault assembly is very closely linked to MVP translation as shown in **Figure 35**. Indeed, once the translation of the 78 copies of MVP is completed, the newly formed vault separates from the polyribosomes (**Figure 35F**) and completes its assembly in a final step (**Figure 35G**)[130]. It is important to mention that each vault is made from the same *MVP* mRNA.

1.4.4.3 Vault components integration

Once the vault self-assembled, the other vault components (vPARP, TEP1 and vRNA) need to be integrated. These processes are not yet fully understood but when we look at the structure of vault-shell, the only readily visible accesses to its interior are the holes formed by the cap ring at both the extremities of the vault. However, its relatively small inner diameter of approximately 50Å and the highly stable interaction of the α-helix at the cap, do not allow the entrance of big molecules, such as vPARP or TEP1.

Furthermore, the vault is not as rigid as primarily expected and allow dynamic exchange processes, through creation of pivotal doors between opposite petals to allow the integration of the particle components[121]. The way TEP1 is targeted to the vault are not yet known, however the presence at the N-terminal of the TROVE module, a part of the p80-homologous domain, is sufficient for its binding with the vRNA and integration in the cap[131]. vPARP insertion is governed by its INT domain, located at the C-terminal, which interacts with the third and fourth MVP domains at the inner core of vault shell and allow its installation[117].

1.4.4.4 MVP/Vault degradation

MVP monomers are not present in the cytoplasm of cells, since they are directly incorporated within the vault. The highly stable structure of the complex makes it difficult to access for degradation processes, which may lead to low turnover, and non-linear association with the amount of *Mvp* RNA and protein within a cell. Indeed, MVP/Vault present a relatively long half-life *in vivo* (up to three days)[132].

To my knowledge, only one publication stated on this subject. Sutovsky and collaborators, a team working on porcine and human oocytes, discovered that MVP levels were increased in the presence of proteasome inhibitors[133]. They demonstrated that MVP was polyubiquitinylated and then associated with the proteasome and degraded by proteolysis. No data about MVP degradation through autophagy was reported so far.

1.4.5 MVP/Vault function

1.4.5.1 MVP/vault expression pattern and cellular distribution

MVP/Vault are found in the cytoplasm of most eukaryotic cells, as reviewed by Scheffer *et al.*[134]. However, MVP/vault seem highly enriched in some specific tissues as scavenging cell types, such as macrophages or microglia[135] or epithelial cells. This may reflect specific functions needed in these kind of cells like detoxification processes[136]. Vaults are abundant in the cytosol of mammalian cells and is uniformly distributed in the cytoplasm with a punctate pattern (**Figure 36A**)[137].

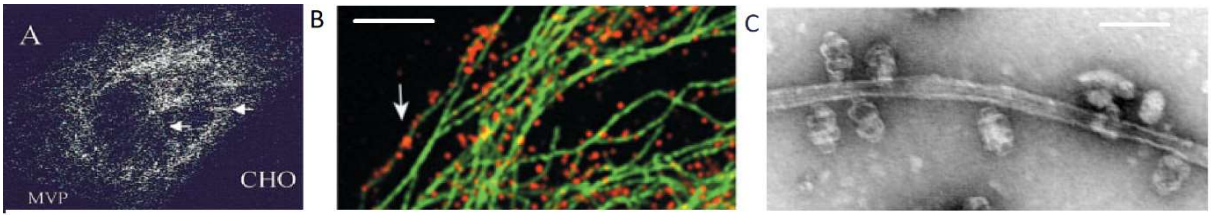


Figure 36 MVP cytoplasmic expression and association with microtubules.

MVP labelling in CHO cells (A). MVP (red) colocalize with microtubules (tubulin in green) in HeLA cells (B). CryoEM micrographs of vaults binding to the microtubules directly with their cap in vitro (C). Adapted from Herrmann *et al.* 1999. and Eichenmuller *et al.* 2003.

MVP/vault was shown to interact directly with microtubules (**Figure 36B and C**)[138], and exhibit anterograde and retrograde transport along them[139].

Mammalian MVP has previously been reported to be associated with the nucleus (about 5% of the total amount), particularly its cytosolic surface on which vaults are thought to dock at or near the nuclear pore complex[140].

Mice deleted of MVP are fertile without any apparent developmental defects[115,141,142]. However, it was associated with diverse cellular pathways including malignant transformation[143], differentiation and senescence[144], anti-viral immunity[145], chemotherapy resistance[146] and more recently synaptic plasticity[147]. A review of known MVP/vault functions is summarized in Berger *et al.*(2009)[148].

1.4.5.2 MVP/vault in multi-drug resistance

Drug resistance is one of the main causes of cancer treatment failure and this physiological mechanism, initially aimed at protecting the cell from xenobiotic attacks,

are an obstacle to the implementation of effective therapeutic strategies. Thus, the comprehension of the process that prevent the efficacy of chemotherapy has been extensively studied.

Numerous studies identified MVP/vault as a factor of multi-drug resistance in various cancer types such as lung cancer[134], breast cancer[149,150], hepatocarcinoma[151] and glioblastoma[2,143]. Indeed, overexpression of MVP is correlated with a poor response to chemotherapy, and MVP/vault knockdown by siRNA in human bladder cancer cells inhibited doxorubicin (a common chemotherapeutic drug) sequestration and enhanced its accumulation in the nucleus[152]. The subcellular localization of MVP/vault near the nuclear pore argue for its role in the trapping and the efflux of the drug[153].

Despite these *in vitro* observations, upregulation of MVP/vault was shown not to be sufficient to induce multi-drug resistance, and its disruption in mice did not induce chemotherapeutic hypersensitivity neither[115]. Thus, the direct link between MVP/vault and multi-drug resistance have not been fully established yet, and may stem out require more complex cellular process.

1.4.5.3 Vault act as a scaffold protein for MAPKinase pathway

The MAPKinase pathway is involved in many cellular processes such as cell survival, proliferation and growth through the integration of extracellular signals (see **1.3 Mitogen-activated protein kinase 3 (MAPK3)**).

The first indication that MVP might regulate MAPK pathway was reported by Kolli *et al.*[4]. In this study, the authors demonstrated that MVP/vault was a basal substrate of SHP-2, a tyrosine phosphatase which acts upstream of Ras[154]. This enzymatic interaction is triggered by the Protein Tyrosine Phosphatase (PTP) domain of SHP-2. Upon EGF stimulation, the substrate-enzyme complex is dissociated thus leading to an increase of MVP/vault tyrosyl-phosphorylation level. In this state, MVP/vault is able to recruit other members of MAPKinase pathway, as Ras or ERK1/2, but also SHP-2, this time in a non-enzymatic manner by interacting with its SH2 domains. This recruitment leads to the increased activation of Elk-1, the transcription factor which is one the final effector of MAPKinase pathway (**Figure 37**) [4].

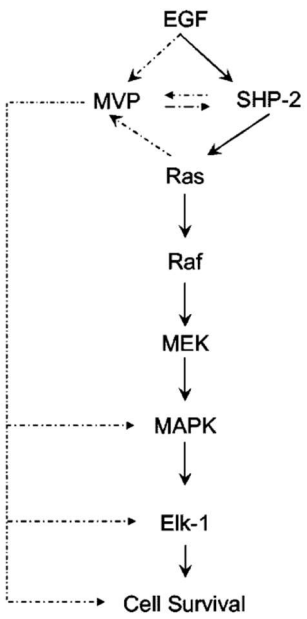


Figure 37 SHP-2 regulation of MVP phosphorylation in EGF-mediated signalling.

SHP-2 is activated by EGF and activate MAPKinase pathway through Raf. MVP/vault is activated by SHP-2 and then act as a scaffold protein by recruiting SHP-2 and MAPK, thus modulating Elk1 activity and cell survival. Adapted from Kolli *et al.* 2004.

MVP/vault has also been shown to interact with the tyrosine kinase, Src (also containing SH2 domain) and to regulate ERK subsequent to EGF stimulation (Kim et al. 2006). Another report, identified a new MVP/vault interacting partner involved in MAPKinase pathway[155]. In this study, MVP/vault inhibits YPEL4 thus resulting in a decrease of Elk1.

Previous studies done on *Dictyostelium* indirectly support the implication of MVP/vault in growth factor signal transduction[156]. In this study, the authors observed no morphological nor developmental defects in between WT *Mvp*-deleted *Dictyostelium* on normal culture condition. However, when placed under nutritional stress the disruption of MVP/vault triggered cell growth defects, possibly through MAPKinase pathway. Taken together, MVP/vault acts as a scaffold for MAPKinase pathways and modulates its fine-tuned regulation.

1.4.5.4 MVP/Vault function in the nervous system

Neurons are highly specified cells, in which a tight spatial and temporal regulation is required to ensure proper function[157]. Three studies addressed the putative role that MVP/vault in the brain and more specifically in neurons.

First, *in vitro* study in PC12 cells treated with nerve growth factors, showed that MVP/vault may be active nucleocytoplasmic shuttles implicated in intracellular transport[137].

Second, Paspalas and collaborators confirmed these observations ten years later and described the expression pattern of MVP/vault along the nucleus-neurite axis in rat primary neuronal cultures (**Figure 38A**)[116].

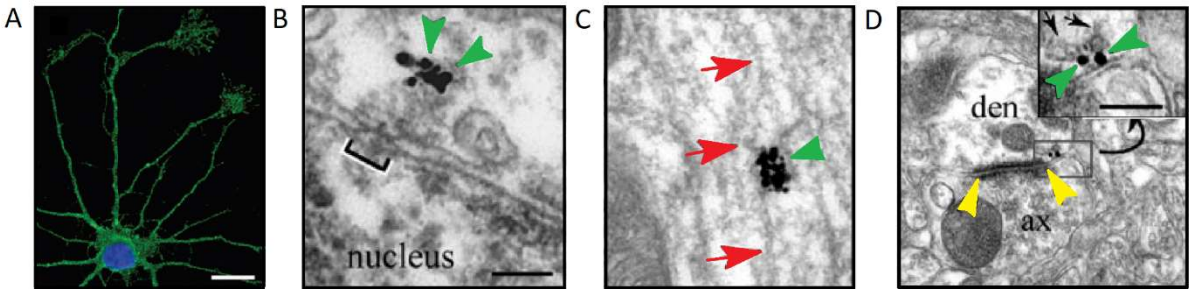


Figure 38 MVP expression in rat neurons.

MVP/vault staining in cortical neuron cultures in the cytoplasm from the soma, neurites and growth cones (**A**). Immunogold labelling of MVP/vault (green arrowhead) close to nuclear pore (bracket) (**B**), along microtubules (red arrow) (**C**) and at the synapse, but not directly associated with the synaptic specification (delimited by the yellow arrowheads) (**D**). Adapted from Paspalas *et al.* 2009.

MVP/vault staining is present in the cytoplasm of cortical neurons from the soma to the growth cones. Then they use immunogold labelling of the vault to observe its distribution pattern at a higher magnification with electronic microscopy, on rat brain slices. They revealed that MVP/vault localized in the vicinity of the nuclear pore (**Figure 38B**), along the neurites in close association with the microtubules (**Figure 38C**) and at the synapse, in both pre- and postsynaptic compartments (**Figure 38D**). However, they did not find labelling directly in the synaptic specialization, thus suggesting that MVP/Vault are not implicated directly in synaptic transmission (**Figure 38D**). Then they performed isolation of synaptosomal fractions and proceeded to a pull-down assay, with or without RNase treatment, and their results suggested that the vault may be protective to and helped in the transport of mRNA to the synapse[116].

Third, a more recent study associated MVP/vault with synaptic plasticity in the mice[147]. They used monocular deprivation as a model for studying plasticity in the primary visual cortex in *Mvp*-heterozygous juvenile mice. After 7-days of eyelid-suture, they observed

that *Mvp* heterozygous mice present a decrease in ocular dominance plasticity, as well as a reduction of the number of functional synapses, with reduced levels of AMPA receptors and dysregulated ERK signaling. This demonstrate a specific role for MVP as a key molecule influencing the homeostatic component of activity-dependent synaptic plasticity[147]. Interestingly, no differences were observed between *Mvp*-heterozygotes and WT, in mice that were not submitted to the monocular deprivation, thus suggesting that the manifestation of *Mvp*-deprivation phenotypes only appear in challenging conditions.

2. Thesis rationale

With the advances of next generation sequencing technologies, the impact of genetics is more and more recognized in the field of neurodevelopmental disorders. In autism spectrum disorders (ASDs), hundreds of different genetic risk factors have already been found in human patients. A large majority of these genetic causes are due to Copy Number Variations (CNVs), often containing many potentially causative genes.

The human 16p11.2 locus is susceptible to a 600 Kb deletion (OMIM #611913) which is among the most frequent genetic cause in autistic population. Besides autism, patients bearing the 16p11.2 microdeletion also present penetrant obesity and neuroanatomic alterations. Such syndromic clinics levers the question of common pathogenicity between symptoms, but also questions about underlying genetic architecture. How many genes drive the phenotypes? Is there one key gene or several genes responsible for the clinical manifestations? These are important questions I will tackle in my thesis.

So far, drosophila and zebrafish were used as attractive animal models towards the screening of causative genes and the understanding of genomic interactions. However, these animal models remain evolutionary far from humans and present drastic limitations for further exploration into underlying pathophysiological mechanisms including gender differences. To the contrary, the mouse extensively used in biomedical research is an evolutionary more relevant model. This is where my doctoral work will contribute.

The first aim of my PhD is to decipher the implication of each of the 30 protein-coding genes within the 16p11.2 locus on brain morphology by using a heterozygous approach in single-gene mouse models. This is now achievable thanks to the growing numbers of knockout mouse models available in the scientific community. To do this, I will take advantage of the quantitative and highly precise neuroanatomical protocol developed in my research team. This is an original gateway to grasp, *in vivo*, the implication of the genes constitutive of a CNV in neuroanatomical phenotypes. Among the 30 genes of the interval, *Mvp*, *Mapk3* and *Kctd13* received a particular focus, since previous research suspected their common implication in 16p11.2 pathophysiology.

The second aim of my PhD is to characterize the function of the Major Vault Protein (MVP) in brain morphology. MVP is the main component of the vault organelle, which function remains largely unknown to date. I thus decided to focus on its role in the central nervous system, in order to understand the potential link between the vault organelle, brain anatomy and ultimately autism spectrum disorders.

3. Scientific contribution

My main objective was to search for neuroanatomical defects in mice deleted for genes pertaining to the 16p11.2 locus in order to identify main drivers and subsequently study how main drivers such as the major vault protein could alter brain morphology.

I was involved in every aspect of the neuroanatomical studies, at an intellectual as well as technical level. This implied the daily management and analysis of data as well as the supervision of interns and Master students that teamed up with me in the lab.

I conducted the detailed characterisation of *Mvp* mice including expression profiling (RT- and qRT-PCR, western blot, immunohistofluorescence), follow-up neuroanatomical studies, primary neuronal cultures and Golgi-Cox staining. Although I did not carry out the experimental behavioural assays, I was very much involved in the statistical analysis of the behavioral data, their interpretation and visualization.

For collaborative studies (RNAseq, electrophysiology, electronic microscopy, phospho-ERK staining), I was involved in the rationale of the experiment, in sample preparation, data analysis and integration.

I was also involved in the management and genotyping of the different mouse lines housed in our animal facility (*Mvp*, *Mapk3*, *Kctd13*, *Ino80e*, *Doc2a*, *Tbx6* and double-heterozygotes).

My findings and the methods I used to generate the data are presented in the main article of my thesis work which I included in the next section.

Beside my main work on the 16p11.2 project, I was also involved in three side projects within the lab: 1) I contributed to the adaptation of a standardised neuroanatomical protocol from the coronal to the sagittal plane[158], 2) I participated to the neuromorphological screen of 26 WDR genes implicated in intellectual disability (WDRopathies)[159], and 3) I investigated the role of KPTN in primary macrocephaly (paper in preparation). These results are not reported in this document.

4. Main Publication

Unravelling the implication of the major vault protein in neuroanatomical phenotypes

Perrine F. Kretz,[1] Christel Wagner,[1] Charlotte Montillot,[2] Sylvain Hugel,[3] Ilaria Morella,[4] Meghna Kannan,[1] Anna Mikhaleva,[5] Marie-Christine Fischer,[1] Maxence Milhau,[2] Riccardo Brambilla,[4] Yann Herault,[1,6], Alexandre Reymond,[5] Mohammed Selloum,[1,6] Stephan C. Collins,[1,7] Binnaz Yalcin,[1,7,*]

[1]IGBMC, UMR7104, Inserm, U964, 67400 Illkirch, France

[2]Inserm UMR1231, University of Bourgogne Franche-Comté, 21000 Dijon, France

[3]INCI, UPR3212, CNRS, 67000 Strasbourg, France

[4]NMHRI, Division of Neuroscience, Cardiff University, CF24 4HQ Cardiff, UK

[5]CIG, University of Lausanne, CH-1015 Lausanne, Switzerland

[6]CELPHEDIA, PHENOMIN, ICS, 67400 Illkirch, France.

[7]Current address: Inserm UMR1231, University of Bourgogne Franche-Comté, 21000 Dijon, France

[*]Correspondence: binnaz.yalcin@inserm.fr

4.1 Abstract

Major Vault Protein (MVP), the main component of the vault organelle, is a highly conserved protein found in higher and lower eukaryotic cells, yet its function is not understood. Using mouse genetic studies, we set out to identify which of the 30 genes causes brain size and other NeuroAnatomical Phenotypes (NAPs) at the autism-associated 16p11.2 locus. Here we show that multiple genes regulate brain size (*Mvp*, *Ppp4c*, *Zg16*, *Taok2*, *Slx1b*) in contrast to previous studies, with female significantly less affected. While we find MVP expression highly specific to the limbic system, *Mvp* stood out as the top driver of NAPs, regulating the morphology of neurons, postnatally and specifically in male. Finally, we demonstrate that the double deletion *Mvp*::*Mapk3* rescues NAPs and alters behavioral performances, suggesting that MVP and ERK share the same pathway, *in vivo*. Our results highlight that sex-specific neuroanatomical mechanisms must be considered in neurological disorders such as autism and provide the first evidence for the involvement of the vault organelle in the regulation of the mammalian brain size.

Keywords

The vault organelle, major vault protein, mouse genetics, brain neuroanatomy, ERK pathway, sex differences, autism

4.2 Introduction

Autism spectrum disorders (ASDs) are a group of complex neurodevelopmental diseases characterized by restricted/repetitive behaviors and a deficit in social communication. Affected children usually express autistic behaviors after 24 months of age. ASDs are well known to be sex-biased with four males diagnosed for every female[160]. Apart from a recent study on the impact of androgens on human neural stem cells[21], the underlying sex-specific biology remains largely unknown.

The human 16p11.2 locus is susceptible to a 600 Kb deletion (OMIM #611913) which is among the most frequent known etiologies of ASDs[58]. This 16p11.2 deletion arises *de novo* and partly associates with brain size and other neuroanatomical traits including a decreased cortical thickness restricted to male carriers[64]. One of the main challenges in the field is to be able to decipher which of the 30 protein-coding genes harboring this region could be the causative gene underlying NeuroAnatomical Phenotypes (NAPs) at the locus and how the genes could interact with each other leading to ASDs.

To evaluate the contribution of each gene within the 16p11.2 locus, a systematic morpholino-mediated knockdown method in the zebrafish model identified *KCTD13* (potassium channel tetramerization domain containing 13) as a major driver of brain size[5]. However, recent studies in mice did not confirm this association[161,162], questioning the relevance of the zebrafish model in this particular syndrome.

Intriguingly, *MVP*, which encodes the 100kDa major vault protein in the interval, was found to modify NAPs via *cis* genetic epistasis with *KCTD13*[5,161]. Described in 1986 by Leonard H Rome and named after its arched shape reminiscent of the ceilings of cathedrals, the vault is a highly abundant and conserved organelle present in many eukaryotic cells[109]. The major vault protein is the main constituent of the vault organelle and is sufficient to give the vault its hollow barrel-like shape[125]. Furthermore, MVP is not visible as free monomers at any given time in cells but instead appears only as vault particles[130]. Indeed, Tanaka *et al*. have resolved the structure of the rat vault at 3.5Å resolution as a particle made of 78 identical MVP[120]. This makes the mammalian vault the biggest ribonucleoprotein organelle in cells, three times the size of a ribosome that can potentially be used as a drug delivery system[163].

Despite its discovery more than 30 years ago, the function of the vault organelle remains elusive. *In vivo*, MVP/vault is expressed in the cytoplasm, neurites and growth cone of rat cortical neurons[116], and associated with cortical plasticity in a monocular deprivation mouse model[147]. *In vitro*, the vault organelle binds to microtubules[138] and MVP interacts with ERK through epidermal growth factor signaling[4,164]. Interestingly, *MAPK3* (mitogen-activated protein kinase 3) encoding the extracellular-signal related kinase 1 or ERK1, maps to the 16p11.2 locus.

Here we set out to assess, in an unbiased and systematic manner, the neuroanatomical implication of each individual gene of the 16p11.2 interval using mouse mutants to inform the genetic architecture of this autism-associated locus. We find that the major vault protein (*Mvp*) gene is one of the top drivers of neuroanatomical phenotypes, regulates neuronal size, and interacts with *Mapk3* at the locus.

4.3 Results

4.3.1 Mouse genetic studies unravel the implications of multiple drivers in NeuroAnatomical Phenotypes at the 16p11.2 locus

To identify which gene(s) regulate(s) mammalian brain architecture at the *de novo* 16p11.2 deletion locus, we aim to assess NeuroAnatomical Phenotypes (named NAPs by us[165]), independently in male and female heterozygous (het) knockout (KO) mice, for each of the 30 protein-coding genes in the interval.

We developed, or acquired through collaboration, 236 adult mutant and 204 colony-matched WT mice, representing 20 unique genes (highlighted in red in **Figure 39A**), each studied with an average of four biological replicates. For the remaining 10 genes, the germline transmission of the mutation failed despite multiple attempts or no mouse model was available during the course of the study**.** For one gene of interest (*Kctd13*), multiple allelic strategies were used. To ensure high comparability between the results, mouse mutants assessed in this study were all processed on identical genetic background (C57BL/6) and at same age using the same pipelines. A detailed description of study samples and allelic constructions is provided in and **Table 3**.

Using a highly robust approach for the assessment of 67 neuroanatomical parameters (**Table 4**), described in details elsewhere[1], we systematically quantified the same two coronal brain sections (Bregma +0.98 mm and Bregma −1.34 mm), and collected neuroanatomical measurements blind to the genotype (**Dataset 1**). These parameters were grouped into five main categories: brain size, commissure, ventricle, cortex and subcortex (**Figure 39B** and **Table 4**). After multiple quality control steps and critical evaluation of each phenotype, gene association was carried out within our internal database using a standardized statistical pipeline (**Figure 42**). Heat maps of the assessed genes comprising percentage change relative to WTs and p-value are provided in **Dataset 2**, **Figure 43** for male and **Figure 44** for female.

We identified 13 genes associated with NAPs (hereafter named as NAP genes), when using a relaxed significance threshold of 0.05, associated with defects in commissure (10 genes), cortex (8 genes), subcortical structures (7 genes), brain size (5 genes) and ventricle (3 genes) in male (**Figure 39C**). Eight genes (*Bola2*, *Qprt*, *Maz*, *Mvp*, *Ppp4c*, *Slx1b*, *Taok2* and *Zg16*) gave significant results affecting two or more categories**.** *Mvp* was the only gene affecting all five main categories. The remaining five genes (*Doc2a*, *Fam57b*, *Hirip3*, *Spn* and *Tbx6*) presented specific phenotypes in one brain category. In the 13 NAP genes, 38.5% decreased the size of the affected brain structures, while 38.5% increased their sizes and 23% had bidirectional effects (**Figure 39C** and **Dataset 2**). When using a stringent significance threshold of 0.0001, *Mvp* and *Tbx6* remained NAP genes.

Figure 39D shows the number of NAPs for each of the 21 individually deleted allele in male. While *Mvp* still stood out as the strongest candidate gene based on the number of affected parameters (n=14), *Ppp4c*, *Zg16* and *Taok2*, showed ten, seven and six NAPs, respectively. These neuroanatomical phenotypes are described in details in **4.8 Supplementary Results** and a heat map provided in **Figure 43**. Our top driver *Mvp* was associated with small brain size (-9%, p=0.016) concomitant with small brain nuclei associated to the limbic system such as the cingulate gyrus (-13%, p=0.0016), the somatosensory cortex (-12%, p=0.0025) and the hippocampus (-20%, p=0.023). Parameters pertaining to brain commissures were also reduced in size, for example the soma of the corpus callosum was reduced by -22% (p=0.044). By contrast, the size of the ventricles was enlarged by 61% (p=0.045) (**Figure 45A**).

Next, sex differences were assessed to determine the impact of each mutation on the female brain. Overall, female mice displayed reduced number of NAP genes (10 as opposed to 13 in male) and for each of the nine genes in common between male and female, there were less affected brain parameters in female (**Figure 39E**). Directionality of the phenotypes was consistent between sex but NAPs were significantly different with an excess of parameters at Bregma -1.34mm for female (p=0.00006, Fisher test) (**Figure 39D-E**). *Gdpd3* was a female-only NAP gene, *Bola2* showed more severe anomalies in female and *Mvp* no anomalies in female (**Figure 45B**). Female NAPs are described in details in **4.8 Supplementary Results** and **Figure 44**.

Four genes (*Coro1a*, *Ino80e*, *Mapk3* and *Kctd13*) were categorized as non-NAP, both in male and female (**Figure 45C**). Considering the discrepancies in the literature[5,161,162], *Kctd13* was assessed twice using two independent allelic constructions (**Table 3**), which gave identical results. We also studied the homozygous (hom) *Kctd13* mice and found a reduction in the size of the hippocampus by 10% (p=0.015) for male (**Figure 43** and **Figure 45D-E**) and 6% (p=0.014) for female (**Figure 44**). This reinforces the existing link between *Kctd13* and hippocampal biology[161,162].

Finally, to assess the genetic interactions between the 30 protein-coding genes of the interval, we studied brain anatomy in two previously published 16p11.2$^{+/Del}$ mice[74], in male and female (**Material and Methods**5. Material and Methods). We were unable to replicate previously reported phenotypes, which ranged from increased volume of the hypothalamus[73] to decreased length of cortices[72], as we found no NAPs (**Figure 39D-E**). Phenotypic variation in body weight was however consistent between our study (**Dataset 2**) and previous reports[72,73,107].

To summarize, our neuroanatomical studies demonstrate the implication of multiple major genes that drive brain size and other NAPs at the 16p11.2 locus with the major vault protein, *Mvp*, gene being one of the top drivers. Our work also indicates profound sex differences with more NAPs in male.

4.3.2 The major vault protein is expressed in the limbic system and is implicated in neuronal morphology

MVP is important for the regulation of brain anatomy as evidenced by the large number of the associated NAPs when deleted in the mouse (**Figure 45A**). However, very little is known about how MVP could be implicated in mammalian brain biology.

To gain a comprehensive understanding of MVP/vault expression in the brain, we established its distribution at multiple scales, independently in male and female WT mice (**Material and Methods**) MVP transcripts, assessed at several developmental stages from embryonic day 16.5 (E16.5) to 30 weeks of age, were constant. Expression was higher in the cerebellum and peripheral tissues than in the cortex and hippocampus (**Figure 40A** and **Figure 46A-D**). The neuroanatomical distribution of the vault organelle was then quantified by immunofluorescence using an anti-vault antibody, throughout the entire brain on consecutive histological sections. MVP/vault signal was refined to the granular layer and the arbor vitae of the cerebellum (**Figure 40B**), the oriens layer of the CA3 region of the ventral hippocampus (**Figure 40C**) and the deep layers of the cingulate gyrus (**Figure 40D**). We noticed MVP/vault signal pertained to specific nuclei of the limbic system, for example the triangular septal nucleus, the zona incerta, the solitary nucleus, the vagal nucleus, the paraventricular hypothalamic nuclei and the dorsal medial thalamic nucleus (**Table 5** and **Figure 46E-N**). At the subcellular level, MVP/vault signal was limited to the cytoplasm of neurons (**Figure 40E** and **Figure 46O-Q**). It was noteworthy that the patterns observed were consistent across sectioning planes (coronal and sagittal) and replicates (**Table 6**) and that no sex differences were detected in the pattern of MVP/vault expression.

The MVP mouse model used in this study was validated as loss-of-function (LoF) of *Mvp*, assessed at the transcript and protein levels (**Figure 47A-D** and **Material and Methods** 5. Material and Methods). We also verified that the expression of neighboring genes was unaffected (**Figure 47E**). Among 617 successfully genotyped animals, we observed Mendelian ratio inheritance, indicating that the loss of *Mvp* has no effect on survival (**Figure 47F**).

To discriminate primary microcephaly from acquired microcephaly (when the brain size is normal at birth but reduced afterwards), we studied brain size in *Mvp*$^{-/-}$ mice across four time points: embryonic day 18.5 (E18.5), postnatal day 10 (P10), P45 and P120,

using an adaptation of the procedure described in the previous section (**Table 7**). The total brain area measurement in male *Mvp*$^{-/-}$ was normal at E18.5 and P10 but smaller at P45 (-7%, *p*=0.0061) and P120 (-9%, *p*=0.046), suggesting that the microcephaly is acquired between P10 and P45 (**Figure 40F**). At P120, in addition to the total brain area being decreased, 16 parameters were significantly smaller including the area of the cingulate gyrus (-8%, *p*=0.007), the thickness of the somatosensory cortex (-12%, *p*=0.0016) and the area of the hippocampus (-24%, *p*=0.038) (**Figure 40G**). In female *Mvp*$^{-/-}$, no change was detected at any given time (**Figure 48A-B**). Number and effect size of NAPs being similar between *Mvp*$^{-/-}$ and *Mvp*$^{+/-}$ male (17 NAPs versus 14, respectively), we performed subsequent studies on *Mvp*$^{-/-}$ only. Given the changes of cortical thickness, we first asked whether alterations in one specific cortical layer may contribute to this phenotype but found none (**Figure 49A-D**). NAPs in male *Mvp*$^{-/-}$ were confirmed on sagittal planes using a previously described procedure[158] showing two new NAPs including small thalamus (-24%, *p*=0.00033) (**Figure 49E** and **Table 8**).

To determine why brain regions were smaller in *Mvp*$^{-/-}$ mice, we took advantage of the high resolution of our approach and developed a suite of automated tools to count the number of cells and calculate the average cell size within each brain region (**Material and Methods**). No significant change was detected in the number of cells in male *Mvp*$^{-/-}$, however cells were significantly smaller in size in all affected brain regions including the cingulate gyrus (-15%, *p*=0.0053) (**Figure 40H**) and somatosensory cortex (-15%, *p*=0.001) (**Figure 40I**). Cell size was normal in unaffected brain region such as the retrosplenial cortex (**Figure 49F-M**). The same set of studies were conducted in female *Mvp*$^{-/-}$ but no defects in cell count nor size were found (**Figure 48C-H**).

To further characterise the cellular phenotype, we conducted hippocampal neuronal cultures, independently in male and female (**Material and Methods**). Consistently, neurons derived from male *Mvp*$^{-/-}$ showed a reduction of the soma size by 6% (*p*=0.0003). The growth cones were smaller by 23% (*p*=0.0081) (**Figure 40J**) and no differences detected for the axonal length (**Figure 40J**). Neurons derived from female *Mvp*$^{-/-}$ showed no morphological defects (**Figure 48I-J**).

Finally, to test if the anatomical changes relate to neural connectivity defects in male *Mvp*$^{-/-}$, we measured miniature excitatory post-synaptic currents (mEPSCs) in pyramidal neurons of the cingulate gyrus (**Material and Methods**). Interestingly, the

mEPSCs amplitude was smaller (p=0.023) (**Figure 50A**), consistent with a reduction in the density of post-synaptic dendritic spines (-5%, p=0.02) (**Figure 40K** and **Dataset 3**). Neuronal ultrastructure was examined but no anomalies were seen in the cingulate gyrus (**Figure 50C-D**). To identify potential biological pathways that might explain NAPs, we carried out transcriptomic analyses in the cingulate gyrus, and although the complete loss of *Mvp* was confirmed, no differentially expressed genes were found (**Figure 50E**).

All together, these findings show that *Mvp* is not essential for survival but has a highly specific pattern of expression in the limbic system and is implicated in the regulation of neuronal size and dendritic spines, postnatally.

4.3.3 The double deletion of Mvp and Mapk3 rescues neuroanatomical phenotypes and alters behavioral performances in mice

MVP-mediated regulation of ERK signaling has been demonstrated in two previous studies[16,17]. To explore the relevance of this interaction, we generated double KOs of *Mvp* and *Mapk3* and tested four genotypes ($Mvp^{+/+}$::$Mapk3^{+/+}$, $Mvp^{+/+}$::$Mapk3^{+/-}$, $Mvp^{+/-}$::$Mapk3^{+/+}$ and $Mvp^{+/-}$::$Mapk3^{+/-}$), both in male and female (**Material and Methods**).

We first examined ERK activity by measuring phospho-ERK in the cortex of the various groups. Consistently with a previous report[84], $Mvp^{+/+}$::$Mapk3^{+/-}$ exhibited an increase in ERK activity (**Figure 41A**), upon ablation of the gene *Mapk3* confirmed by a 50% reduction of ERK1 protein levels (**Figure 51A-B**). Female $Mvp^{+/+}$::$Mapk3^{+/-}$ did not show altered ERK activity (**Figure 41B**). Our quantification of ERK activity revealed a three-fold increase of phospho-ERK in the cortex of $Mvp^{+/-}$::$Mapk3^{+/+}$, indicating that MVP could be an inhibitor of ERK signaling. Accordingly, ERK activity was reduced by more than a third in the double hets $Mvp^{+/-}$::$Mapk3^{+/-}$. These preliminary findings provide evidence of MVP-mediated regulation of ERK signaling, *in vivo*.

Next, we produced new cohorts of double KOs of *Mvp* and *Mapk3* to assess NAPs using our standard procedures for histomorphology[1]. We found no differences in $Mvp^{+/-}$::$Mapk3^{+/-}$ when compared to $Mvp^{+/+}$::$Mapk3^{+/+}$, indicating that the decreased

expression of *Mapk3* rescues the impact of *Mvp*-deficiency on brain size and anatomy (**Figure 41C**).

Furthermore, we ran a range of sixteen behavioral tests evaluating eleven core behaviors (anxiety, depression, anhedonia, memory, locomotion, coordination, motricity, sociability, schizophrenia, autism and epilepsy) in double het and single-gene adult KOs. The behavioral pipeline is described in details in the **Material and Methods** and the raw data available in **Datasets 4-6**. The double hets showed normal body weight (**Figure 51C**) as well as normal motricity, coordination, reward-seeking behaviors, working and short-term memory, associative learning, circadian activity, and no signs of autism and schizophrenic-like behaviors (**Figure 51D-M**). While the open-field test indicated no increase in traveled distance suggesting normal basic activity, it showed increased time spent in the center of the apparatus and decreased latency to enter the center (**Figure 41D**), suggesting a resistance to anxiogenic behavior. The elevated plus maze confirmed the resistance to anxiety, shown by more time spent in the open arm of the device (**Figure 41E**). Fear conditioning in contextual and cued testing showed decreased freezing (**Figure 41F**), indicating improved hazard perception skills. The tail suspension test showed a trend to increased latency to immobility (**Figure 41G**), while the forced swim did not show any apparent differences (**Figure 41H**). Finally, the duration of clonic seizures was dramatically reduced in the pentylenetetrazol (PTZ)-induced seizure paradigm (**Figure 41I**), supporting the idea of a resistance to epilepsy also. Additionally, we studied the same series of eleven core functions and assessed the phenotypes of single-gene mutation of *Mvp* and *Mapk3.* We found no behavioral anomalies in $Mvp^{-/-}$ and $Mapk3^{+/-}$ (**Figure 52** and **Figure 53**, respectively), suggesting that the het deletion of *Mvp* or *Mapk3* is alone not sufficient to potentiate the enhanced behaviors seen in the double het KOs. A summary of behavioral findings is provided in **Figure 41J**.

Taken together, these results confirm the *in vivo* interaction between MVP and ERK1, indicate neuroanatomical correction of *Mvp*-deficiency by reducing expression of *Mapk3*, and reveal the benefit of a double deletion of *Mvp* and *Mapk3* in the neurobiology of fearfulness and epilepsy.

4.4 Discussion

Previous phenotypic screens of 16p11.2 were exclusively conducted in fruit fly[166] and zebrafish[5,167]. Advances in large-scale mouse KO programs[168], combined with our fine-grained analysis of NAPs [1,158], allow for neurogenetic screens in a more suitable model. While many stand-alone mouse functional studies have been conducted in the context of microdeletion syndromes such as 22q11.2 [169], they often yield discrepant findings due to lack of standardized innovative methodology. To overcome this, mouse mutants assessed in this study were all processed on identical genetic background and at same age using the same pipelines, independently in male and female.

Our findings are important in at least three respects. First, we found multiple drivers within the 16p11.2 deletion that underpin brain morphology, revealing an even more complex genetic architecture than the working model of one single primary driver[170]. Among the 20 syntenic genes assessed, two were associated with decreased (-9% for *Mvp* and -10% *Ppp4c*) and three with increased (+7% for *Slx1b*, +6% for *Taok2* and +7% for *Zg16*) brain size. No neuroanatomical changes were detected in the double ablation of *Mvp*::*Mapk3* nor in *16p11.2*$^{+/Del}$ mice. The 16p11.2 locus is thus likely to undergo complex additive and synergistic epistasis between genes within the interval, where drivers can have opposite effect, with the potential to modulate neuroanatomical phenotypes. This could explain the variable MRI findings in patients of different genetic background harboring the 16p11.2 deletion, with one-third[171] to two-thirds[66] not showing brain anatomical changes.

We did not find any brain morphological changes in *Kctd13* het mice but a subtle reduction of the size of the hippocampus in *Kctd13* hom mice, possibly caused by deficit in spine maturation[161] and reduced synaptic transmission [162] in the CA1 area of the hippocampus. At the exception of *Taok2*, linked to enlarged brain size in a recent report[172] and this study, NAP genes identified were new from the literature, both in mouse and human. Among non-NAP genes (genes whose disruption do not yield to neuroanatomical phenotypes), findings were consistent with a report of *Prrt2* KO mice that did not find any overt abnormalities in brain size[173], approving our approach's specificity and sensitivity.

Second, we identify the major vault protein *Mvp* gene as the top driver of NAPs at the 16p11.2 locus. The major vault protein is unviable in cells and only occurs as part of

the vault organelle [116,130]. We thus provide the first evidence that the vault organelle modulates brain and neuronal size in key structures of the limbic system where MVP is specifically expressed including the cingulate gyrus, a structure associated with the management of emotions[174]. We found no differential expression changes in the cingulate gyrus, arguing against a major transcriptional effect of the loss of *Mvp*. MVP has previously been associated with synaptic transmission in other models[147,175]. The vault organelle might also be implicated in the transport of mRNAs along neurites for translation at the synapse[116], although the totality of mRNA species that bind to the vault organelle remains to be identified. Only two vault binding mRNAs translated in response to synaptic activity, are known[116], including the brain-enriched protein tyrosine phosphatase STEP known to inactivate ERK1/2 kinases[176]. We hypothesize that the vault organelle might transport key mRNAs along microtubules for translation at the synapse, to modulate synaptic transmission. Disruptions in neural connections could in turn affect neuronal and brain size. Furthermore, we found a decrease in postsynaptic spine density and mEPSCs amplitude in pyramidal neurons, suggesting that the loss of *Mvp* disrupts neural connections also.

Third, we provide insights into the mechanisms of sex differences with the overall excess of males at the 16p11.2 locus in a suitable model, sexing being not possible in fruit fly[166] and zebrafish[5,167]. By studying males and females independently in each mouse line carrying the same genetic mutation, we found that females show less NAPs than males (26 as opposed to 67). This suggests that, from an anatomical perspective, the female mouse brain is more robust while the male brain is more vulnerable when a genetic mutation is involved. This validates the “female protective effect” previously predicted in human neurodevelopmental disorders[177], and to some extent excludes the possibilities of female underdiagnosis[178]. Sex-specific neuroanatomical endophenotypes were previously studied in 16p11.2$^{+/-}$ mice where the expression patterns of the genes within the 16p11.2 interval overlapped with brain structural changes[179]. However, our findings on MVP exclude expression differences between male and female as a potential cause responsible for the underlying sex differences in NAPs, favoring the hormonal hypothesis[178].

ERK1/2 kinase signaling cascades are central in the control of neuronal plasticity and size[83], and can be activated in response to testosterone[180]. While hyperphosphorylation of ERK1 has been previously reported in several studies of

16p11.2$^{+/Del}$ mice, sex differences in ERK phosphorylation were overlooked[84,107], at the exception of one recent study[181]. On this ground, pharmacological ERK inhibitors have been used to improve neurological traits in 16p11.2$^{+/Del}$ mice[107]. In the current study, ablation of *Mapk3* showed increased ERK1 activation in male cortex not female, suggesting a potential hormone-dependent regulation of kinase cascades. Our double heterozygous knockout strategy of *Mvp* and *Mapk3* supports the idea that the vault organelle could act as a naturally-occurring inhibitor of ERK1 signaling pathway, possibly through its association with STEP [176]. On this ground, ERK inhibitors have been used to improve neurological traits in 16p11.2$^{+/-}$ mice[107].

Finally, human and mouse MVP sharing approximately 90% of their amino acid residues, the vault organelle likely plays a similar role in the human brain. Human MVP alone has however not yet been implicated in abnormal brain size which could be due to small effect size missed by conventional human brain imaging technologies or to a polygenic model of inheritance involving other genes of the 16p11.2 interval such as *Mapk3*. Interestingly, a smaller region encompassing five out of the 30 genes have been narrowed down[182] and included *MVP*. In summary, we have begun to identify what may be an underlying reason why neurodevelopmental disorders such as autism predominantly affect boys with the identification of the vault organelle and its implication in ERK1 signaling pathway.

Acknowledgements

We are very grateful to Leonard H Rome, Nancy Lynn Kedersha and Valerie Kickhoefer for exciting discussions on the vault organelle and kindly providing anti-MVP IgG antibodies (George and Charlie) and anti-vault antibody (N2-B15). We thank Ipek Yalcin for her guidance in isolating the anterior cingulate gyrus. For histological work, we thank Sylvie Nguyen as well as Dorinda Wright and Lindsey Proctor at HistologiX Ltd. We thank the students involved in the study (Jules Roussey, Emeline Aguilar, Paula Hahn, Durna Kumruoglu, David Gualberto, Jonathan Delevoye, Sarah Arthur, Rebecca Balz and Léo Gagliardi). We thank Renee Araiza, Taylor J Lunceford and Kevin C. Kent Lloyd for the shipping of the Hirip3 mouse. We thank Gilles Pagès (University of Nice), Virginia E. Papaioannou (University of Columbia), Yoshimi Takai (Kobe University), and Nicholas Katsanis and Christelle Golzio (Duke University Medical Center) for kindly providing the *Mapk3*, *Tbx6*, *Doc2a* and *Kcdt13* mice, respectively. We thank Jean Pieters (University of Basel) and Tsutomu Fukuwatari (Kyoto University) for kindly providing brain samples for Coro1a and Qprt, respectively. We thank members of the International Mouse Phenotyping Consortium (IMPC) including the Italian National Research Council (Roma), the Mouse Clinical Institute (Illkirch), the Mouse Biology Program (MBP) at UC Davis (California), the Texas A&M Institute for Genomic Medicine (TIGM) and the Wellcome Sanger Institute (Hinxton). We thank members of core facilities at the Institute of Genetics and Molecular and Cellular Biology (Illkirch) and the Mouse Clinical Institute (Illkirch) involved in the study including the team of Sylvie Jacquot for genotyping, Nadia Messadeq for electron microscopy, Dalila Ali-Hadji for animal care, Hamid Meziane for behavioural testing, and Christelle Thibault-Carpentier for transcriptomics. Sequencing was performed by the GenomEast platform, a member of the « France Génomique » consortium (ANR-10-INBS-0009). The study was funded by the French National Research Agency (ANR-11-PDOC-0029 and ANR-18-CE12-0009), the Gutenberg Circle, the grant ANR-10-LABX-0030-INRT, a French State fund managed by the Agence Nationale de la Recherche under the frame program Investissements d'Avenir ANR-10-IDEX-0002-02, and the Infrafrontier Research Infrastructure (TAOK2), to B.Y.

Author contributions

B.Y. directed the study and acquired funding for the project. P.F.K., C.W., C.M., S.H., M.M., M.C.F., M.K., and S.C.C. performed/analyzed data for mouse neuroanatomy, neuronal cultures, behavior and electrophysiology; I.M. performed/analyzed phospho-ERK in brain sections with supervision by R.B.; Y.H., A.R. and M.S. provided resources. S.C.C managed all aspects of the project. B.Y. wrote the paper together with P.F.K. All authors edited and approved the manuscript.

4.5 Figures and legends

Figure 39 Mouse neuroanatomical screen at the 16p11.2 autism-associated locus.

(**A**) Schematic representation of the human 16p11.2 region showing gene content and order with genes that underwent mouse neuroanatomical studies in red. (**B**) 67 brain parameters (listed in **Table 4**) are grouped into five categories (brain size, ventricle, cortex, commissure and subcortex) on two coronal sections at Bregma +0.98mm and Bregma -1.34mm. (**C**) Venn diagram illustrating NeuroAnatomical Phenotype (NAP) genes (genes whose mutations yield neuroanatomical phenotypes) in male mice positioned on each category. Phenotypic directionality is color-coded. Green font corresponds to reduction in structure size, red to increase and yellow to both. Bar plots showing the number of NAPs per section (Bregma +0.98mm in orange and Bregma -1.34mm in blue) for each gene assessed for male (**D**) and female (**E**), respectively. Genes are listed on the x-axis and sorted according to the number of NAPs.

Figure 39

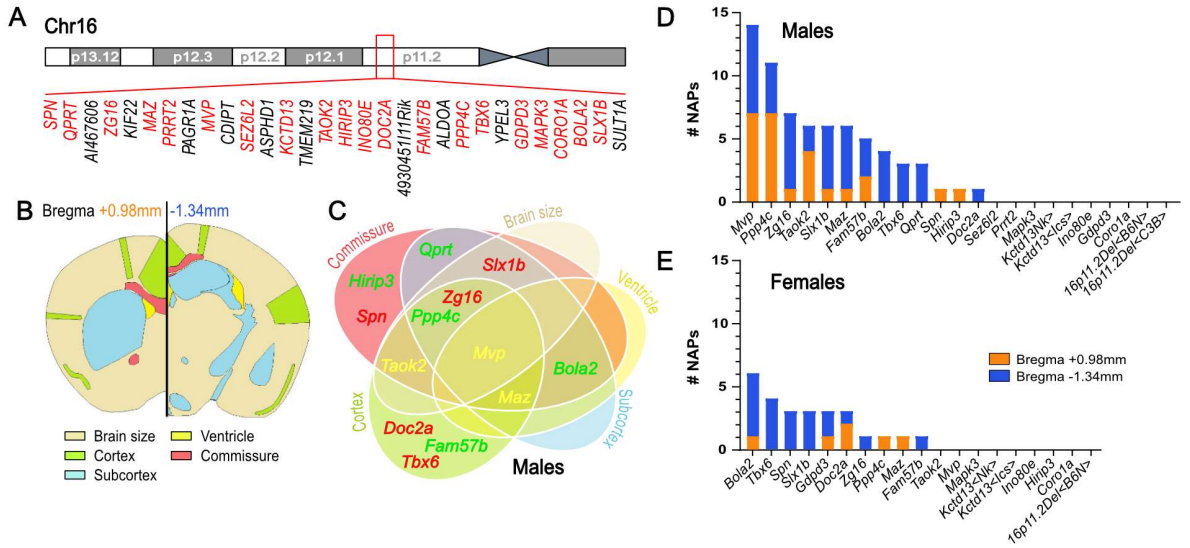


Figure 40 Involvement of the *Mvp* gene in male brain morphology.

(**A**) *Mvp* qRT-PCR expression in liver (LIV), cerebellum (CRB), hippocampus (HP) and cortex (CTX). Normalization was done using *Gnas*/*Hprt* ratio. Immunofluorescence of MVP (green) in CRB (**B**), ventral HP (**C**) and deep layers of cingulate gyrus (Cg) (**D**). Arrow heads point to MVP-positive cells. AV, arbor vitae; gr/mole, granular/molecular layer of CRB; or/py, oriens/pyramidal layer of HP. (**E**) Image of one single MVP-positive cell (top left), co-labeled with NeuN (red) and Hoechst (blue). For detailed description of MVP expression studies, see **Table 5, Table 6** and **Figure 46**. (**F**) *Left.* Graph showing the total brain area (TBA) measured at four time points, embryonic age 18.5 (E18.5), postnatal day 10 (P10), P45 and P120. *Right.* Half images of coronal brain sections stained with Nissl-luxol at the four time points from $Mvp^{+/+}$ and $Mvp^{-/-}$. (**G**) *Left*. Schematic representation of 24 brain regions assessed at Bregma +0.98mm and -1.34mm in male $Mvp^{-/-}$ mice at P120. Colored regions indicate the presence of at least one significant parameter within the brain region at the 0.05 level. White indicates a p-value >0.05, grey shows not enough data to calculate a p-value. *Right*. Histograms of percentage change relative to $Mvp^{+/+}$ for each of the 42 parameters (see **Table 4**). (**H-I**). *Left*. Coronal sections of Cg (**H**) and somatosensory cortex S2 (**I**) from $Mvp^{+/+}$ and $Mvp^{-/-}$ at P120. *Right.* Cell count and average cell area measures within Cg and S2. (**J**) *Left.* Graphs showing measures of soma area, growth cone area and axonal length from hippocampal neurons from $Mvp^{+/+}$ and $Mvp^{-/-}$. *Right.* Neurons stained with MAP2 (red) and SMI312 (green). (**K**) *Left*. Spine density from dendrites of pyramidal neurons in the cortex from $Mvp^{+/+}$ and $Mvp^{-/-}$. *Right*. Representative Golgi images. Plots are shown as mean ± SEM (except **G**). One-way ANOVA with Tukey *post-hoc* (**A**), two-tailed Student *t* test equal variance (**F-K**). *p<0.05 **p<0.01 ***p<0.001.

Figure 40

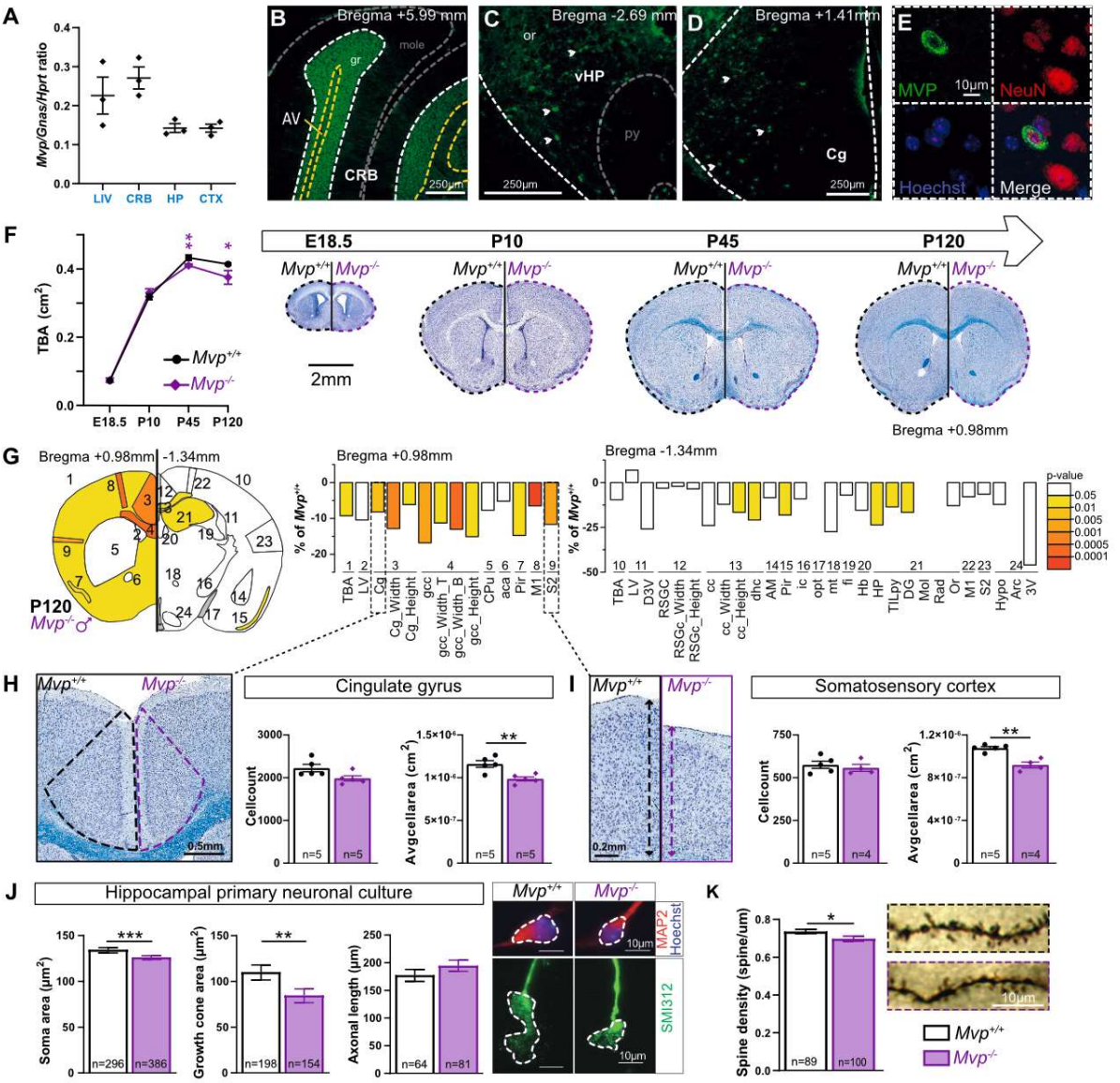

A
Mvp/Gnas/Hprt ratio
LIV CRB HP CTX
B Bregma +5.99 mm
gr
mole
AV
CRB
250µm
C Bregma -2.69 mm
or
vHP
py
250µm
D Bregma +1.41mm
Cg
250µm
E
MVP
10µm
NeuN
Hoechst
Merge
F
TBA (cm²)
Mvp+/+
Mvp-/-
E18.5 P10 P45 P120
E18.5
P10
P45
P120
2mm
Bregma +0.98mm
G
Bregma +0.98mm -1.34mm
P120
Mvp-/- ♂
Bregma +0.98mm
% of Mvp+/+
TBA LV Cg Cg_Width Cg_Height gcc gcc_Width_T gcc_Width_B gcc_Height CPu aca Pir M1 S2
Bregma -1.34mm
TBA LV D3V RSGC RSGc_Width RSGc_Height cc cc_Width cc_Height dhc AM Pir ic opt mt fi Hb HP TlLpy DG Mol Rad Or M1 S2 Hypo Arc 3V
p-value
0.05
0.01
0.005
0.001
0.0005
0.0001
H
Mvp+/+
Mvp-/-
0.5mm
Cingulate gyrus
Cellcount
Avgcellarea (cm²)
n=5
n=5
I
Mvp+/+
Mvp-/-
0.2mm
Somatosensory cortex
Cellcount
Avgcellarea (cm²)
n=5
n=4
J
Hippocampal primary neuronal culture
Soma area (µm²)
n=296
n=386
Growth cone area (µm²)
n=198
n=154
Axonal length (µm)
n=64
n=81
Mvp+/+
Mvp-/-
MAP2
Hoechst
SMI312
10µm
K
Spine density (spine/um)
n=89
n=100
10µm
Mvp+/+
Mvp-/-

Figure 41 Analysis of double deletion of *Mvp* and *Mapk3* genes.

(**A-B**) Number of phospho-ERK positive neurons quantified through the prefrontal cortex by immunofluorescence in male (**A**) and female (**B**). (**C**) Selection of three histological parameters in *Mvp*$^{+/-}$*::Mapk3*$^{+/-}$ mice: total brain area (TBA), cingulate gyrus (Cg) area and average Cg cell area (see **Dataset 1** and **Dataset 3** for neuroanatomical and cellular raw data, respectively). Selection of behavioral paradigms among 17 tested in 12 *Mvp*$^{+/+}$*::Mapk3*$^{+/+}$ and 12 *Mvp*$^{+/-}$*::Mapk3*$^{+/-}$ mice, from 11 to 25 weeks of age: open field (**D**), elevated plus maze (**E**), fear conditioning (**F**), tail suspension (**G**), forced swim (**H**) and PTZ-test (**I**). (**J**) Summary table of core functions assessed in various cohorts. Green indicates significance, yellow trend, grey no difference and white no data. Arrows indicate directionality of effect. A comprehensive description of behavioral tests, raw datasets and additional findings are provided in **Material and Methods**, **Datasets 4-6** and **Figure 51-53**, respectively. Mean and standard error of the mean are shown in the graphs and sample size indicated within each graph. Two-tailed Student *t* test equal variance (**C**, **E**, **G-I**), one-way ANOVA with Tukey post hoc (**A-B**) and two-way ANOVA with Sidak post hoc (**D**, **F**). **p*<0.05 ***p*<0.01 ****p*<0.001.

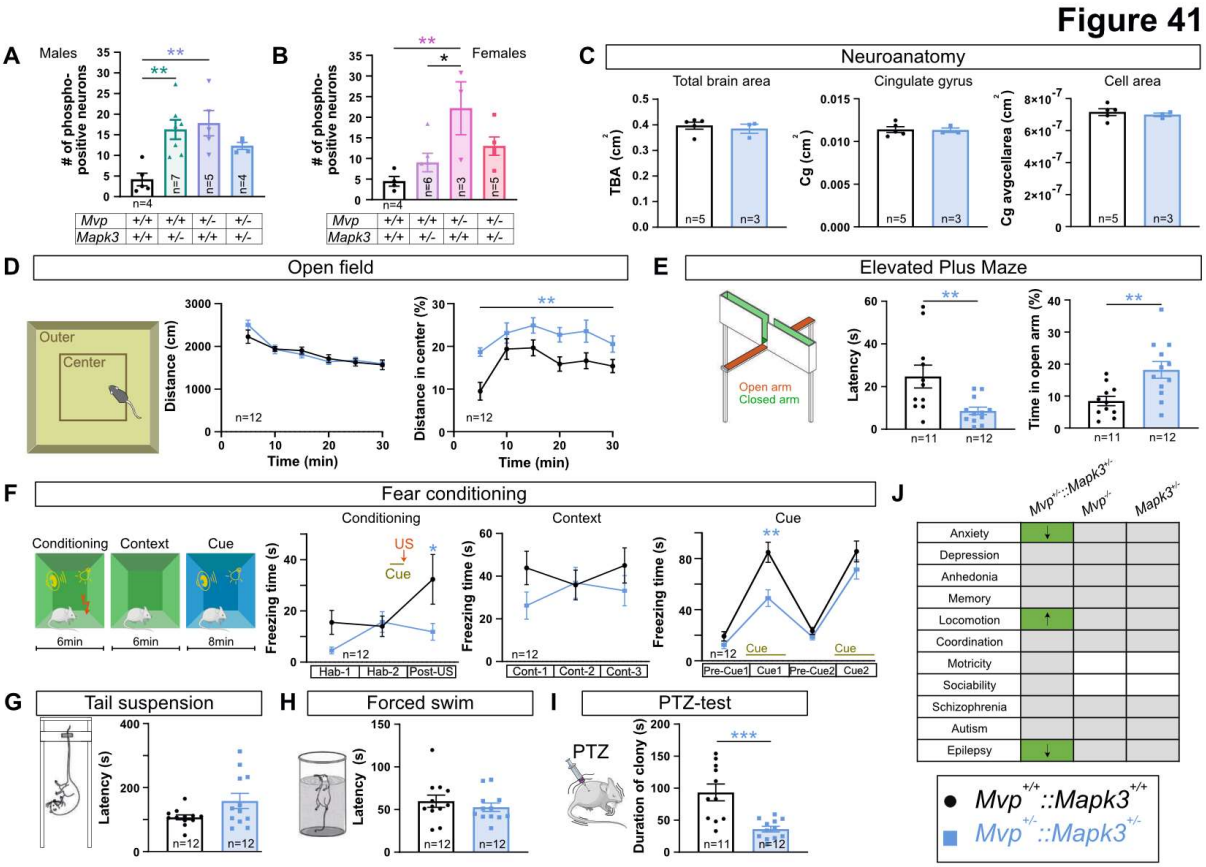


Figure 42 Standardized workflow for neurophenotyping.

The workflow contains six items corresponding to the mouse itself, the brain sample, the paraffin-embedded blocs, the brain images, the quantitative measures and the heat map. Each of these items are interconnected by specific standard operating procedures. i) Mouse: Mutant mice were either produced by the partners of the International Mouse Phenotyping Consortium (IMPC) or acquired through collaboration (**Material and Methods** and **Table 3)**. Mouse colonies were maintained on a pure inbred C57BL/6 background and brain tissues were carefully harvested at 16 weeks of age. ii) Brain: Samples were fixed in 10% buffered formalin for 48 hours and trimmed for paraffin embedding. iii) Blocs: Paraffin blocs were sectioned with a sliding microtome (Microm HM 450) at a thickness of 5µm. Sections were double-stained using Nissl (violet) for neurons and Luxol (blue) for myelin, generating a color contrast between white and grey matters. iv) Images: Whole slides were scanned using a digital slide Hamamatsu Nanozoomer scanner (2.0HT C9600 series), producing whole brain images at cell-level resolution. Each image was checked for stringent quality parameters including symmetry and position relative to the defined sections. v) Measures: Neurophenotyping parameters were measured using Fiji 1.51e across the two stereotactically defined coronal sections (**Table 4**). At Bregma +0.98 mm and Bregma -1.34mm, 22 and 45 parameters were measured, respectively. Measurements and quality control were made within a relational database interface in FileMaker Pro 14.0.6. To test for any potential co-variate structure effect, experimental co-variates were also recorded, for example sample processing dates and usernames were collected at every step of the procedure. Extensive quality control checks and critical evaluation of the dataset were carried out. Gene association was considered as significant when the p-value was smaller than 0.05. All significant gene associations were critically verified by an independent experimenter with extensive expertise in brain anatomy. vi) Heat map: A color-coded heat map of neuroanatomical defects including percentage change and corresponding p-value was generated once the above steps were completed. NAP genes were classified based on the number of brain categories affected as one or multiple.

Figure 42

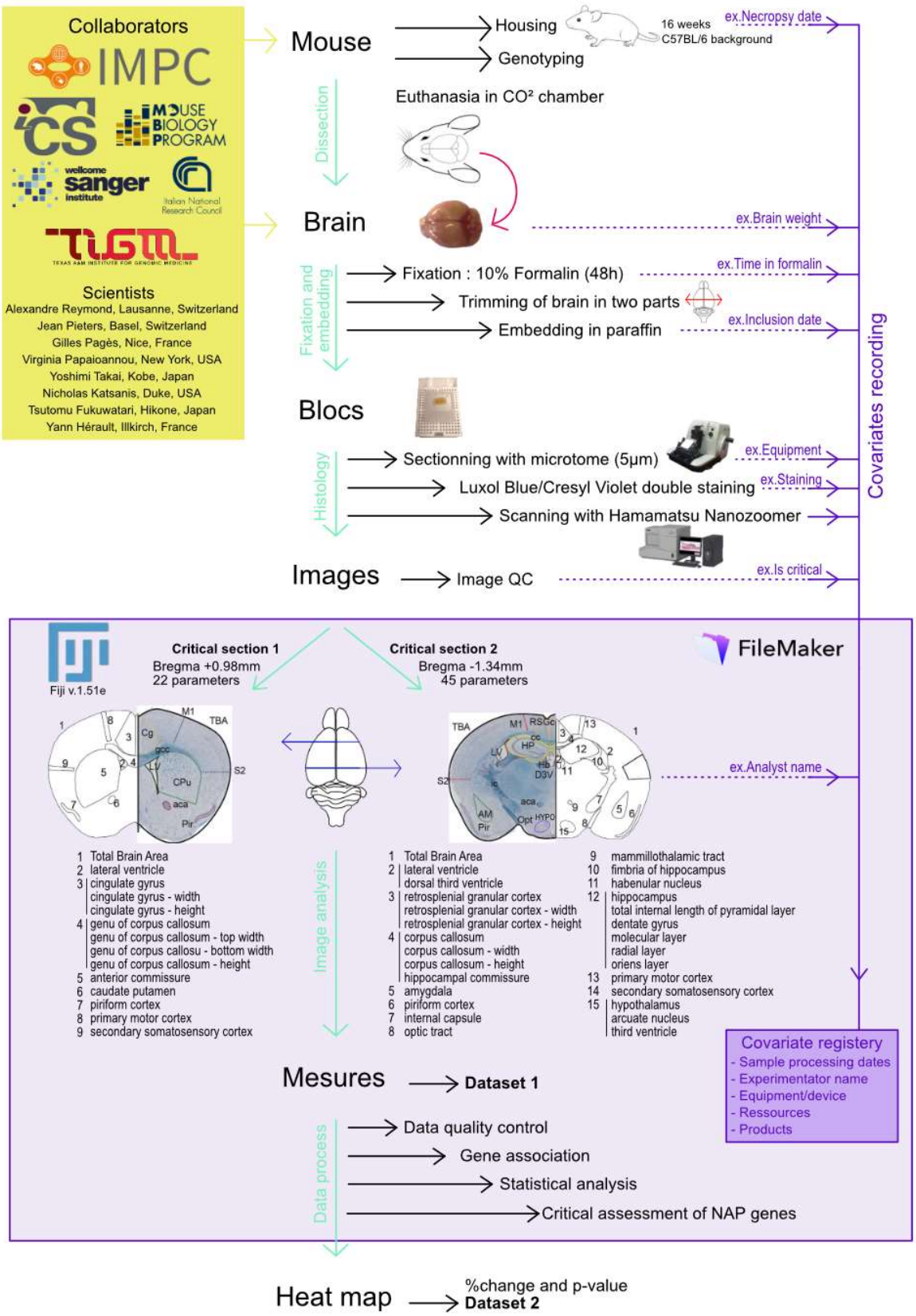

Collaborators
IMPC
ICS
MOUSE BIOLOGY PROGRAM
wellcome sanger institute
Italian National Research Council
TIGM
Scientists
Alexandre Reymond, Lausanne, Switzerland
Jean Pieters, Basel, Switzerland
Gilles Pagès, Nice, France
Virginia Papaioannou, New York, USA
Yoshimi Takai, Kobe, Japan
Nicholas Katsanis, Duke, USA
Tsutomu Fukuwatari, Hikone, Japan
Yann Hérault, Illkirch, France
Mouse
Housing
16 weeks
C57BL/6 background
ex.Necropsy date
Genotyping
Dissection
Euthanasia in CO² chamber
Brain
ex.Brain weight
Fixation and embedding
Fixation : 10% Formalin (48h)
ex.Time in formalin
Trimming of brain in two parts
Embedding in paraffin
ex.Inclusion date
Blocs
Histology
Sectionning with microtome (5µm)
ex.Equipment
Luxol Blue/Cresyl Violet double staining
ex.Staining
Scanning with Hamamatsu Nanozoomer
Images
Image QC
ex.Is critical
Covariates recording
Fiji v.1.51e
FileMaker
Critical section 1
Bregma +0.98mm
22 parameters
Critical section 2
Bregma -1.34mm
45 parameters
ex.Analyst name
1 Total Brain Area
2 lateral ventricle
3 cingulate gyrus
cingulate gyrus - width
cingulate gyrus - height
4 genu of corpus callosum
genu of corpus callosum - top width
genu of corpus callosu - bottom width
genu of corpus callosum - height
5 anterior commissure
6 caudate putamen
7 piriform cortex
8 primary motor cortex
9 secondary somatosensory cortex
Image analysis
1 Total Brain Area
2 lateral ventricle
dorsal third ventricle
3 retrosplenial granular cortex
retrosplenial granular cortex - width
retrosplenial granular cortex - height
4 corpus callosum
corpus callosum - width
corpus callosum - height
hippocampal commissure
5 amygdala
6 piriform cortex
7 internal capsule
8 optic tract
9 mammillothalamic tract
10 fimbria of hippocampus
11 habenular nucleus
12 hippocampus
total internal length of pyramidal layer
dentate gyrus
molecular layer
radial layer
oriens layer
13 primary motor cortex
14 secondary somatosensory cortex
15 hypothalamus
arcuate nucleus
third ventricle
Covariate registery
- Sample processing dates
- Experimentator name
- Equipment/device
- Ressources
- Products
Mesures
Dataset 1
Data process
Data quality control
Gene association
Statistical analysis
Critical assessment of NAP genes
Heat map
%change and p-value
Dataset 2

Figure 43 Heat map of gene association with NAPs in males.

Brain parameters are shown along the horizontal axis and studied alleles along the vertical axis. Color intensity indicates the strength of the association based on the p-value significance threshold. Blue color refers to a brain region that is increased in size while red shows a region decreased in size. More details are available in **Dataset 2**.

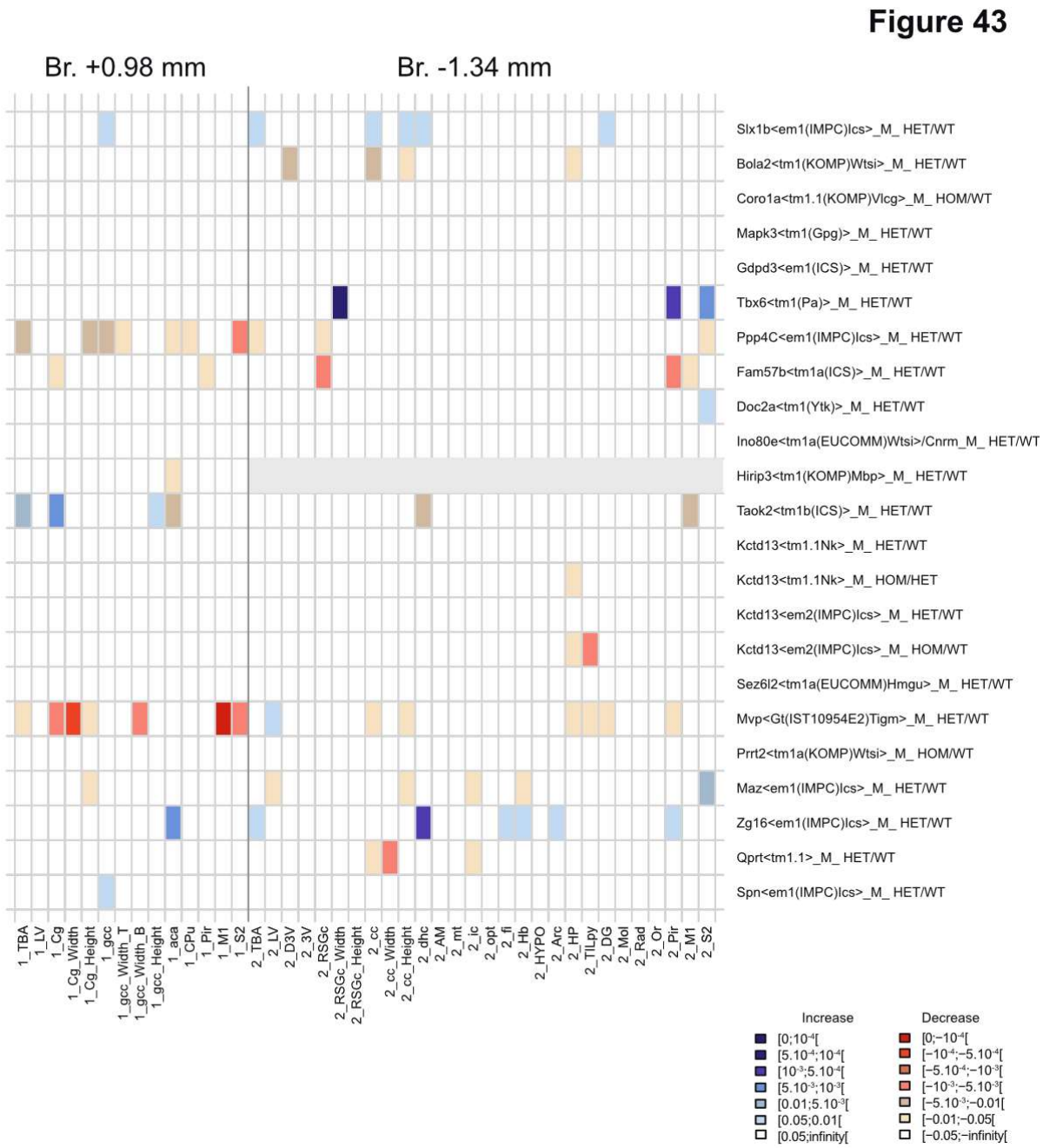


Figure 44 Heat map of gene association with NAPs in females.

Brain parameters are shown along the horizontal axis and studied alleles along the vertical axis. Color intensity indicates the strength of the association based on the p-value significance threshold. Blue color refers to a brain region that is increased in size while red shows a region decreased in size. More details are available in **Dataset 2**.

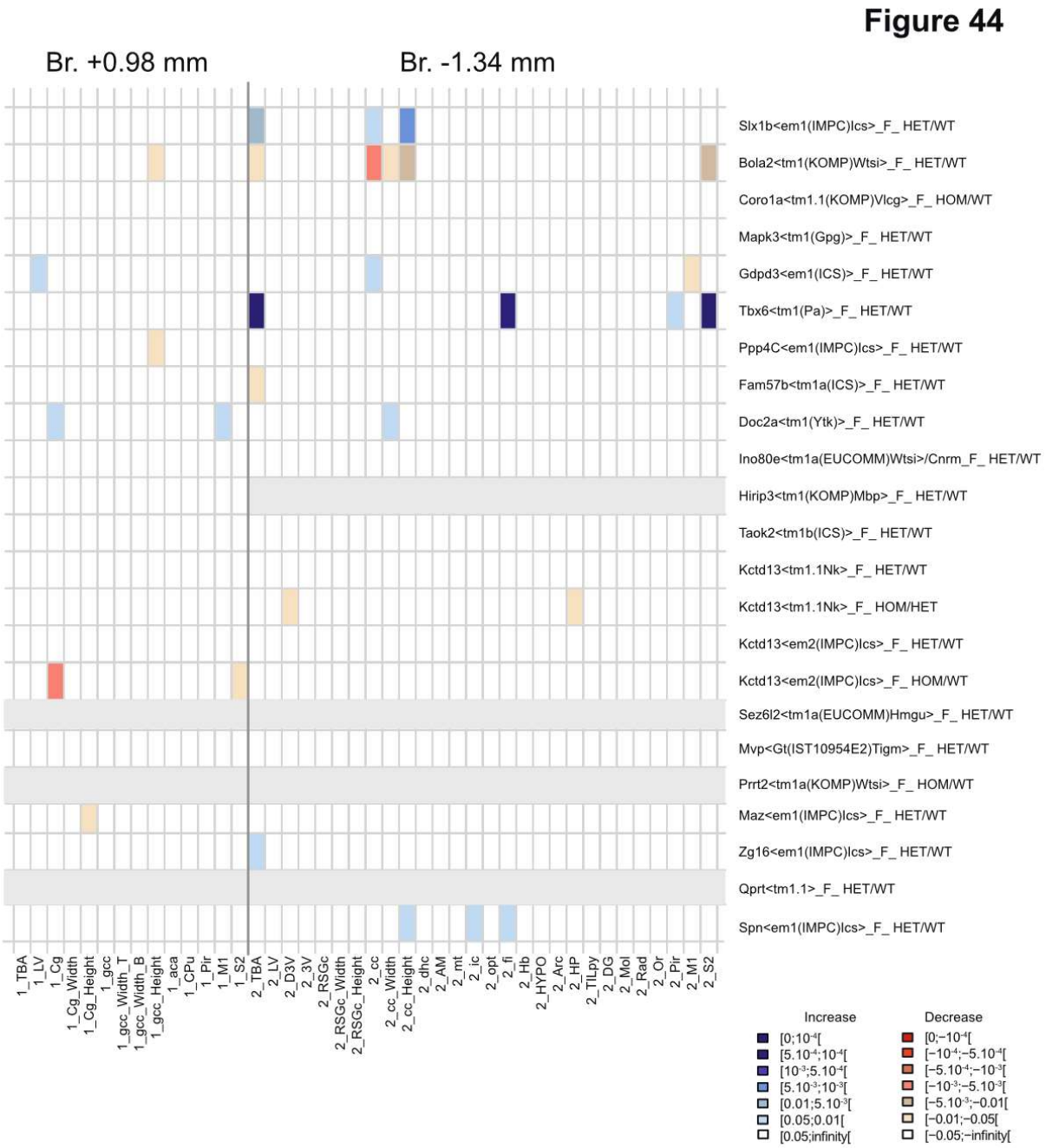


Figure 45 Comparative analysis of *Mvp*, *Mapk3* and *Kctd13* mice.

Each panel represents results obtained from quantitative histomorphological analysis of heterozygous male (**A**) and female (**B**) *Mvp*$^{+/-}$, heterozygous male *Mapk3*$^{+/-}$ (**C**) and heterozygous (**D**) and homozygous (**E**) male *Kctd13*. Two different genetic models of *Kctd13* were used (<Nk> and <Ics> explained in **Table 3**). *Left*. Schematic representation of the 24 brain regions quantified at Bregma +0.98mm and -1.34mm on coronal sections. Colored regions indicate the presence of at least one significant parameter within the brain region at the 0.05 level. White indicates a p-value > 0.05, grey shows not enough data to calculate a p-value. *Right.* Histograms of percentage change relative to matched WT (set as 0) for each of the measured parameters (listed in **Table 4**). For complete data sets (% change and p-value) of male and female neuroanatomical studies, see **Dataset 2**.

Figure 45

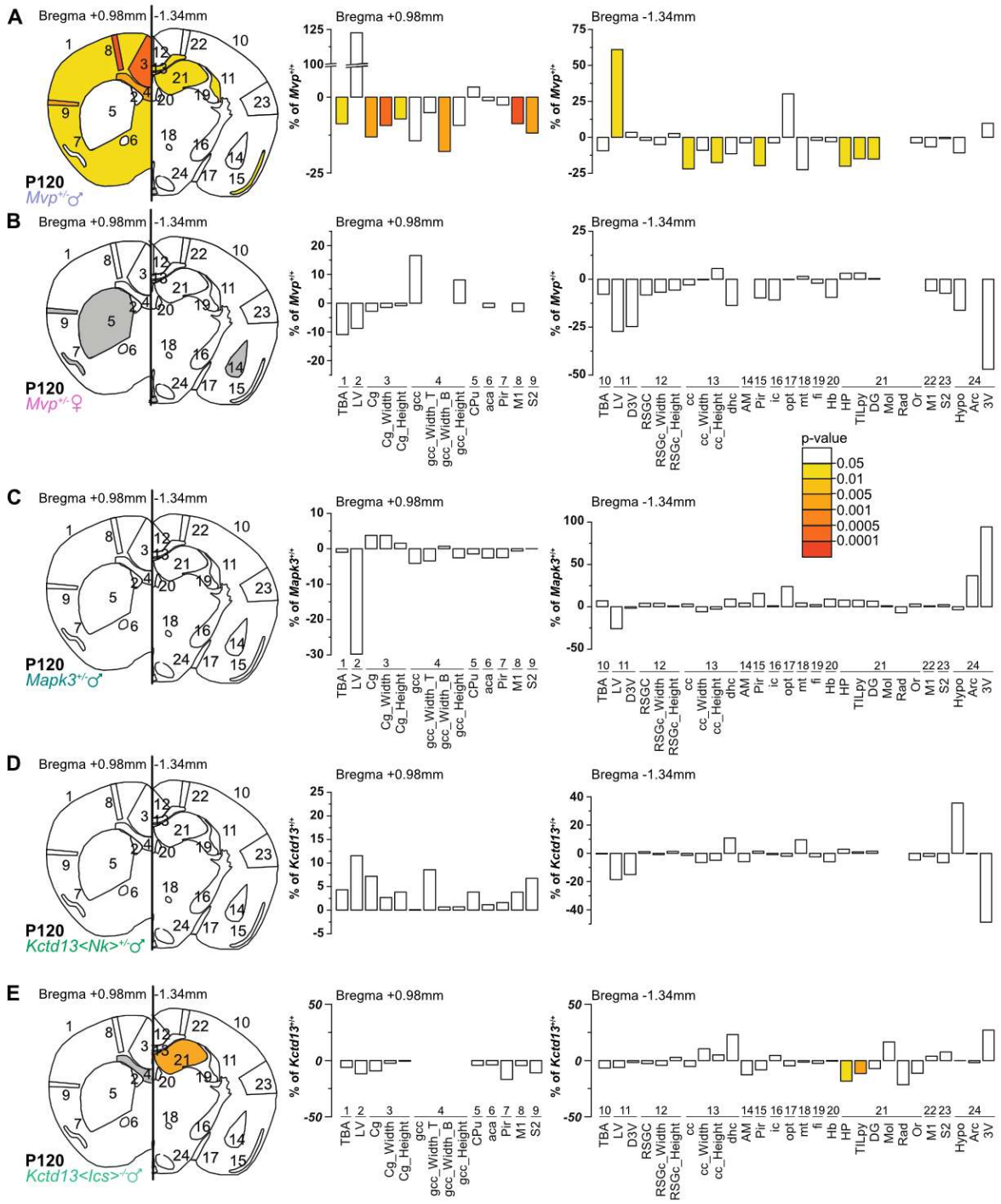

A
Bregma +0.98mm
-1.34mm
P120
Mvp+/-♂
% of Mvp+/+
B
P120
Mvp+/-♀
C
P120
Mapk3+/-♂
% of Mapk3+/+
D
P120
Kctd13<Nk>+/-♂
% of Kctd13+/+
E
P120
Kctd13<Ics>+/-♂
TBA LV D3V RSGC RSGc_Width RSGc_Height cc cc_Width cc_Height dhc AM Pir ic opt mt fi Hb HP TiLpy DG Mol Rad Or M1 S2 Hypo Arc 3V
Cg Cg_Width Cg_Height gcc gcc_Width_T gcc_Width_B gcc_Height CPu aca Pir M1 S2
p-value
0.05
0.01
0.005
0.001
0.0005
0.0001

Figure 46 *Mvp* expression profiling.

(**A**) Reverse transcriptase PCR assessing *Mvp* expression in whole murine brain extracts from WT mice at different developmental stages, ranging from embryonic day 16.5 (E16.5), to E18.5, post-natal day 0 (P0), P10, P30 and up to adulthood (P210). Blue and pink bars indicate male and female samples, respectively. (**B-C**) Reverse transcriptase PCR assessing *Mvp* expression in male (**B**) and female (**C**) extracts. Four brain regions were tested including the cortex (CTX), the hippocampus (HP), the striatum (STR) and the cerebellum (CRB), as well as four peripheral organs: the liver (LIV), the lung (LUN), the heart (HRT) and the kidney (KID). Three biological replicates were used. *Gapdh* was used as loading control. (**D**) Quantitative reverse transcriptase PCR was conducted to evaluate *Mvp* expression in female WT mice from extracts of liver, cerebellum, hippocampus and cortex. *Mvp* transcript levels were normalized over *Gnas* and *Hprt* ratio. One-way ANOVA with Tukey *post-hoc*. $*p<0.05$ $**p<0.01$. (**E-N**) Selection of MVP immuno-histo-fluorescence (IHF) images on coronal brain sections, each of 30 µm in thickness, performed throughout the entire brain from Bregma +1.41mm to Bregma -7.67mm in three biological WT replicates. MVP presence was revealed using an anti-vault antibody (N2-B15) and secondary antibodies fused to Alexa488 fluorochrome (green). Dashed lines delineate regions with MVP-positive cells (white), MVP-positive tract (yellow) and MVP-negative cells (grey) (also see **Figure 40A**). It is noteworthy to mention that the experiment was repeated on a new set of brain samples on sagittal orientation, ranging from Lateral +0.24mm to Lateral +0.84mm. Quantitative analysis both on coronal and sagittal sections was done and MVP expression was scored as mild (+), moderate (++) or strong (+++) across all the positive brain regions. The complete analysis is available in **Table 6**. Pir, piriform cortex; MS, medial septum; DBB, diagonal band of Broca; TS, triangular septal nucleus; ZI, zona incerta; RM, retromammillary nucleus; py/mol, pyramidal/molecular layer of the hippocampus (HP); DG, dentate gyrus of the hippocampus; 4V, fourth ventricle; CP, choroid plexus; MVe, medial vestibular nucleus; Sol, solitary nucleus; AP, area postrema; 10N, vagus nucleus; Raphe, raphe magnus nucleus. (**O-Q**) IHF of MVP (green), NeuN neurons (red) and Hoechst (blue) on brain sections from WT male (**O**, **P** and **Q**) and female (**O'**, **P'** and **Q'**) mice. Pictures from somatosensory cortex (**O** and **O'**), piriform cortex (**P** and **P'**) and cerebellum (**Q** and **Q'**) shown. Dashed square delineates the region presented at higher-magnification on individual channels for MVP, NeuN and Hoechst, and merge of the three. In all structures, MVP signal came from the cytoplasm with no overlap with the neuronal nuclei, NeuN, marker.

Figure 46

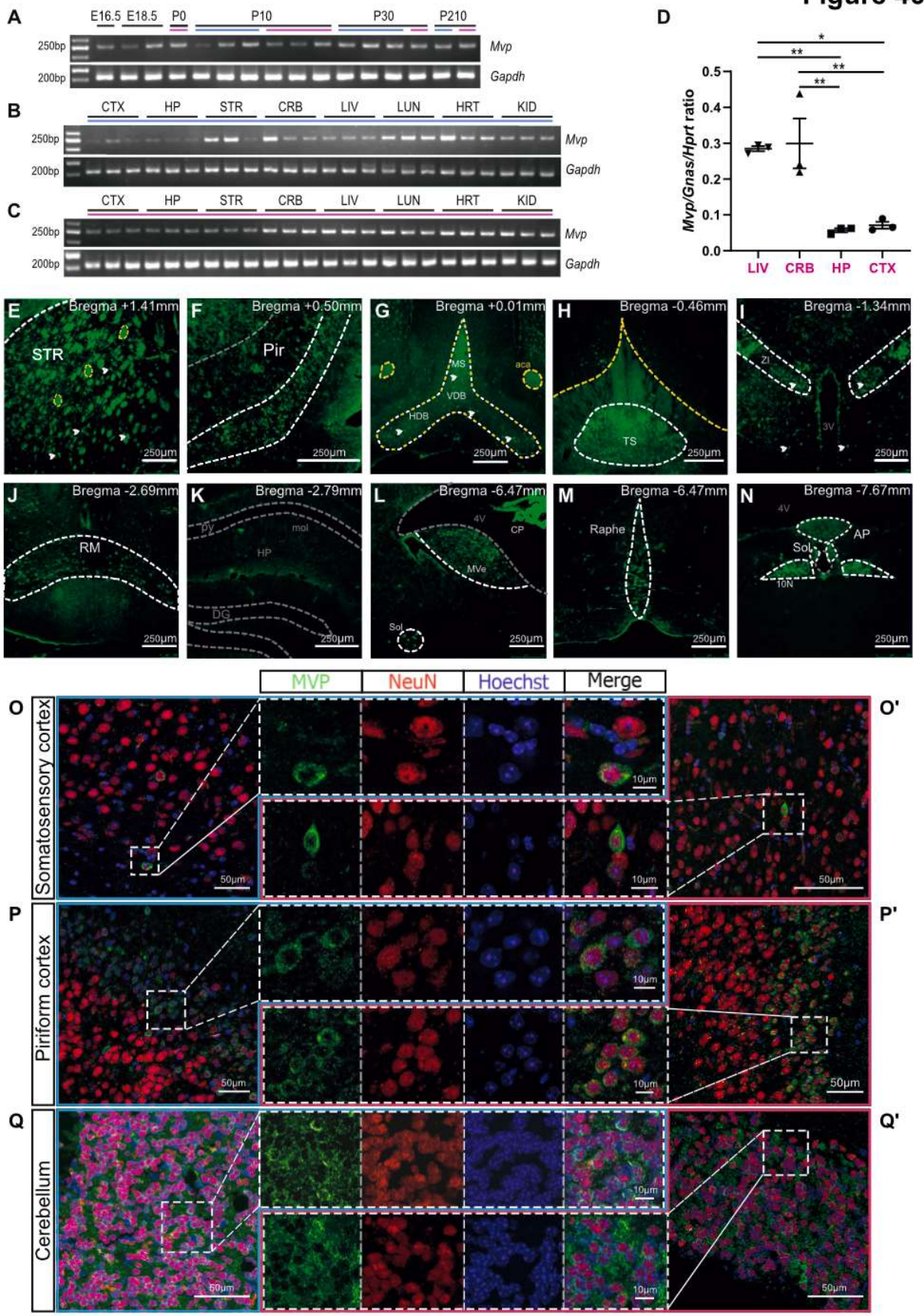
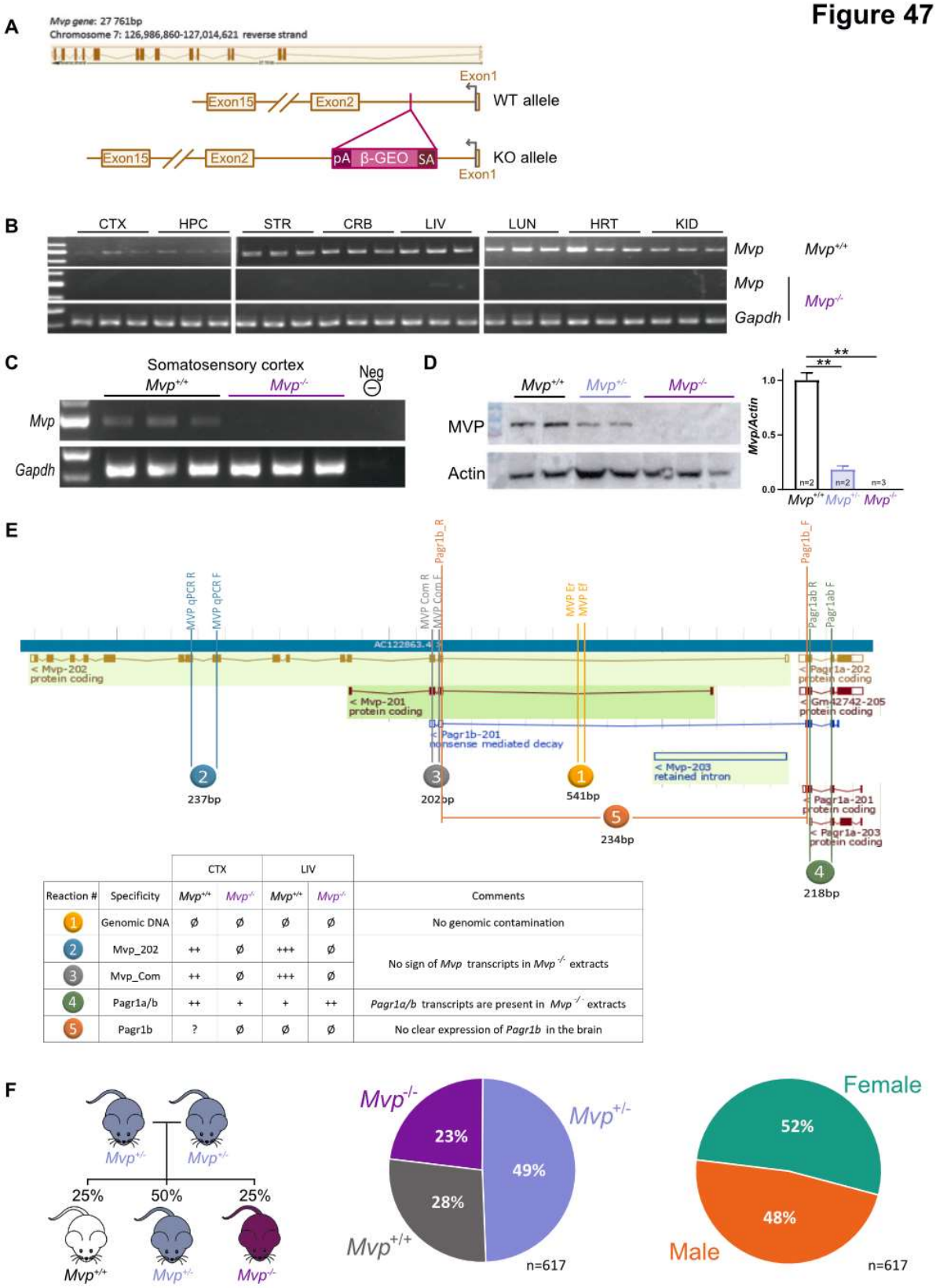

A
E16.5 E18.5 P0 P10 P30 P210
250bp
200bp
Mvp
Gapdh
B
CTX HP STR CRB LIV LUN HRT KID
C
D
Mvp/Gnas/Hprt ratio
LIV CRB HP CTX
E Bregma +1.41mm
STR
F Bregma +0.50mm
Pir
G Bregma +0.01mm
MS
aca
VDB
HDB
H Bregma -0.46mm
TS
I Bregma -1.34mm
ZI
3V
J Bregma -2.69mm
RM
K Bregma -2.79mm
mol
HP
DG
L Bregma -6.47mm
4V
CP
MVe
Sol
M Bregma -6.47mm
Raphe
N Bregma -7.67mm
4V
AP
Sol
10N
250µm
MVP
NeuN
Hoechst
Merge
O
O'
Somatosensory cortex
P
P'
Piriform cortex
Q
Q'
Cerebellum
10µm
50µm

Figure 47 *Mvp* knockout model validation and viability screen.

(**A**) Schematic representation of *Mvp* knockout model construction. Murine *Mvp* gene, located on the reverse strand on chromosome 7 from position 126,986,860 to 127,014,621 (spanning 27,761bp) on Build GRCm38.p6 of the mouse genome, contains 15 exons. The knockout mutants were obtained by a promoter gene trap design, resulting in the insertion of a β-galactosidase/neomycin phosphotransferase (β-GEO) cassette within the first intron of *Mvp* gene. The β-GEO cassette comprises a splice acceptor (SA) allowing the expression of the reporter as well as drug resistance. The polyA-tail (pA) serves to end the transcription, and thus discard the transcription of *Mvp* gene. (**B**) Validation of *Mvp* knockout at the transcript level. RT-PCR for *Mvp* transcript from $Mvp^{+/+}$ (top panel) and $Mvp^{-/-}$ (middle and bottom panels). Four brain regions (cortex (CTX), hippocampus (HP), striatum (STR) and cerebellum (CRB)) and four peripheral organs (liver (LIV), lung (LUN), heart (HRT) and kidney (KID)) were tested on three biological replicates. *Gapdh* transcripts were used as positive control. (**C**) RT-PCR from somatosensory cortex extracts of $Mvp^{+/+}$ and $Mvp^{-/-}$ mice (three replicates each) providing further evidence of a complete full knockout model. Negative control (Neg) was set with water. (**D**) Validation of *Mvp* knockout at the protein level. *Left*. Western blot (WB) from liver extracts of $Mvp^{+/+}$ (n=2), $Mvp^{+/-}$ (n=2) and $Mvp^{-/-}$ (n=3) mice. *Right*. Quantification of the WB membrane. Actin was used as loading control for normalization and $Mvp^{+/+}$ mean set as one for ease of readability of the graph. (**E**) Specificity of *Mvp* deletion on transcription. Design of the different primers (indicated in color at the top of the scheme) and reactions (defined by a number with the corresponding color according to primer association) used to attest the presence of the different transcripts. The size of the expected band is indicated below each reaction. Reaction 1 (Genomic DNA) was specific to genomic DNA, and served as negative control. Reaction 2 (Mvp_202) enabled to see *Mvp_202* (long isoform) expression. Reaction 3 (Mvp_Com) amplified a fragment of transcript that is common to *Mvp_202* and *Mvp_201* isoforms. Reaction 4 (Pagr1a/b) amplified a fragment of transcript that is common to *Pagr1a* and *Pagr1b* isoforms. Reaction 5 (Pagr1b) is specific to *Pagr1b* transcript. RT-PCR for each of these reactions were conducted according to protocols defined in **Material and Methods** on two biological replicates in male $Mvp^{+/+}$ and $Mvp^{-/-}$ mice, in cortex (CTX) and liver (LIV) extracts. The table recapitulates the results obtained: Ø when no bands were seen, +, ++ and +++, for faint, medium and strong band, respectively. (**F**) Viability screen of *Mvp* knockout mice. *Left.* Schematic representation of heterozygous by heterozygous breeding scheme, with expected Mendelian ratio of segregation of the offspring (25% $^{+/+}$; 50% $^{+/-}$; 25% $^{-/-}$). *Center.* Graph represents the ratio of each genotype of successfully genotyped mice obtained from heterozygous by heterozygous breeding (n=617). Proportions followed Mendelian's ratio, thus excluding any mortality due to the presence of the mutation. *Right.* The sex of each mouse also segregated according to 50%/50% expected sex ratio, thus excluding influence of the mutation on viability dependent upon gender.

Figure 47



		CTX		LIV		
Reaction #	Specificity	*Mvp*$^{+/+}$	*Mvp*$^{-/-}$	*Mvp*$^{+/+}$	*Mvp*$^{-/-}$	Comments
1	Genomic DNA	Ø	Ø	Ø	Ø	No genomic contamination
2	Mvp_202	++	Ø	+++	Ø	No sign of *Mvp* transcripts in *Mvp*$^{-/-}$ extracts
3	Mvp_Com	++	Ø	+++	Ø	
4	Pagr1a/b	++	+	+	++	*Pagr1a/b* transcripts are present in *Mvp*$^{-/-}$ extracts
5	Pagr1b	?	Ø	Ø	Ø	No clear expression of *Pagr1b* in the brain

Figure 48 Impact of *Mvp* deletion on neuromorphology in females.

(**A**) Evolution of total brain area during development at three time points, from embryonic age 18.5 (E18.5), to post-natal day 10 (P10) and 120 (P120) in $Mvp^{+/+}$ and $Mvp^{-/-}$. (**B**) *Left*. Schematic representation of the 24 brain regions quantified at Bregma +0.98mm and -1.34mm on coronal section from $Mvp^{+/+}$ and $Mvp^{-/-}$ mice. Colored regions indicate the presence of at least one significant parameter within the brain region at the 0.05 level. White indicates a p-value > 0.05, grey shows not enough data to calculate a p-value. *Right.* Histograms of percentage change relative to $Mvp^{+/+}$ (set as 0) for each of the measured parameters. (**C-H**) Cell count and average cell area determined in different regions of the murine brain: anterior cingulate cortex (**C**), anterior somatosensory cortex (**D**), posterior somatosensory cortex (**E**), caudate putamen (**F**), motor cortex (**G**) and retrosplenial cortex (**H**). (**I**) Schematic representation of a hippocampal neuron in culture and the different measures taken. (**J**) Primary hippocampal neuron measurements of soma area, growth cone area and axonal length cultured from $Mvp^{+/+}$ and $Mvp^{-/-}$. Plots are represented as mean ± SEM. Two-tailed Student *t* test equal variance. $^{*}p<0.05$.

Figure 48

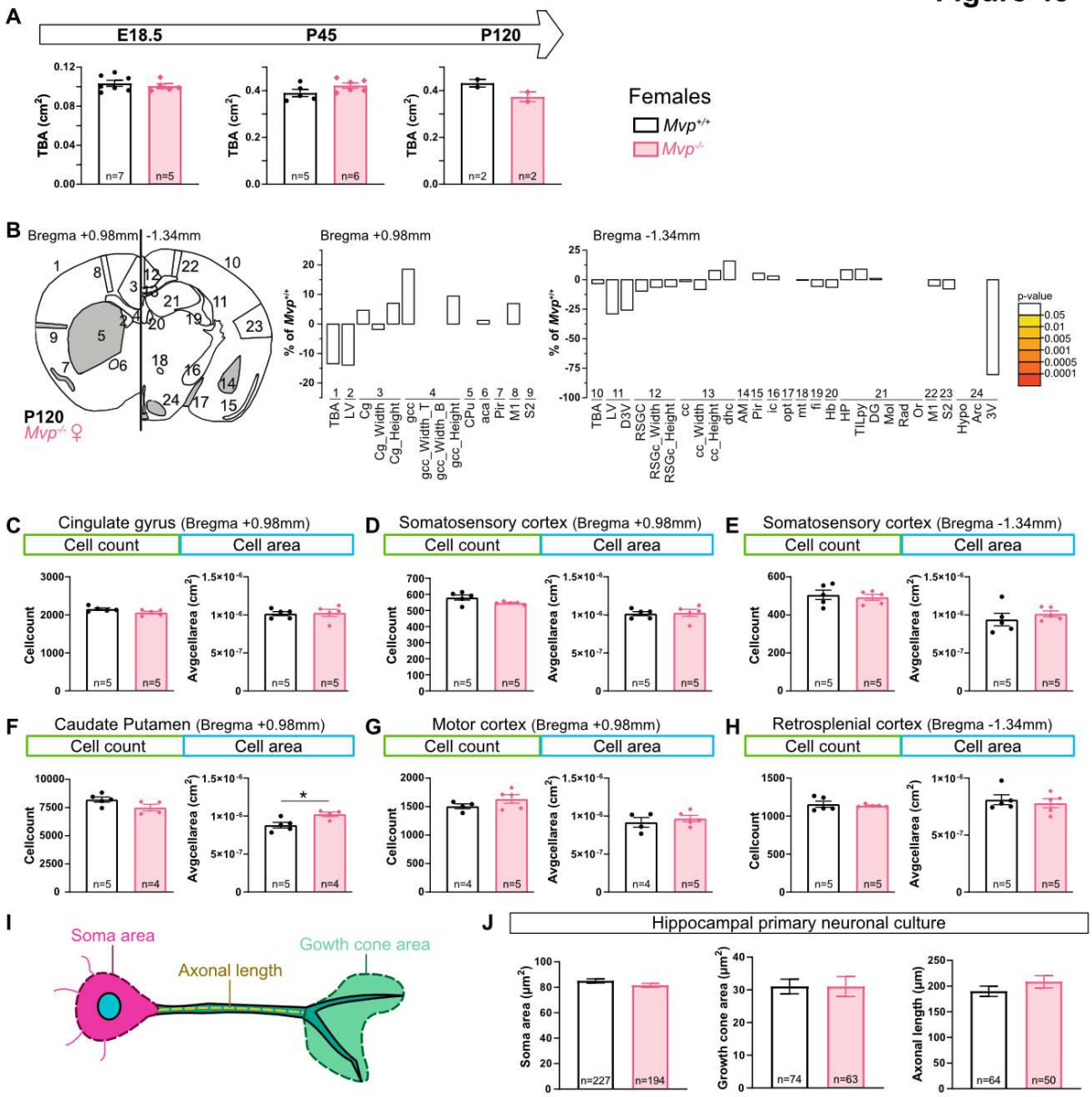

A
E18.5
P45
P120
TBA (cm²)
n=7
n=5
n=5
n=6
n=2
n=2
Females
Mvp+/+
Mvp-/-
B
Bregma +0.98mm
-1.34mm
P120
Mvp-/- ♀
Bregma +0.98mm
Bregma -1.34mm
% of Mvp+/+
p-value
0.05
0.01
0.005
0.001
0.0005
0.0001
TBA
LV
Cg
Cg_Width
Cg_Height
gcc
gcc_Width_T
gcc_Width_B
gcc_Height
CPu
aca
Pir
M1
S2
D3V
RSGC
RSGc_Width
RSGc_Height
cc
cc_Width
cc_Height
dhc
AM
ic
opt
mt
fi
Hb
HP
TILpy
DG
Mol
Rad
Or
Hypo
Arc
3V
C
Cingulate gyrus (Bregma +0.98mm)
D
Somatosensory cortex (Bregma +0.98mm)
E
Somatosensory cortex (Bregma -1.34mm)
F
Caudate Putamen (Bregma +0.98mm)
G
Motor cortex (Bregma +0.98mm)
H
Retrosplenial cortex (Bregma -1.34mm)
Cell count
Cell area
Cellcount
Avgcellarea (cm²)
n=5
n=4
*
I
Soma area
Axonal length
Gowth cone area
J
Hippocampal primary neuronal culture
Soma area (µm²)
Growth cone area (µm²)
Axonal length (µm)
n=227
n=194
n=74
n=63
n=64
n=50

Figure 49 Impact of *Mvp* deletion on neuromorphology in males.

(**A**) Procedure for cortical layer analysis on histological sections involving the measurements of the total area of a cortical layer of interest (red), the number of cells (green) and the average cell area (blue) within the cortical layer of interest (for more details, see **Material and Methods**). (**B-D**) Area analysis of the six cortical layers in the somatosensory cortex, anterior at Bregma +0.98mm (**B**) and posterior at Bregma -1.34mm (**D**), as well as in the motor cortex (Bregma +0.98mm) (**C**) in $Mvp^{+/+}$ and $Mvp^{-/-}$. (**E**) *Left.* Schematic representation of the 22 brain regions quantified at Lateral +0.60mm on sagittal section from $Mvp^{+/+}$ and $Mvp^{-/-}$. Colored regions indicate the presence of at least one significant parameter within the brain region at the 0.05 level. White indicates a p-value > 0.05, grey shows not enough data to calculate a p-value. *Right.* Histograms of percentage change relative to $Mvp^{+/+}$ (set as 0) for each of the measured parameters. (**F-K**) Cell count and average cell area determined in different regions of the murine brain: layers of the anterior somatosensory cortex (**F**), posterior somatosensory cortex (**G**), Caudate Putamen at Bregma +0.98mm (**H**), layers of the motor cortex at Bregma +0.98mm (**I**), motor cortex at Bregma +0.98mm (**J**) and retrosplenial cortex at Bregma -1.34mm (**K**). (**L**) Cell count and average cell area in the cingulate gyrus of $Mvp^{+/+}$ and $Mvp^{-/-}$ mice on sagittal section at Lateral +0.60mm. (**M**) Cell density of Purkinje cells in cerebellum was calculated by dividing the number of Purkinje cells (PC; blue dots) by the total length of the granular layer of the cerebellum (red line) from $Mvp^{+/+}$ and $Mvp^{-/-}$ mice. Plots are represented as mean ± SEM. Two-tailed Student *t* test equal variance (**G-H, J-M**). For layer analysis (**B-D**, **F** and **I**), two-tailed Student *t* tests were corrected for multiple comparisons using the Holm-Sidak method. *$p<0.05$.

Figure 49

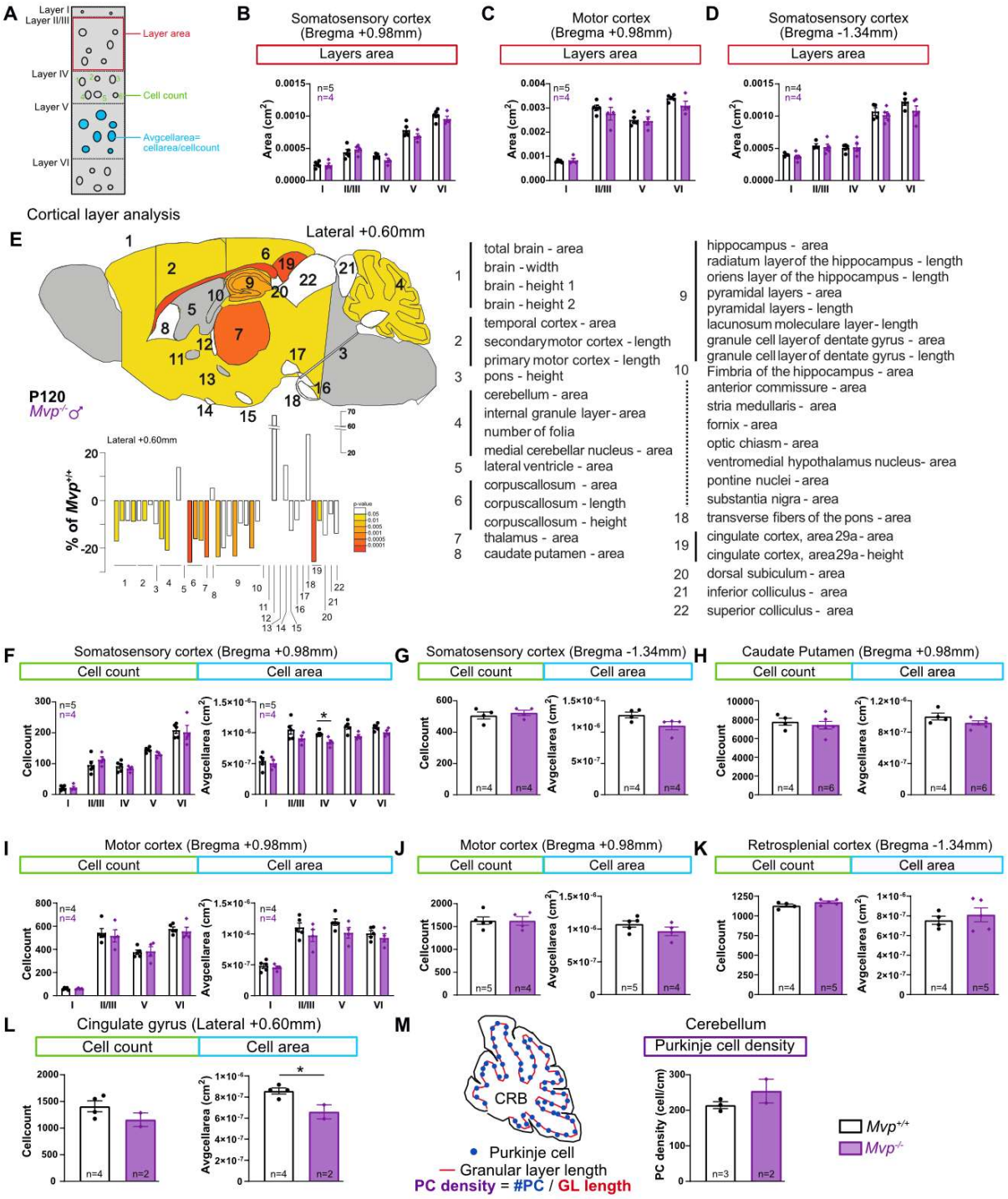

A
Layer I
Layer II/III
Layer IV
Layer V
Layer VI
Layer area
Cell count
Avgcellarea= cellarea/cellcount
B
Somatosensory cortex (Bregma +0.98mm)
Layers area
C
Motor cortex (Bregma +0.98mm)
Layers area
D
Somatosensory cortex (Bregma -1.34mm)
Layers area
Area (cm²)
n=5
n=4
I
II/III
IV
V
VI
Cortical layer analysis
E
Lateral +0.60mm
P120
Mvp-/- ♂
% of Mvp+/+
p-values
1 total brain - area
brain - width
brain - height 1
brain - height 2
2 temporal cortex - area
secondary motor cortex - length
primary motor cortex - length
3 pons - height
4 cerebellum - area
internal granule layer - area
number of folia
medial cerebellar nucleus - area
5 lateral ventricle - area
6 corpuscallosum - area
corpuscallosum - length
corpuscallosum - height
7 thalamus - area
8 caudate putamen - area
9 hippocampus - area
radiatum layer of the hippocampus - length
oriens layer of the hippocampus - length
pyramidal layers - area
pyramidal layers - length
lacunosum moleculare layer - length
granule cell layer of dentate gyrus - area
granule cell layer of dentate gyrus - length
10 Fimbria of the hippocampus - area
anterior commissure - area
stria medullaris - area
fornix - area
optic chiasm - area
ventromedial hypothalamus nucleus - area
pontine nuclei - area
substantia nigra - area
18 transverse fibers of the pons - area
19 cingulate cortex, area 29a - area
cingulate cortex, area 29a - height
20 dorsal subiculum - area
21 inferior colliculus - area
22 superior colliculus - area
F
Somatosensory cortex (Bregma +0.98mm)
Cell count
Cell area
Cellcount
Avgcellarea (cm²)
G
Somatosensory cortex (Bregma -1.34mm)
H
Caudate Putamen (Bregma +0.98mm)
I
Motor cortex (Bregma +0.98mm)
J
Motor cortex (Bregma +0.98mm)
K
Retrosplenial cortex (Bregma -1.34mm)
L
Cingulate gyrus (Lateral +0.60mm)
M
CRB
Purkinje cell
Granular layer length
PC density = #PC / GL length
Cerebellum
Purkinje cell density
PC density (cell/cm)
Mvp+/+
Mvp-/-

Figure 50 Analysis of *Mvp* loss in the cingulate gyrus.

(**A-B**) Electrophysiological readouts: amplitude (**A**) and inter-event interval (**B**) of miniature excitatory post-synaptic currents (mEPSCs) measured by patch-clamp on pyramidal neurons from deep layers of the cingulate gyrus from *Mvp*$^{+/+}$ and *Mvp*$^{-/-}$ mice. (**C**) Pictures of sections (2µm thick) of cingulate gyrus of *Mvp*$^{+/+}$ and *Mvp*$^{-/-}$ mice stained with toluidine blue. (**D**) Electronic microscopy from cingulate gyrus of *Mvp*$^{+/+}$ and *Mvp*$^{-/-}$ mice. N, nucleus; M, mitochondria; Ly, Lysosome. (**E**) Scatter-plot comparing the mean normalized counts for each condition. A pseudo count of 1 was added to all values in order to represent genes that are not expressed in one condition. Significant genes were selected using the following thresholds: adjusted p-value lower than 0.05 and absolute value of log2 Fold-Change greater than 1. Only *Mvp* (red dot) expression was significantly affected in *Mvp*$^{-/-}$ (n=6) compared to *Mvp*$^{+/+}$ (n=6) mice. Two-tailed Welch's unequal variances t-test. *p<0.05.

Figure 50

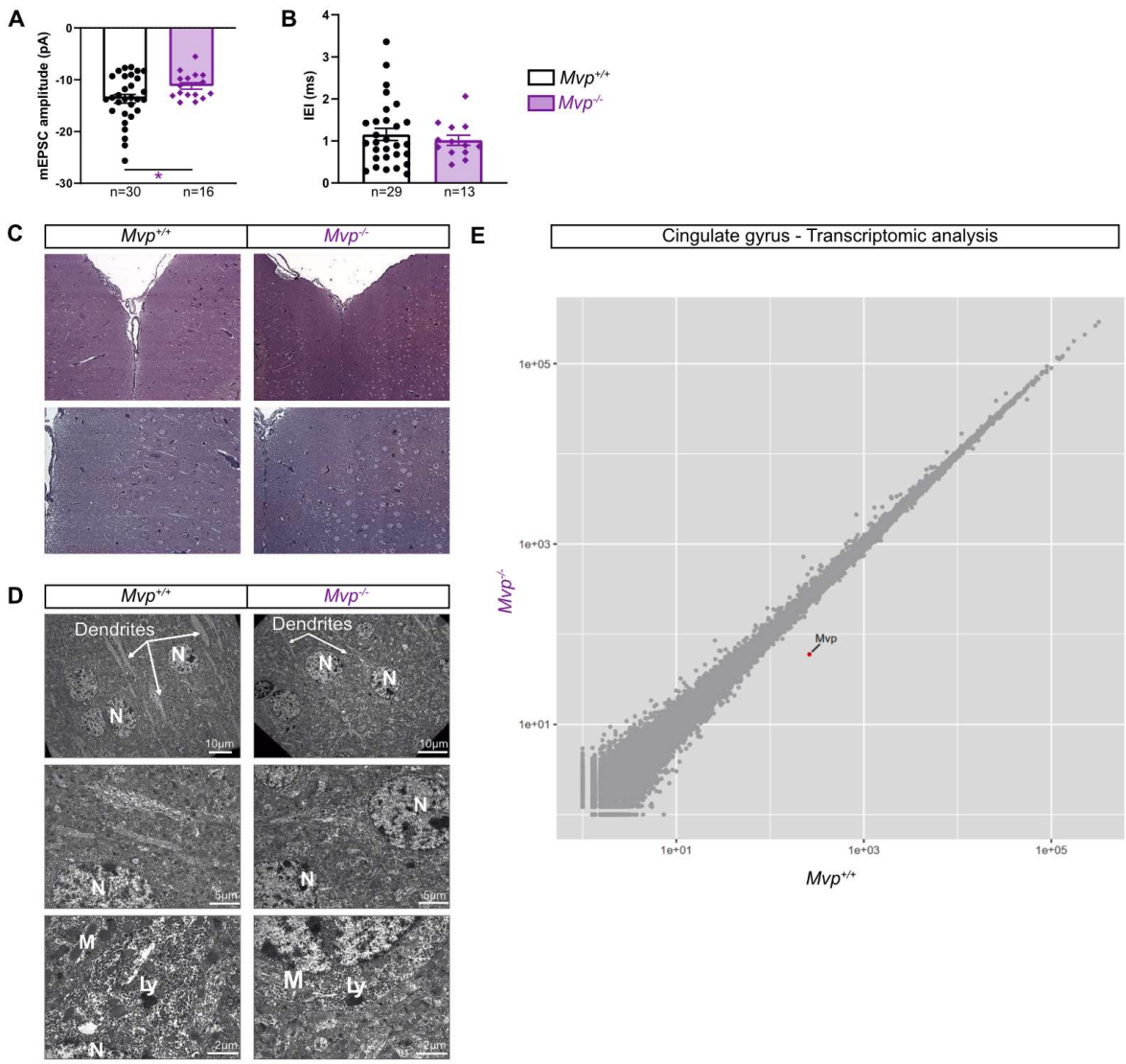


Figure 51 Behavioral characterization of *Mvp*::*Mapk3* double heterozygous.

(**A**-**B**) *Left*. Western blot analysis of ERK-expression on proteins extracted from cortical tissues in male (**A**) and female (**B**) showing ERK1 and ERK2 differential expression levels in five *Mvp*$^{+/-}$*::Mapk3*$^{+/-}$ compared to five matched *Mvp*$^{+/+}$*::Mapk3*$^{+/+}$ in each group. *Right*. Quantification of the Western blots relative to *Mvp*$^{+/+}$*::Mapk3*$^{+/+}$. Normalization was done on vinculin expression. Bullet points represents *Mvp*$^{+/+}$*::Mapk3*$^{+/+}$ mice and square *Mvp*$^{+/-}$*::Mapk3*$^{+/-}$. Males are represented in blue and females in pink. (**C**) Weekly body weight tracking from age 11- to 25 weeks in male *Mvp*$^{+/-}$*::Mapk3*$^{+/-}$ and *Mvp*$^{+/+}$*::Mapk3*$^{+/+}$ mice (n=12 per group). (**D**) Grip strength was evaluated on either the four (*left*) or the two anterior (*right*) paws, and normalized on mouse body weight. (**E**) Rotarod test. *Learning*. Time (in seconds) spent on the rotarod at increasing speed from 4 to 40 rotation per minute (rpm) over five minutes, during four sessions on three consecutive days. *Latency before falling*. On the day of testing, latency before falling off the rotarod at rotation speeds (4, 10, 16, 22, 28, 34 and 40 rpm). Maximum per trial was set to 120 seconds. (**F**) Sucrose preference was measured as the percentage of sucrose consumption over the total consumption during, habituation (Hab, 1h), and testing days (Test-Day1 and Test–Day2, 12h during dark cycle). (**G**) Y-maze. Percentage of alternation was calculated by dividing the number of alternations by the number of possible triads x 100. (**H**) Novel Object Recognition (NOR). *Exploration*. Exploration time (in seconds) of the different objects (ObjectA in 1st phase, and ObjectFam and ObjectNew, in 2nd phase). *Recognition*. Percentage of exploration of the new object over total object exploration (ObjectNew/ (ObjectNew+ObjectFam)). 50% is considered as no recognition of the familiar object. (**I**) Three-chamber test. Percentage of the duration of interaction with the Old Individual (*OldInd*) or New Individual (*NewInd*) over total interaction time during the 3rd phase. (**J**) Social Interaction. Graphs plot the duration of pawing, following, interaction and sniffing behaviors between two individuals from the same genotype. (**K**) Marble. Plot represents the number of marbles according to their level of coverage (75%cover, 100%Cover or NoCover). (**L**) Circadian Activity. *Locomotion.* Tracking of locomotor activity (in cm) every hour over the duration of the test (32h). Statistical analysis is performed over combined Dark and Light phase (from 7pm to 7am). *Pellet* and *Water*. Weight of pellets and water in grams consumed over 32 hours. (**M**) Pre-Pulse Inhibition. *Startle response*. Averaged amplitude of the acoustic startle reflex (arbitrary unit) following 60dB (Background Noise), 70dB (P70), 75dB (P75), 85dB (P85), 90dB (P90) and 110dB (ST110) acoustic stimulus. *PPI*. Percentage of Pre-Pulse Inhibition calculated as PPX %=(PPX-ST110)/ST110. A comprehensive description of behavioral tests, raw datasets and additional findings are provided in **Material and Methods**, **Dataset 4** and **Figure 41**, respectively. Mean and standard error of the mean are shown in the graphs. Two-tailed Student *t* test equal variance (**A**, **B**, **D**, **F**, **G**, **H**, **I**, **J**, **K**, **L** and **M**) and two-way ANOVA with Sidak post hoc (**C**, **E** and **M**). *p<0.05.

Figure 51

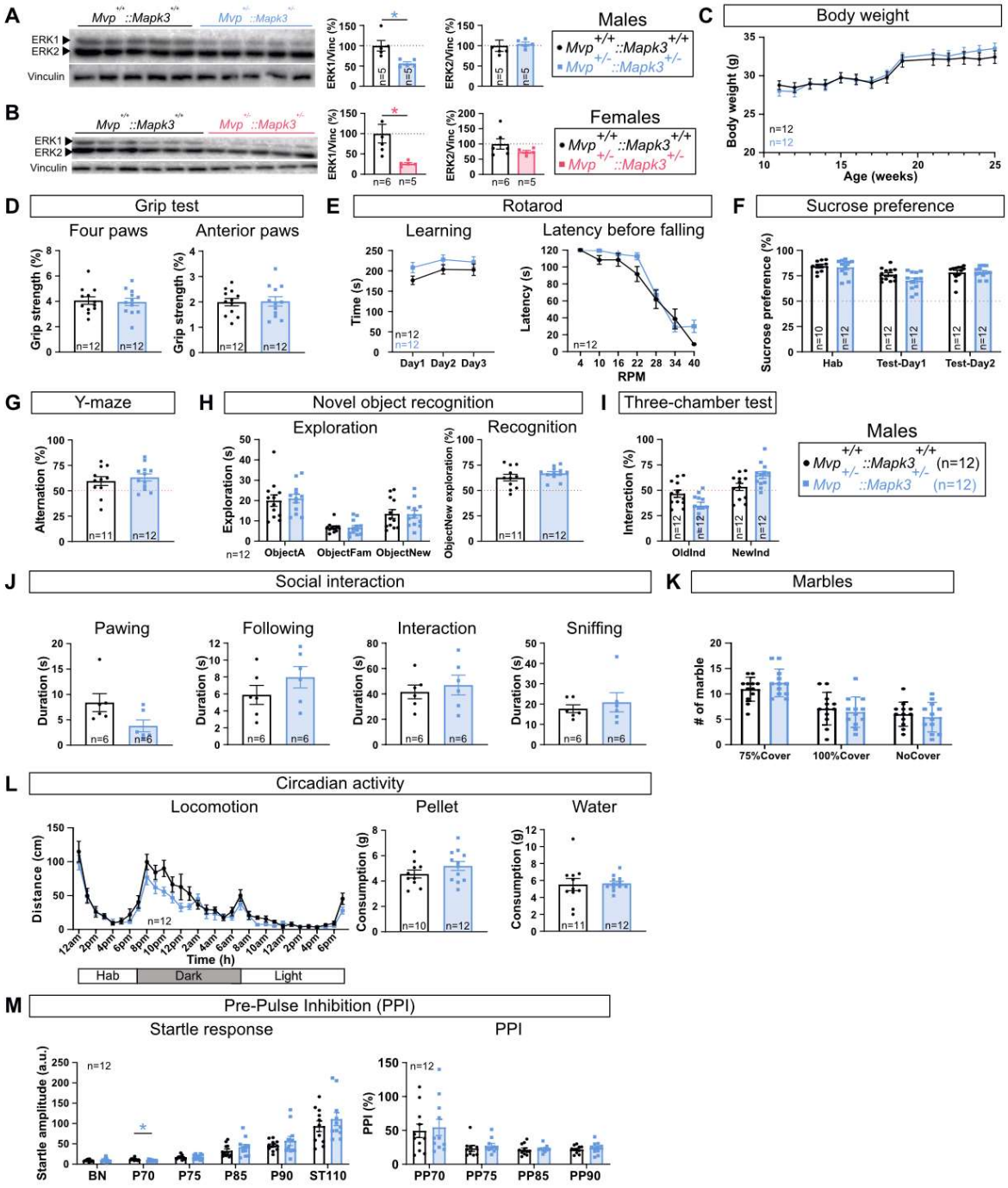

A
Mvp +/+::Mapk3 +/+
Mvp +/-::Mapk3 +/-
ERK1
ERK2
Vinculin
ERK1/Vinc (%)
ERK2/Vinc (%)
n=5
Males
Mvp +/+::Mapk3 +/+
Mvp +/-::Mapk3 +/-
B
n=6
Females
C
Body weight
Body weight (g)
Age (weeks)
n=12
D
Grip test
Four paws
Anterior paws
Grip strength (%)
E
Rotarod
Learning
Latency before falling
Time (s)
Latency (s)
Day1
Day2
Day3
RPM
F
Sucrose preference
Sucrose preference (%)
Hab
Test-Day1
Test-Day2
G
Y-maze
Alternation (%)
H
Novel object recognition
Exploration
Recognition
Exploration (s)
ObjectA
ObjectFam
ObjectNew
ObjectNew exploration (%)
I
Three-chamber test
Interaction (%)
OldInd
NewInd
Males
Mvp +/+::Mapk3 +/+ (n=12)
Mvp +/-::Mapk3 +/- (n=12)
J
Social interaction
Pawing
Following
Interaction
Sniffing
Duration (s)
K
Marbles
# of marble
75%Cover
100%Cover
NoCover
L
Circadian activity
Locomotion
Pellet
Water
Distance (cm)
Consumption (g)
Time (h)
Hab
Dark
Light
M
Pre-Pulse Inhibition (PPI)
Startle response
PPI
Startle amplitude (a.u.)
PPI (%)
BN
P70
P75
P85
P90
ST110
PP70
PP75
PP85
PP90

Figure 52 Behavioral studies of *Mvp* homozygous mice.

(**A**) Open field test. *Left*. Distance travelled, recorded in centimeters every five minutes, over the time of the test (30 minutes). *Middle*. Percentage of distance travelled in the center of the arena over total distance travelled. *Right*. Latency in seconds before the first entrance in the center of the arena. (**B**) Elevated plus maze. *Left*. Number of entrances in the open arm. *Middle*. Percentage of time spent in the open arm over total time of the test (5 minutes). *Right*. Number of times the mice bend its nose to the ground over the open arm. (**C**) Fear conditioning test. *Conditioning.* Freezing time (in seconds) during the first 4 minutes (Hab-1 and Hab-2, two minutes each) and after unconditional stimulus (POST-US, 2 minutes). *Context.* Freezing time (in seconds) of the mice when replaced in the same context after 24 hours (Cont-1-3, two minutes each). *Cue*. Freezing time (in seconds) during the different phase (two minutes each) of the cued test. The cue (light and sound) is presented during the entire Cue1 and Cue2 phase. (**D**) Latency in seconds before the first immobility of the mouse in the tail suspension test. (**E**) Pentylenetetrazol (PTZ)-test. Duration of the clonic phase of status epilepticus after PTZ intraperitoneal injection. (**F**) Rotarod test. *Learning*. Time (in seconds) spent on the rotarod at increasing speed from 4 to 40 rotation per minute (rpm) over five minutes, during four sessions on three consecutive days. *Test.* On the day of testing, latency before falling off the rotarod during different fixed rotation speeds (4, 10, 16, 22, 28, 34 and 40 rpm). Maximum time per trial was set to 120 seconds. (**G**) Sucrose preference was measured as the percentage of sucrose consumption over the total consumption during habituation (Hab, 1h), and testing days (Test-Day1 and Test–Day2, 12 hours during dark cycle). (**H**) Y-maze. Percentage of alternation was calculated by dividing the number of alternations by the number of possible triads x 100. (**I**) Novel Object Recognition (NOR). *Exploration.* Exploration time (in seconds) of the different objects (ObjectA in 1st phase, and ObjectFam and ObjectNew, in 2nd phase). *Recognition*. Percentage of exploration of the new object over total object exploration (ObjectNew/(ObjectNew+ObjectFam)). 50% was considered as no recognition of the familiar object. (**J**) Three-chamber test. Percentage of the duration of interaction with the Old Individual (OldInd) or New Individual (NewInd) over total interaction time during the 3rd phase of the test. (**K**) Marble: Plot represents the number of marbles according to their level of coverage (75%Cover, 100%Cover or NoCover). (**L**) Circadian Activity. *Locomotion.* Tracking of locomotor activity (in centimeters) every hour over the duration of the test (32 hours). Statistical analysis was performed over the combined Dark and Light phase (from 7pm to 7pm). *Pellet* and *Water* consumptions. Weight of pellets and water in grams consumed over 32 hours. (**M**) Pre-Pulse Inhibition. *Startle response*. Averaged amplitude of the acoustic startle reflex (arbitrary unit) following 60 dB (Background Noise = BN), 70 dB pulse (P70), 75 dB (P75), 85 dB (P85), 90 dB (P90) and 110 dB (ST110) acoustic stimulus. PPI: Percentage of Pre-Pulse (PP) Inhibition calculated as PPX %=(PPX-ST110)/ST110. (**N**) Body weight. Body weight tracking (in grams) from 15 to 24 weeks of age. *Mvp*$^{+/+}$ are represented by black dots and *Mvp*$^{-/-}$ by purple diamonds. A comprehensive description of behavioral tests and raw datasets are provided in **Material and Methods** and **Dataset 5**, respectively. Mean and standard error of the mean are shown in the graphs. Sample size is indicated within each histogram. Two-tailed Student t test equal variance (**A**, **B**, **D**, **E**, **G**, **H**, **I**, **J**, **K**, **L** and **M**) and two-way ANOVA with Sidak post hoc (**A**, **C**, **F**, **L** and **N**).

Figure 52

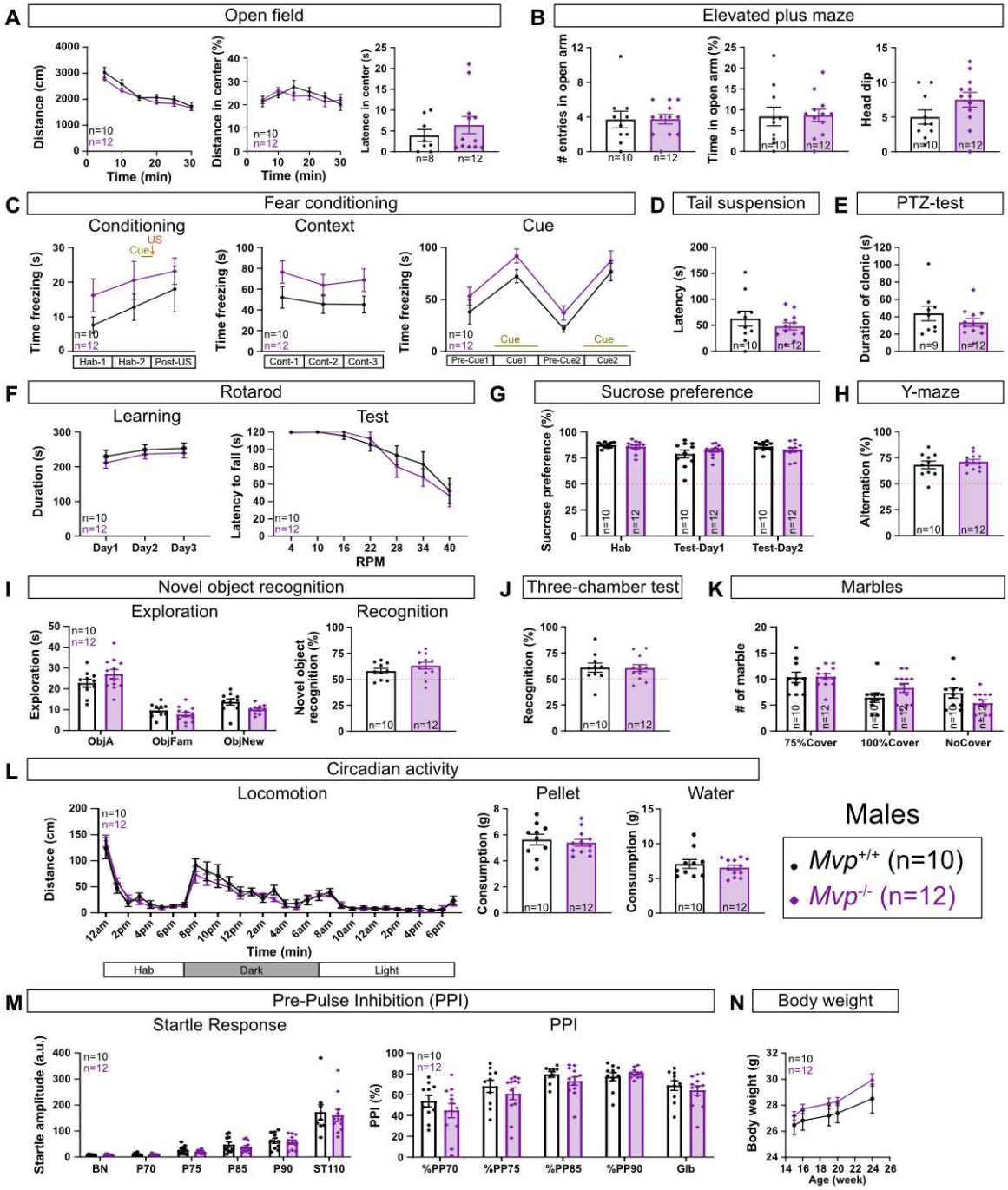
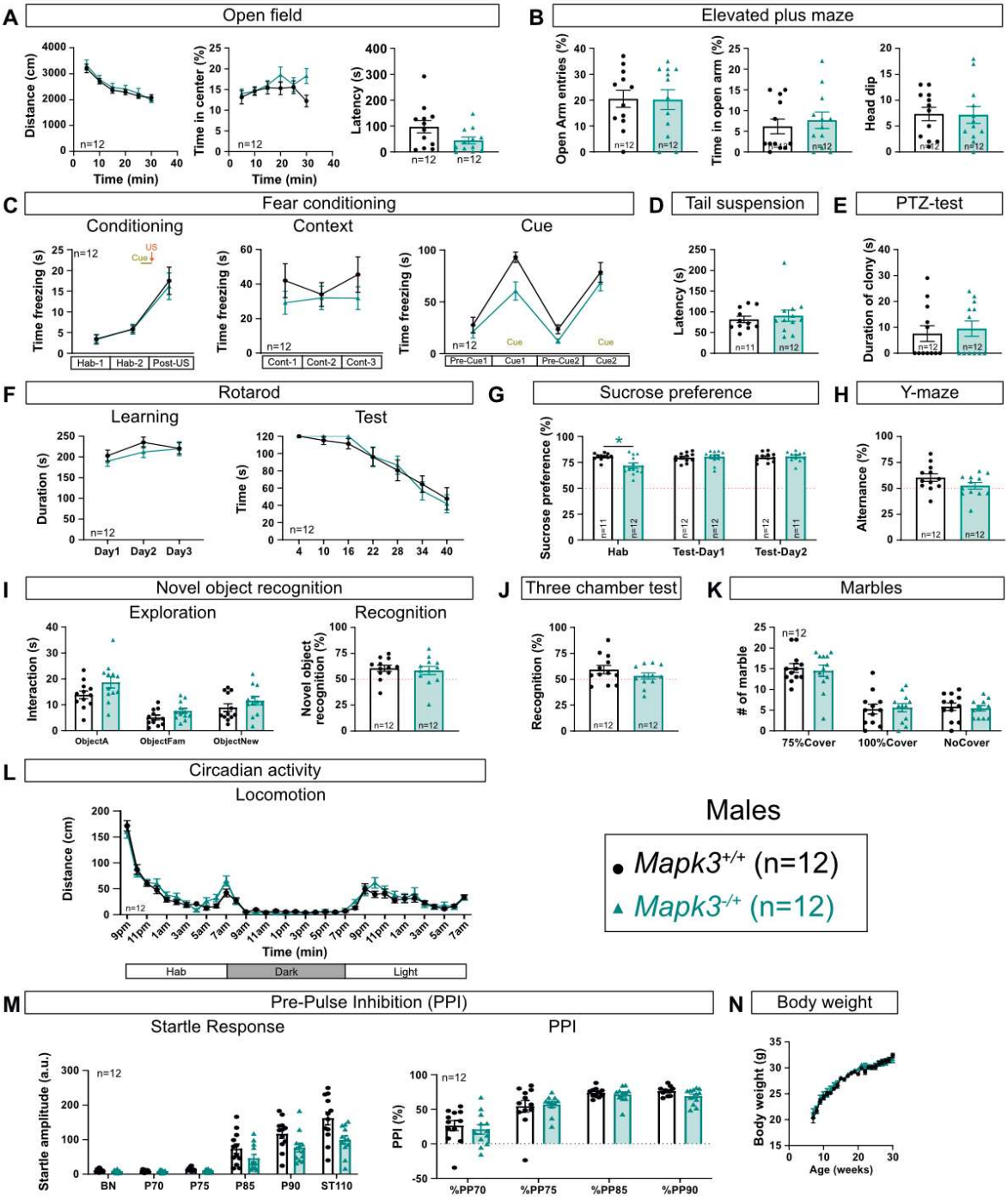

A
Open field
Distance (cm)
Time (min)
Distance in center (%)
Latence in center (s)
n=10
n=12
n=8
B
Elevated plus maze
entries in open arm
Time in open arm (%)
Head dip
C
Fear conditioning
Conditioning
Context
Cue
Time freezing (s)
US
Cue
Hab-1
Hab-2
Post-US
Cont-1
Cont-2
Cont-3
Pre-Cue1
Cue1
Pre-Cue2
Cue2
D
Tail suspension
Latency (s)
E
PTZ-test
Duration of clonic (s)
n=9
F
Rotarod
Learning
Test
Duration (s)
Latency to fall (s)
Day1
Day2
Day3
RPM
G
Sucrose preference
Sucrose preference (%)
Hab
Test-Day1
Test-Day2
H
Y-maze
Alternation (%)
I
Novel object recognition
Exploration
Recognition
Exploration (s)
Novel object recognition (%)
ObjA
ObjFam
ObjNew
J
Three-chamber test
Recognition (%)
K
Marbles
# of marble
75%Cover
100%Cover
NoCover
L
Circadian activity
Locomotion
Pellet
Water
Distance (cm)
Consumption (g)
Time (min)
Hab
Dark
Light
Males
Mvp+/+ (n=10)
Mvp-/- (n=12)
M
Pre-Pulse Inhibition (PPI)
Startle Response
PPI
Startle amplitude (a.u.)
PPI (%)
BN
P70
P75
P85
P90
ST110
%PP70
%PP75
%PP85
%PP90
Glb
N
Body weight
Body weight (g)
Age (week)

Figure 53 Behavioral studies of *Mapk3* heterozygous mice.

(**A**) Open field test. *Left*. Distance travelled, recorded in centimeters every five minutes, over the time of the test (30 minutes). *Middle*. Percentage of distance travelled in the center of the arena over total distance travelled. *Right*. Latency in seconds before the first entrance in the center of the arena. (**B**) Elevated plus maze. *Left*. Number of entrances in the open arm. *Middle*. Percentage of time spent in the open arm over total time of the test (5 minutes). *Right*. Number of times the mice bend its nose to the ground over the open arm. (**C**) Fear conditioning test. *Conditioning.* Freezing time (in seconds) during the first 4 minutes (Hab-1 and Hab-2, two minutes each) and after unconditional stimulus (POST-US, 2 minutes). *Context.* Freezing time (in seconds) of the mice when replaced in the same context after 24 hours (Cont-1-3, two minutes each). *Cue*. Freezing time (in seconds) during the different phase (two minutes each) of the cued test. The cue (light and sound) is presented during the entire Cue1 and Cue2 phase. (**D**) Latency in seconds before the first immobility of the mouse in the tail suspension test. (**E**) Pentylenetetrazol (PTZ)-test. Duration of the clonic phase of status epilepticus after PTZ intraperitoneal injection. (**F**) Rotarod test. *Learning*. Time (in seconds) spent on the rotarod at increasing speed from 4 to 40 rotation per minute (rpm) over five minutes, during four sessions on three consecutive days. *Test.* On the day of testing, latency before falling off the rotarod during different fixed rotation speeds (4, 10, 16, 22, 28, 34 and 40 rpm). Maximum time per trial was set to 120 seconds. (**G**) Sucrose preference was measured as the percentage of sucrose consumption over the total consumption during habituation (Hab, 1h), and testing days (Test-Day1 and Test–Day2, 12 hours during dark cycle). (**H**) Y-maze. Percentage of alternation was calculated by dividing the number of alternations by the number of possible triads x 100. (**I**) Novel Object Recognition (NOR). *Exploration.* Exploration time (in seconds) of the different objects (ObjectA in 1st phase, and ObjectFam and ObjectNew, in 2nd phase). *Recognition*. Percentage of exploration of the new object over total object exploration (ObjectNew/ (ObjectNew+ObjectFam)). 50% was considered as no recognition of the familiar object. (**J**) Three-chamber test. Percentage of the duration of interaction with the Old Individual (OldInd) or New Individual (NewInd) over total interaction time during the 3rd phase of the test. (**K**) Marble: Plot represents the number of marbles according to their level of coverage (75%Cover, 100%Cover or NoCover). (**L**) Circadian Activity. *Locomotion.* Tracking of locomotor activity (in centimeters) every hour over the duration of the test (32 hours). Statistical analysis was performed over the combined Dark and Light phase (from 7pm to 7pm). (**M**) Pre-Pulse Inhibition. *Startle response*. Averaged amplitude of the acoustic startle reflex (arbitrary unit) following 60 dB (Background Noise = BN), 70 dB pulse (P70), 75 dB (P75), 85 dB (P85), 90 dB (P90) and 110 dB (ST110) acoustic stimulus. PPI: Percentage of Pre-Pulse (PP) Inhibition calculated as PPX %=(PPX-ST110)/ST110. (**N**) Body weight. Body weight tracking (in grams) from 15 to 24 weeks of age. *Mapk3*$^{+/+}$ are represented by black dots and *Mapk3*$^{+/-}$ by blue triangle. A comprehensive description of behavioral tests and raw datasets are provided in **Material and Methods** and **Dataset 6**, respectively. Mean and standard error of the mean are shown in the graphs. Sample size is indicated within each histogram. Two-tailed Student t test equal variance (**A**, **B**, **D**, **E**, **G**, **H**, **I**, **J**, **K** and **M**) and two-way ANOVA with Sidak post hoc (**A**, **C**, **F**, **L** and **N**). *p<0.05.

Figure 53


A
Open field
Distance (cm)
Time (min)
Time in center (%)
Latency (s)
n=12
B
Elevated plus maze
Open Arm entries (%)
Time in open arm (%)
Head dip
C
Fear conditioning
Conditioning
Context
Cue
Time freezing (s)
US
Cue
Hab-1
Hab-2
Post-US
Cont-1
Cont-2
Cont-3
Pre-Cue1
Cue1
Pre-Cue2
Cue2
D
Tail suspension
Latency (s)
E
PTZ-test
Duration of clony (s)
F
Rotarod
Learning
Test
Duration (s)
Time (s)
Day1
Day2
Day3
G
Sucrose preference
Sucrose preference (%)
Hab
Test-Day1
Test-Day2
H
Y-maze
Alternance (%)
I
Novel object recognition
Exploration
Recognition
Interaction (s)
Novel object recognition (%)
ObjectA
ObjectFam
ObjectNew
J
Three chamber test
Recognition (%)
K
Marbles
# of marble
75%Cover
100%Cover
NoCover
L
Circadian activity
Locomotion
Distance (cm)
Time (min)
Hab
Dark
Light
Males
Mapk3+/+ (n=12)
Mapk3-/+ (n=12)
M
Pre-Pulse Inhibition (PPI)
Startle Response
PPI
Startle amplitude (a.u.)
BN
P70
P75
P85
P90
ST110
PPI (%)
%PP70
%PP75
%PP85
%PP90
N
Body weight
Body weight (g)
Age (weeks)

4.6 Tables and legends

Table 3 Description of mouse models used in the study.

The 30 protein-coding mouse genes encompassing the 16p11.2-syntenic region are listed in column (A). Details are given according to Ensembl GRCm38.p6 and consist of the accession number, the strand (+: forward; - : reverse), the genomic coordinates (base pair), the total length of the gene (base pair) and the full name description (columns B-G, respectively). The 20 genes that underwent the neuroanatomical screen are indicated in column H. Column I provides the complete allele name. The core strain of the mouse is indicated in column J, the origin of the mice in column K (name of the production center or of the collaborator who provided the material) and finally column L the material received (living mouse or dissected brain sample).

A	B	C	D	E	F	G	H	I	J	K	L
Gene	Accession	Strand	Start (bp)	End (bp)	Length (bp)	Description	Analysed	Allele Name	Core Strain	Source	Material received
Sult1a	ENSMUSG00000030711	-	126672794	126676503	3709	sulfotransferase family 1A, phenol-preferring, member 1					
Slx1b	ENSMUSG00000059772	-	126689468	126695784	6316	SLX1 structure-specific endonuclease subunit homolog B	X	*Slx1b<em1(IMPC)Ics>*	B6N	ICS Illkirch France	Mouse
Bola2	ENSMUSG00000047721	-	126695401	126699798	4397	bolA-like 2 (E. coli)	X	*Bola2<tm1(KOMP)Wtsi>*	B6N	Alexandre Reymond Lausanne Switzerland	Brain
Coro1a	ENSMUSG00000030707	-	126699773	126707787	8014	Coronin, actin binding protein 1A	X	*Coro1a<tm1.1(KOMP)Vlcg>*	B6N	Jean Pieters Basel Switzerland	Brain
Mapk3	ENSMUSG00000063065	+	126759601	126765819	6218	Mitogen-activated protein kinase 3	X	*Mapk3<tm1(Gpg)>*	B6J	Gilles Pagès Nice France	Mouse
Gdpd3	ENSMUSG00000030703	+	126766334	126775649	9315	glycerophosphodiester phosphodiesterase domain containing 3	X	*Gdpd3<em1(ICS)>*	B6N	ICS Illkirch France	Mouse
Ypel3	ENSMUSG00000042675	+	126776955	126780514	3559	yippee like 3					
Tbx6	ENSMUSG00000030699	+	126781483	126785560	4077	T-box 6	X	*Tbx6<tm1(Pa)>*	B6J	Virginia Papaioannou New York USA	Mouse
Ppp4c	ENSMUSG00000030697	-	126785866	126792496	6630	protein phosphatase 4, catalytic subunit	X	*Ppp4c<em1(IMPC)Ics>*	B6N	ICS Illkirch France	Mouse
Aldoa	ENSMUSG00000030695	-	126795234	126800751	5517	aldolase A, fructose-bisphosphate					
Fam57b	ENSMUSG00000058966	+	126797668	126830219	32551	family with sequence similarity 57, member B	X	*Fam57b<tm1a(ICS)>*	B6N	ICS Illkirch France	Mouse
4930451I11Rik	ENSMUSG00000045989	-	126830468	126831639	1171	RIKEN cDNA 4930451I11 gene					
Doc2a	ENSMUSG00000052301	+	126847416	126852705	5289	double C2, alpha	X	*Doc2a<tm1(Ytk)>*	B6J	Yoshimi Takai Kobe Japan	Mouse
Ino80e	ENSMUSG00000030689	-	126850960	126862377	11417	INO80 complex subunit E	X	*Ino80e<tm1a(EUCOMM)Wtsi>/Cnrm*	B6N	CNR Rome Italy	Mouse
Hirip3	ENSMUSG00000042606	+	126861972	126865377	3405	HIRA interacting protein 3	X	*Hirip3<tm1(KOMP)Mbp>*	B6N	UC Davis USA	Mouse
Taok2	ENSMUSG00000059981	-	126865678	126884703	19025	TAO kinase 2	X	*Taok2<tm1b(ICS)>*	B6N	ICS Illkirch France	Mouse
Tmem219	ENSMUSG00000060538	-	126886171	126922917	36746	transmembrane protein 219					
Kctd13	ENSMUSG00000030685	+	126928879	126945631	16752	potassium channel tetramerisation domain containing 13	X	*Kctd13<em2(IMPC)Ics>*	B6N	ICS Illkirch France	Mouse
								Kctd13<tm1.1Nk>	B6J	Nicholas Katsanis Duke USA	Mouse
Asphd1	ENSMUSG00000046378	-	126945567	126949582	4015	aspartate beta-hydroxylase domain containing 1					
Sez6l2	ENSMUSG00000030683	+	126950563	126970606	20043	seizure related 6 homolog like 2	X	*Sez6l2<tm1a(EUCOMM)Hmgu>*	B6N	ICS Illkirch France	Mouse
Cdipt	ENSMUSG00000030682	+	126975914	126980501	4587	CDP-diacylglycerol--inositol 3-phosphatidyltransferase					
Mvp	ENSMUSG00000030681	-	126986860	127014621	27761	major vault protein	X	*Mvp<Gt(IST10954E2)Tigm>*	B6N	TIGM USA	Mouse
Pagr1a	ENSMUSG00000030680	-	127015033	127017352	2319	PAXIP1 associated glutamate rich protein 1A					
Prrt2	ENSMUSG00000045114	-	127017531	127021211	3680	proline-rich transmembrane protein 2	X	*Prrt2<tm1a(KOMP)Wtsi>*	B6N	WSI Cambridge UK	Brain
Maz	ENSMUSG00000030678	-	127022130	127027037	4907	MYC-associated zinc finger protein (purine-binding transcription factor)	X	*Maz<em1(IMPC)Ics>*	B6N	ICS Illkirch France	Mouse
Kif22	ENSMUSG00000030677	-	127027729	127042471	14742	kinesin family member 22					
Zg16	ENSMUSG00000049350	-	127050156	127087328	37172	zymogen granule protein 16	X	*Zg16<em1(IMPC)Ics>*	B6N	ICS Illkirch France	Mouse
AI467606	ENSMUSG00000045165	+	127091359	127093986	2627	expressed sequence AI467606					
Qprt	ENSMUSG00000030674	-	127107114	127122226	15112	quinolinate phosphoribosyltransferase	X	*Qprt<tm1.1>*	B6J	Tsutomu Fukuwatari Hikone Japan	Brain
Spn	ENSMUSG00000051457	-	127132232	127137823	5591	sialophorin	X	*Spn<em1(IMPC)Ics>*	B6N	ICS Illkirch France	Mouse

Table 4 List of 67 coronal neuroanatomical coronal measurements.

This table enlists combined morphological phenotypes studied in the coronal neuroanatomical screen. Stereotaxic coordinates (column A) of the two sections of interest are indicated. Association of each parameter with a brain region is indicated in column B. Column C shows the number attributed to each region (**Figure 40**, **Figure 45** and **Figure 48**). Column D gives the merged name of the parameter (used in **Dataset 2**), column E the full name, column F the description and column G the unit the measurement.

A	B	C	D	E	F	G
Section of interest	**Brain_region**	**ID**	**Acronym**	**Full Acronym**	**Description**	**Unit**
Bregma +0.98mm	Brain size	1	1_TBA	1_TBA	Total Brain Area	cm²
	Ventricle	2	1_LV	1_LV_L	Area of Lateral Ventricle_Left	cm²
				1_LV_R	Area of Lateral Ventricle_Right	cm²
	Cortex	3	1_Cg	1_Cg_L	Area of Cingulate gyrus_Left	cm²
				1_Cg_R	Area of Cingulate gyrus_Right	cm²
			1_Cg_Width	1_Cg_Width_L	Width of Cingulate gyrus_Left	cm
				1_Cg_Width_R	Width of Cingulate gyrus_Right	cm
			1_Cg_Height	1_Cg_Height	Height of Cingulate gyrus	cm
	Commissure	4	1_gcc	1_gcc	Area of genu of corpus callosum	cm²
			1_gcc_Width_T	1_gcc_Width_T	Width of genu of corpus callosum_Top	cm
			1_gcc_Width_B	1_gcc_Width_B	Width of genu of corpus callosum_Bottom	cm
			1_gcc_Height	1_gcc_Height	Height of genu of corpus callosum	cm
		6	1_aca	1_aca_L	Area of anterior commissure-anterior part_Left	cm²
				1_aca_R	Area of anterior commissure-anterior part_Right	cm²
	Subcortical structure	5	1_CPu	1_CPu_L	Area of Caudate Putamen_Left	cm²
				1_CPu_R	Area of Caudate Putamen_Right	cm²
	Cortex	7	1_Pir	1_Pir_L	Area of Piriform cortex_Left	cm²
				1_Pir_R	Area of Piriform cortex_Right	cm²
		8	1_M1	1_M1_L	Height of primary Motor cortex_Left	cm
				1_M1_R	Height of primary Motor cortex_Right	cm
		9	1_S2	1_S2_L	Height of secondary Somatosensory cortex_Left	cm
				1_S2_R	Height of secondary Somatosensory cortex_Right	cm
Bregma -1.34mm	Brain size	10	2_TBA	2_TBA	Total Brain Area	cm²
	Ventricle	11	2_LV	2_LV_L	Area of Lateral Ventricle_Left	cm²
				2_LV_R	Area of Lateral Ventricle_Right	cm²
			2_D3V	2_D3V	Area of dorsal third ventricle	cm²
		24	2_3V	2_3V	Area of third ventricle	cm²
	Cortex	12	2_RSGc	2_RSGc_L	Area of Retrosplenial Granular cortex_Left	cm²
				2_RSGc_R	Area of Retrosplenial Granular cortex_Right	cm²
			2_RSGc_Width	2_RSGc_Width_L	Width of Retrosplenial Granular cortex_Left	cm
				2_RSGc_Width_R	Width of Retrosplenial Granular cortex_Right	cm
			2_RSGc_Height	2_RSGc_Height	Height of Retrosplenial Granular cortex	cm
	Commissure	13	2_cc	2_cc	Area of corpus callosum	cm²
			2_cc_Width	2_cc_Width	Width of corpus callosum	cm
			2_cc_Height	2_cc_Height	Height of corpus callosum	cm
			2_dhc	2_dhc	Area of dorsal hippocampal commissure	cm²
	Subcortical structure	14	2_AM	2_AM_L	Area of Amygdala_Left	cm²
				2_AM_R	Area of Amygdala_Right	cm²
		18	2_mt	2_mt_L	Area of mammillothalamic tract_Left	cm²
				2_mt_R	Area of mammillothalamic tract_Right	cm²
		16	2_ic	2_ic_L	Area of internal capsule_Left	cm²
				2_ic_R	Area of internal capsule_Right	cm²
		17	2_opt	2_opt_L	Area of optic tract_Left	cm²
				2_opt_R	Area of optic tract_Right	cm²
		19	2_fi	2_fi_L	Area of fimbria of hippocampus_Left	cm²
				2_fi_R	Area of fimbria of hippocampus_Right	cm²
		20	2_Hb	2_Hb_L	Area of Habenular nucleus_Left	cm²
				2_Hb_R	Area of Habenular nucleus_Right	cm²
		24	2_HYPO	2_HYPO_L	Area of Hypothalamus_Left	cm²
				2_HYPO_R	Area of Hypothalamus_Right	cm²
			2_Arc	2_Arc	Area of Arcuate nucleus	cm²
		21	2_HP	2_HP	Area of Hippocampus	cm²
			2_TILpy	2_TILpy	Total Internal Length of pyramidal layer	cm
			2_DG	2_DG_L	Length of Dentate Gyrus_Left	cm
				2_DG_R	Length of Dentate Gyrus_Right	cm
			2_Mol	2_Mol_L	Length of Molecular layer_Left	cm
				2_Mol_R	Length of Molecular layer_Right	cm
			2_Rad	2_Rad_L	Length of Radial layer_Left	cm
				2_Rad_R	Length of Radial layer_Right	cm
			2_Or	2_Or_L	Length of Oriens layer_Left	cm
				2_Or_R	Length of Oriens layer_Right	cm
	Cortex	15	2_Pir	2_Pir_L	Area of Piriform cortex_Left	cm²
				2_Pir_R	Area of Piriform cortex_Right	cm²
		22	2_M1	2_M1_L	Height of primary Motor cortex_Left	cm
				2_M1_R	Height of primary Motor cortex_Right	cm
		23	2_S2	2_S2_L	Height of secondary Somatosensory cortex_Left	cm
				2_S2_R	Height of secondary Somatosensory cortex_Right	cm

Table 5 MVP-positive brain structures and link to limbic system

MVP-positive regions identified with immunohistofluorescence on wild-type brain sections are indicated in columns A-B and their known function in column C. (**Figure 40**, **Figure 46** and **Table 6**). The link with the limbic system (defined as Established, Putative or None) are indicated in column D. Direct connections between the different MVP-positive structures are summarized in column E. The last column (ref. for Reference) gives the literature sources used to elaborate this table.

A	B	C	D	E	F
Acronym	**Description**	**Known function**	**Link to limbic system**	**Known connection between MVP-positive structures**	**Ref**
7N	Facial nerve nucleus	Facial expression - Salivation	Putative		
10N	Vagus nerve nucleus	Autonomic modulation	Established	Sol	183
aca	anterior commissure - anterior part	Connection between amygdalae, and olfactory system	Established	Pir	
AcNuc	Acumbens Nucleus	Processing of rewarding and reinforcing - Regulation of slow wave sleep	Established	f, CPu	
AP	Area Postrema	Blood chemoreception - Autonomic control	None	Sol	
Arc	Arcuate hypothalamic nucleus	Autonomic control - Homeostasis	None	PVH, Sol, HP, RSGc	
AV	Arbor vitae	Cerebellum white matter	None		
cc	Corpus callosum	Connection of the cerebral cortex	Putative	Cg, S2, RSGc	
Cg	Cingulate gyrus	Emotion processing - Learning - Memory	Established	ZI, RM, Raphe, cc, DMT, CMT, MS, DBB	184,185
CMT	Central medial thalamic nucleus	Attention modulation	Putative	CPu, MVe, Cg, S2	
CN	Cuneate nucleus	Upper body sensitivity	None		
CP	Choroid plexus	Cerebrospinal fluid production	None		
CPu	Caudate-Putamen	Goal directed actions - Motor learning - slow wave sleep - attachment behaviors	Putative	AcNuc, CMT	
DBB	Diagonal band of Broca	connects with DG in HP, theta waves with MS	Established	PVH, f, MS, ZI, HP, Cg	186
Dk	Darkschewitsch nucleus	Pupillary light reflex	None	ZI, PAG	187
DMH	Dorsal medial hypothalamic nucleus	Circadian rhythm - Feeding behavior - Autonomous regulation	None	LPO, PVH	
DMT	Dorsal medial thalamic nucleus	Attention modulation - Memory - Emotional pain response - Oculomotor	Established	S2, Cg, HP	
f	Fornix	Memory - Theta rhythm	Established	RM, MS, AcNuc, DBB	
Lith	Lithoid nucleus		None	PAG	
LPO	Lateral preoptic area	Modulation of sleep and thirst - Reward-related behaviors	Established	ZI, Raphe, DMH	188
MS	Medial septal nucleus	Theta waves generation	Established	PHV, Cg	186
MVe	Medial vestibular nucleus	Balance - Vestibular integration	Putative		189
PAG	Periaqueductal gray	Pain modulation	Established	Dk, Lith	
Pir	Piriform cortex	Olfaction	Putative	aca	
PVH	Paraventricular hypothalamic nucleus	Autonomic control - Appetite	Established	SFO, Arc, MS, DBB, Sol, ZI, DMH	190
Raphe	Raphe	Pain modulation - Thermoregulation - Emotion regulation - Sleep/wake state	Established	Sol, Cg, LPO, TS	191
RM	Retromammillary nucleus	Memory - Theta rhythm	Established	f, HP, Cg	
RSGc	Retrosplenial cortex	Episodic memory - navigation - slow wave theta rhythmicity	Established	cc, Arc, HP,	192
S2	Somatosensory cortex	Integration of somesthesy (light touch, visceral sensation)	Putative	cc, DMT, CMT	
Sb	Subiculum	Memory - Addiction	Putative	HP	
SFO	Subfornical organ	Energy homeostasis - Cardiovascular modulation	None	PVH	
Sol	Solitary nucleus	Autonomic modulation	Established	AP, PVH, Raphe, 10N	183
TS	Triangular septal nucleus	Control of anxiety and fear	Established	Raphe	193,194
vHP	Hippocampus - ventral part	Memory - Large scale spatial navigation	Established	Arc, DBB, RM, DMT, RSGc, DS	195
Xi	Xiphoid thalamic nucleus	Visual threats adaptive response	Putative	Amygdala	196
ZI	Zona incerta	Autonomic control - Slow-wave sleep - Attention	Established	Cg, LPO, PVH, DBB, Dk	197,198

Table 6 Expression scores of MVP in murine brain

Summary of MVP expression obtained from immunohistofluorescence on serial coronal (from Bregma +1.94mm to Bregma -7.67mm) and sagittal (Lat +0.24mm to Lat +0.72mm) sections. The stereotaxic coordinates are indicated in the first column along with the corresponding plane number of the Franklin and Paxinos Atlas[199]. Each brain part is indicated at the top of each table, and more precision about the identity of each structure is indicated in bracket. Abbreviations of MVP-positive structures are given in **Table 5.** MVP/vault expression is scored as mild (+), moderate (++) or strong (+++) across all the positive brain regions. ¶ indicate the structure where only fibers are studied. NA: not applicable. NV: not visible. Cells highlighted in pale blue correspond to illustrations presented in **Figure 40** and **Figure 46**.

Coronal - Bloc 1													
Stereotaxic coordinates (mm)	Plane on Franklin and Paxinos Atlas	Cingulate gyrus	Corpus callosum ¶	Striatum	Striatum ¶	Anterior commissure anterior part ¶	Piriform cortex	Somatosensory cortex	Bands of Brocca	Septum	Hypothalamus	Choroid plexus	Fornix
Bregma +1.94	15	NA	NA	NA	NA	++	++	NA	NA	NA	NA	NA	NA
Bregma +1.41	19	++	NA	+(AcNuc) +(CPu)	++	++	+++	+	++	NA	NA	NA	NA
Bregma +0.85	24	+	+++	+	++	++	+++	+	+++	+++ (MS)	NA	+++	NA
Bregma +0.50	27	+	+++	+	++	++	+++	+	+++	+++ (MS)	NA	+++	NA
Bregma +0.14	30	+	+++	+	++	++	+++	+	++	NA	++ (LPO)	+++	NA
Bregma +0.01	31	+	+++	+	++	++	+++	+	++	NA	+	+++	+++

Coronal - Bloc 2															
Stereotaxic coordinates (mm)	Plane on Franklin and Paxinos Atlas	Retrosplenial cortex	Body of the corpus callosum ¶	Piriform cortex	Amygdala	Dorsal hippocampus	Ventral hippocampus	Somatosensory cortex	Choroid plexus	Septum	Thalamus	Midbrain	Fornix	Hypothalamus	Other
Bregma -0.46	35	+	NA	+++	NA	NA	NA	+	+++	+++ (TS)	NA	NA	+++	++ (SFO) +++ (PVH)	NA
Bregma -0.58	36	+	NA	+++	NA	NA	NA	+	+++	+++ (TS)	++ (Xi)	NA	NA	+++ (SFO)	NA
Bregma -0.70	37	+	++	+++	NA	NA	NA	+	+++	NA	NA	NA	NA	NA	NA
Bregma -1.07	40	+	++	+++	NA	NA	NA	+	+++	NA	NA	NA	NA	NA	NA
Bregma -1.34	42	+	NA	+++	+	+	NA	+	+++	NA	++ (CMT) ++ (DMT)	NA	NA	+++ (Arc) +++ (ZI)	NA
Bregma -1.46	43	+	++	+++	-	-	NA	-	+++	NA	NA	NA	NA	+++ (DMH) +++ (ZI)	NA
Bregma -1.55	44	+	++	+++	-	-	NA	+	+++	NA	NA	NA	NA	+++ (Arc) +++ (DMH) +++ (ZI)	NA
Bregma - 1.67	45	NA	NA	+++	-	-	NA	+	+++	NA	NA	NA	NA	+++ (DMH) +++ (ZI)	NA
Bregma -2.15	49	+	NA	+++	-	NA	NA	+	+++	NA	NA	NA	NA	++ (ZI)+++	NA
Bregma -2.27	50	NA	NA	NA	NA	-	NA	-	+++	NA	NA	NA	NA	++ (ZI)++	NA
Bregma -2.53	52	+	++	+++	NA	-	+	+	+++	NA	NA	NA	NA	++ (ZI) +++	NA
Bregma -2.69	53	+	++	+++	NA	+	++	+	+++	NA	NA	++ PAG, Lith, Dk	NA	+ (ZI) +++ (RM)	NA
Bregma -2.79	54	+	++	+++	NA	NA	++	+	+++	NA	NA	++ PAG, Lith, Dk	NA	+ (ZI) +++ (RM)	NA
Bregma -2.91	55	+	++	NA	NA	+	++	+	+++	NA	NA	++ PAG, Lith, Dk	NA	+++ (RM)	++ (Sb)

Coronal - Bloc 3								
Stereotaxic coordinates (mm)	Plane on Franklin and Paxinos Atlas	Lobules	Arbor vitae ¶	Raphe	Nerve nucleus	Reticular area	Vestibular nucleus	Other
Bregma -5.80	79	++	++	++	NA	++	++	NA
Bregma -5.88	80	++	++	++	++ (7N)	-	++	NA
Bregma -5.99	81	++	++	++	++ (7N)	+	++	NA
Bregma -6.47	85	++	++	+	++(Sol)	+	+++ (Mve)	+++ (CP)
Bregma -6.83	88	NA	NA	+	++ (Sol)	-	+	NA
Bregma -7.43	93	NA	NA	NA	+++ (10N)	+	NA	+ (CN)
Bregma -7.67	96	NA	NA	NA	+++ (10N) +++(Sol)	+	++	++ (AP)

Sagittal									
Stereotaxic coordinates (mm)	Plane on Franklin and Paxinos Atlas	Cingulate Gyrus	Corpus callosum ¶	Hippocampus	Cortex	Striatum	Hypothalamus	Cerebellum	Medulla/pons
Lat +0.24	103	-	-	-	-	+	++	+++	+++
Lat +0.60	106	+	-	-	++	-	++	++	+++
Lat +0.72	107	-	-	-	-	+	++	+++	+++

Table 7 List of 42 coronal neuroanatomical measurements in E18.5 embryos.

This table enlists morphological phenotypes, adapted from the postnatal neuroanatomical screen, to fit to the analysis of brain anatomy in mouse embryo at embryonic stage 18.5 (E18.5). Stereotaxic coordinates (column A) of the two sections of interest are indicated. Column B gives the merged name of the parameter (used in **Dataset 2**), column C the full name, column D the description and column E the unit of the measurement. An illustrative representation of the measurements are available in **Figure 56**.

A	B	C	D	E
Section of interest	**Acronym**	**Full Acronym**	**Description**	**Unit**
Critical section 5 (2.19mm)	5_TBA	5_TBA	Total Brain Area	cm²
	5_LV	5_LV_L	Area of Lateral Ventricle_Left	cm²
		5_LV_R	Area of Lateral Ventricle_Right	cm²
	5_Cg	5_Cg_L	Area of Cingulate cortex_Left	cm²
		5_Cg_R	Area of Cingulate cortex_Right	cm²
	5_Cg_Width	5_Cg_Width_L	Width of Cingulate cortex_Left	cm
		5_Cg_Width_R	Width of Cingulate cortex_Right	cm
	5_Cg_Height	5_Cg_Height	Height of Cingulate cortex	cm
	5_gcc	5_gcc	Area of genu of corpus callosum	cm²
	5_gcc_Width_T	5_gcc_Width_T	Width of genu of corpus callosum_Top	cm
	5_gcc_Height	5_gcc_Height	Height of genu of corpus callosum	cm
	5_CPu	5_CPu_L	Area of Caudate Putamen_Left	cm²
		5_CPu_R	Area of Caudate Putamen_Right	cm²
	5_aca	5_aca_L	Area of anterior commissure-anterior part_Left	cm²
		5_aca_R	Area of anterior commissure-anterior part_Right	cm²
	5_M1	5_M1_L	Height of Motor cortex_Left	cm
		5_M1_R	Height of Motor cortex_Right	cm
	5_S2	5_S2_L	Height of secondary Somatosensory cortex_Left	cm
		5_S2_R	Height of secondary Somatosensory cortex_Right	cm
Critical Section 6 (3.51mm)	6_TBA	6_TBA	Total Brain Area	cm²
	6_LV	6_LV_L	Area of Lateral Ventricle_Left	cm²
		6_LV_R	Area of Lateral Ventricle_Right	cm²
	6_D3V	6_D3V	Area of Dorsal third Ventricle	cm²
	6_3V	6_3V	Area of third Ventricle	cm²
	6_HP	6_HP_L	Area of Hippocampus_Left	cm²
		6_HP_R	Area of Hippocampus_Right	cm²
	6_TlLpy	6_TlLpy_L	Total Internal Length of pyramidal layer_Left	cm
		6_TlLpy_R	Total Internal Length of pyramidal layer_Right	cm
	6_DG	6_DG_L	Length of Dentate Gyrus_Left	cm
		6_DG_R	Length of Dentate Gyrus_Right	cm
	6_Mol	6_Mol_L	Length of Molecular layer_Left	cm
		6_Mol_R	Length of Molecular layer_Right	cm
	6_Rad	6_Rad_L	Length of Radial layer_left	cm
		6_Rad_R	Length of Radial layer_Right	cm
	6_Or	6_Or_L	Length of Oriens layer_Left	cm
		6_Or_R	Length of Oriens layer_Right	cm
	6_RSGc	6_RSGc_L	Area of Retrosplenial Granular cortex_Left	cm²
		6_RSGc_R	Area of Retrosplenial Granular cortex_Right	cm²
	6_M1	6_M1_L	Height of primay Motor cortex_Left	cm
		6_M1_R	Height of primary Motor cortex_Right	cm
	6_S2	6_S2_L	Height of secondary Somatosensory cortex_Left	cm
		6_S2_R	Height of secondary Somatosensory cortex_Right	cm

Table 8 List of 40 sagittal neuroanatomical measurements in postnatal mice.

This table enlists combined morphological phenotypes studied on sagittal plane. Stereotaxic coordinates (column A) of the section of interest are indicated. Column B shows the number attributed to each region (**Figure 49)**. Column C gives the full name, column D the description and column E the unit of the measurement.

A	B	C	D	E
Section of interest	**ID**	**Full Acronym**	**Description**	**Unit**
Lateral +0.60mm	**1**	4_TB_area	Total Brain Area	cm²
		4_TB_width	Width of the Total Brain	cm
		4_TB_height_CS1	Height of Total Brain at Bregma +0.98mm	cm
		4_TB_height_CS2	Height of Total Brain at Bregma -1.34mm	cm
	2	4_TCTX_area	Area of the secondary motor cortex	cm²
		4_M2_length	Height of secondary Motor cortex	cm
		4_M1_length	Height of primary Motor cortex	cm
	3	4_Pons_height	Height of the Pons	cm
	4	4_TC_area	Total Cerebellar area	cm²
		4_IGL_area	Area of Internal Granular Layer of cerebellum	cm²
		4_Folia	Number of Folia	number
		4_Med_area	Area of Medial cerebellar nucleus	cm²
	5	4_LV_area	Area of Lateral Ventricle	cm²
	6	4_cc_area	Area of corpus callosum	cm²
		4_cc_length	Total outer length of corpus callosum	cm
		4_cc_height	Height of corpus callosum	cm
	7	4_TTh_area	Area of the Thalamus	cm²
	8	4_CPu_area	Area of Caudate Putamen	cm²
	9	4_HP_area	Area of Hippocampus	cm²
		4_Rad_length	Height of Radiatum layer of hippocampus	cm
		4_Or_length	Height of Orien layer of hippocampus	cm
		4_TILpy_area	Area of Pyramidal cell layer of hippocampus	cm²
		4_TILpy_length	Total Internal Length of Pyramidal cell layer of hippocampus	cm
		4_Mol_length	Height of Molecular layer of hippocampus	cm
		4_DG_area	Area of the Dentate Gyrus	cm²
		4_DG_length	Length of Dentate Gyrus	cm
	10	4_fi_area	Area of fimbria of hippocampus	cm²
	11	4_aca_area	Area of anterior commissure	cm²
	12	4_sm_area	Area of stria medullaris	cm²
	13	4_f_area	Area of fornix	cm²
	14	4_och_area	Area of optic chiasm	cm²
	15	4_VMHvl_area	Area of the Ventromedial nucleus of the Hypothalamus	cm²
	16	4_Pn_area	Area of Pontine nuclei	cm²
	17	4_SN_area	Area of Substantia Nigra	cm²
	18	4_fp_area	Area of fasciculus of pons	cm²
	19	4_Cg_area	Area of Cingulate gyrus	cm²
		4_Cg_height	Height of Cingulate gyrus	cm
	20	4_DS_area	Area of Dorsal Subiculum	cm²
	21	4_InfC_area	Area of Inferior Colliculus	cm²
	22	4_SupC_area	Area of Superior Colliculus	cm²

4.7 Dataset legends

Dataset 1 Full neuroanatomical data.

A total of 458 samples were processed and analyzed across two coronal sections. A unique barcode (column A) identifies each sample. Column B and C refer to each subproject that were used to create subgroups for the analysis. Columns D to N give additional elements about study samples. Columns O to CC are the raw data for the 67 brain parameters (see **Table 4** for more information about the parameters), columns CD and CE give body and brain weight, and columns CF to DO provide the co-variates for each sample.

Dataset 2 Description of gene association with neuroanatomical phenotypes.

Sheet 1 HeatMap_SingleGenes_Coronal.
Columns A-I provide general information about each group of mice analyzed in the histology study. More specifically, column A gives the gene name, column B the core background strain, column C the full allele name, column D the age of the mice at necropsy in weeks, column E the gender of studied mice, column F the zygosity, column G the number of mutant mice and column I the number of colony-matched control animals (column H). Column J indicates the significance threshold of the NAP gene defined as NAT (NeuroAnatomical Threshold), column K the directionality of the impact on gene structure, column L the number of neuroanatomical phenotypes at the defined NAT of 0.05, column M the type of the statistics and finally columns N-BC the p-value and percentage change for each brain parameter across each tested allele. Column BD and BE indicate p-value and percentage change for body and brain weight, respectively. Column BF to BK give the number of NAPs according to each NAT. Column BL to BR give additional descriptive statistics. A color-code is used to indicate p-value threshold, ranging from yellow (for 0.05) to red for stringent threshold (of 0.0001) and directionality of percentage change (blue indicates decrease and red increase).
Sheet 2 HeatMap_Mvp_Dev.
Column A to BC and color-code legend are the same as **Sheet 1**. Column BD to BI give the number of NAPs according to each NAT.
Sheet 3 HeatMap_Mvp_Sagittal.
Column A to M and color-code legend are the same as **Sheet 1**. Columns N-BA indicate the p-value and percentage change for each brain parameter across each tested allele. Column BB to BG give the number of NAPs according to each NAT.

Dataset 3 Raw data for the various *Mvp* studies.

Sheet 1 Viability: Calculation of birth ratio for each genotype, and each sex, done on successfully genotyped mice obtained from heterozygous by heterozygous breeding (n=617).
Sheet 2 WesternBlot: Measurements of MVP and Actin levels in liver samples from *Mvp* mutant mice.
Sheet 3 Dev_E18.5: Raw data from male and female $Mvp^{+/+}$ and $Mvp^{-/-}$ E18.5 embryos. See embryos. Extended for description of the parameters.
Sheet 4 Histocount_Coro_M: Cellular parameters for male *Mvp* mutants. Cellular measures done on histological coronal sections (See **Table 4** for corresponding parameter description) are presented in the table. For 1_M1, 1_S2 and 2_S2 analysis of area and cellular parameters was further explored in the different cortical layers (see **Figure 49A** and **Material and Methods** for more explanation). Columns highlighted in purple designate the data obtained by calculation: avgcellarea=cellarea/cellcount; cellareapercent=cellarea/structure_area; celldens=cellcount/structure_area.
Sheet 5 Histocount_Coro_F: Cellular parameters for female *Mvp* mutants. Measures done on histological coronal sections on specific regions (See **Table 4** for corresponding parameter description) are presented in the table. Columns highlighted in purple designate the data obtained by calculation: avgcellarea=cellarea/cellcount; cellareapercent=cellarea/structure_area; celldens=cellcount/structure_area.
Sheet 6 Histocount_Sag_M: Cellular parameters for male *Mvp* mutants. Measures done on histological sagittal sections on motor cortex (4_M1) and cerebellar lobes (Cb) are presented in the table. For Cb, the length of granular layer for each lobule (GLlength, in cm) and the number of Purkinje cells (PC) was measured (See **Figure 49M** for further explanation). Columns highlighted in purple designate the data obtained by calculation: avgcellarea=cellarea/cellcount; cellareapercent=cellarea/M1_area; celldens=cellcount/M1_area for motor cortex and PCdensity=PC/GL_length. P-value of Student *t*-test and percentage change of each interaction and the mean and SEM for each genotype are indicative.
Sheets 7-13 Immunocytofluorescence (ICF): these tabs recapitulates the data obtained by the measurements of cellular parameters done on primary hippocampal neuronal cultures at DIV4 from E18.5 male and female embryos. Soma area (soma), growth cone (GC) area and axonal length (AL) are presented (see **Figure 48I** for visual description). The raw measure for each neuron is given in column D (area (in μm^2) for soma and GC, and length (in µm) for AL). The decimal logarithm (Log10) of each measure is given in column E and is used to conduct statistical tests. For male soma, two different batches were tested, each reported in a different tab.
Sheet 14 Golgi_SpineDensity: Spine density in somatosensory neurons. Column A and B inform on the number and the genotype of the mouse. The length measured on analyzed portion of dendrite, as well as the number of spines counted are provided in column C and D. Spine density (E) corresponds to : SpineDensity = #spines/length.
Sheet 15. Electrophysiology. Patch-clamp data in cingulate gyrus. Inter-event interval (IEI, in ms) and amplitude (in pA) of mEPSCs are recorded. The decimal logarithm (Log10) of each measure is given and was used to conduct statistical tests. The capacitance of each neuron is measured.

Dataset 4 Raw behavioral data for *Mvp*$^{+/-}$::*Mapk3*$^{+/-}$

Raw data sheets for the analysis of *Mvp::Mapk3* double-heterozygotes (n=12) versus WT (n=12) mice. See **5.14 Behavioral Analysis** for more detailed description of test procedures.

Sheet 1 Legend
Sheet 2 DblKO_BW. Body weight longitudinal study. Weight were taken every week between 11 and 26 weeks old.
Sheet 3 DblKO_OF. Open Field raw data. Measured parameters: Distance Traveled (cm), Number of Rearing's, % Time Spent in the Center, Distance (cm) in Center, Latency (s) in Center and % Distance in Center. Data are calculated over five minutes periods along the test (max 30min).
Sheet 4 DblKO_EPM. Elevated Plus Maze. Measured parameters: number of entries and time spent in closed or open arms, number of head dip, rearing and extension, and latency (s) before entering open arm.
Sheet 5 DblKO_FC_Cond. Fear Conditioning – Conditioning session. Immobility duration measured over three consecutive periods of two minutes (Hab-1, Hab-2 and Post-US1). Electrical foot shock (US), preceded by light and tone stimuli is given right before Post-US1.
Sheet 6 DblKO_FC_Cont. Fear Conditioning – Context session. Duration of immobility (s) of the mouse during three consecutive periods of two minutes. Percentage of freezing over the total time (360s) is calculated and indicated.
Sheet 7 DblKO_FC_Cue. Fear Conditioning – Cue session. Immobility (s) of the mouse measured over four consecutive periods of two minutes (Pre-Cue1, Cue1, Pre-Cue2 and Cue2). Percentage of freezing during Cue1 and Cue2 is calculated on the duration of the session.
Sheet 8 DblKO_TST. Tail Suspension Test. Immobility (s) of the mouse measured over three consecutive periods of two minutes (Block1, Block2 and Block3) after suspension by the tail. Latency (s) to immobility is also recorded.
Sheet 9 DblKO_FS. Forced Swim test. Immobility (s) of the mouse measured over three consecutive periods of two minutes (Trial1, Trial2 and Trial3) after immersion in a becher of water. Latency (s) to immobility is also recorded.
Sheet 10 DblKO_PTZ. PTZ-susceptibility test. Latency and duration of the clonic and tonic phases of status epilepticus after PTZ intraperitoneal injection.
Sheet 11 DblKO_GS. Grip Strength. Grip strength values for four or two paws in kg and g. Each condition is repeated 4 times. The mean and the maximal value (highlighted in orange) are calculated and normalized over the mouse body weight.
Sheet 12 DblKO_Rotarod_Learning. Rotarod test – Learning session. Time (in seconds) spent on the rotarod at increasing speed from 4 to 40 rotations per minute (rpm) over five minutes session, during four sessions on three consecutive days.
Sheet 13 DblKO_Rotarod_Test. Rotarod test – Test session. Latency before falling off the rotarod at rotation speeds (4, 10, 16, 22, 28, 34 and 40 rpm). Maximum per trial is set to 120 seconds. Two trials were done and the mean of both was calculated and gathered in the bottom panel.
Sheet 14 DblKO_SP_Data. Sucrose preference test – Test data. Weight of water and 0.8% sucrose bottles before and after each session. Consumption corresponds to the difference between the two weights. Sucrose preference was calculated as the

percentage of sucrose consumption over the total consumption during habituation (Hab, 1h), and testing days (Test-Day1 and Test–Day2, 12 hours during dark cycle).
Sheet 15 DblKO_SP_Consumption. Sucrose preference test – Consumption data. Beverage consumption (addition of sucrose and water consumption) for the three sessions.
Sheet 16 DblKO_YM. Y-Maze test. Poss (Poss=number of total visited arm - 2) is used as an alias of locomotor activity, Real (Real: three consecutive different arms visited) and percentage of alternation (%Alt) as an alias for working memory. %Alt = Real/Poss*100. Same Arm Return (SAR) and Alternance Arm Return (AAR) were recorded and percentage calculated. Test was conducted on a single trial and data over either 6 or 8 minutes were transcribed separately.
Sheet 17 DblKO_NOR. Novel-Object recognition test. Exploration time (s) of the different objects (ObjectA in 1st session, and ObjectFam and ObjectNew, in 2nd session). Discrimination ratio = (ObjectNew – ObjectFam)/ (ObjectNew+ObjectFam). Percentage of exploration of the new object over total object exploration %Tnew=ObjectNew/(ObjectNew+ObjectFam). 50% was considered as no recognition of the familiar object.
Sheet 18-20 DblKO_SR_Phase1/2/3. Social Recognition – Phase 1/2/3. Total distance moved (cm) during each phase of SR test. Cumulative duration of the mouse localization in each of the three-chamber (Left, Center or Right) either Zone1 (the entire chamber), Zone2 (closer to the goal box).
Sheet 21 DblKO_SI. Social interaction test. Recordings of the number and time spent to do different social behaviors: Agressivity, Pawing contact, Individual contact, following and sniffing over the ten minutes of the test.
Sheet 22 DblKO_Marbles. Marbles burying test. Number of marbles that were ¾ covered, completely covered or not covered after 15minutes of test period, and their relative percentage.
Sheet 23 DblKO_CA_All. Tracking of locomotor activity (in centimeters) every hour over the duration of the test (32 hours). Statistical analysis was performed over the combined Dark and Light phase (from 7pm to 7pm). Back and front activity, as well as rearing, licks and pellet distribution are also recorded.
Sheet 24 DblKO_CA_food consumption. Circadian consumption. Weight of pellets and water in grams consumed over 32 hours. The difference between pellet distribution and pellet lost is considered as pellet distribution. The weight of water bottles before and after the trial was also measured to quantify water consumption.
Sheet 25 DblKO_PPI_Results Avg. Pre-Pulse Inhibition test – Average value. Averaged amplitude of the acoustic startle reflex (arbitrary unit) following 60 dB (Background Noise = BN), 70 dB pulse (P70), 75 dB (P75), 85 dB (P85), 90 dB (P90) and 110 dB (ST110) acoustic stimulus. PPI: Percentage of Pre-Pulse (PP) Inhibition calculated as PPX %=(PPX-ST110)/ST110.
Sheet 26 DblKO_PPI_Results Max. Pre-Pulse Inhibition test – Max value. Maximum amplitude of the acoustic startle reflex (arbitrary unit) following 60 dB (Background Noise = BN), 70 dB pulse (P70), 75 dB (P75), 85 dB (P85), 90 dB (P90) and 110 dB (ST110) acoustic stimulus. PPI: Percentage of Pre-Pulse (PP) Inhibition calculated as PPX %=(PPX-ST110)/ST110.

Dataset 5 Raw behavioral data for *Mvp*$^{-/-}$.

Raw data sheets for the analysis of *Mvp*$^{-/-}$ mutant mice (n=12) versus WT (n=10) mice. See **5.14 Behavioral Analysis** for more detailed description of test procedures.

Sheet 1 Legend.
Sheet 2 Mvp_OF. Open Field raw data. Measured parameters: Distance Traveled (cm), Number of rearing, % Time Spent in the Center, Distance (cm) in Center, Latency (s) in Center and % Distance in Center. Data are calculated over five minutes periods along the test (max 30min).
Sheet 3 Mvp_EPM. Elevated Plus Maze. Measured parameters: number of entries and time spent in closed or open arms, number of head dip, rearing and extension, and latency (s) before entering open arm.
Sheet 4 Mvp_FC_Cond. Fear Conditioning – Conditioning session. Immobility duration measured over three consecutive periods of two minutes (Hab-1, Hab-2 and Post-US1). Electrical foot shock, preceded by light and tone stimuli is given right before Post-US1.
Sheet 5 Mvp_FC_Context. Fear Conditioning – Context session. Duration of immobility (s) of the mouse during three consecutive periods of two minutes. Percentage of freezing over the total time (360s) is calculated and indicated.
Sheet 6 Mvp_FC_Cue. Fear Conditioning – Cue session. Immobility (s) of the mouse measured over four consecutive periods of two minutes (Pre-Cue1, Cue1, Pre-Cue2 and Cue2). Percentage of freezing during Cue1 and Cue2 is calculated on the duration of the session.
Sheet 7 Mvp_TST. Tail Suspension Test. Immobility (s) of the mouse measured over three consecutive periods of two minutes (Block1, Block2 and Block3) after suspension by the tail. Latency (s) to immobility is also recorded.
Sheet 8 Mvp_PTZ. PTZ-susceptibility test. Latency and duration of the clonic and tonic phases of status epilepticus after PTZ intraperitoneal injection.
Sheet 9 Mvp_Rotarod_Learning. Rotarod test – Learning session. Time (in seconds) spent on the rotarod at increasing speed from 4 to 40 rotations per minute (rpm) over five minutes session, during four sessions on three consecutive days.
Sheet 10 Mvp_Rotarod_Test. Rotarod test – Test session. Latency before falling off the rotarod at rotation speeds (4, 10, 16, 22, 28, 34 and 40 rpm). Maximum per trial is set to 120 seconds. Two trials were done and the mean of both was calculated and gathered in the bottom panel.
Sheet 11 Mvp_SP_Data. Sucrose preference test – Test data. Weight of water and 0.8% sucrose bottles before and after each session. Consumption corresponds to the difference between the two weights. Sucrose preference was calculated as the percentage of sucrose consumption over the total consumption during habituation (Hab, 1h), and testing days (Test-Day1 and Test–Day2, 12 hours during dark cycle).
Sheet 12 Mvp_SP_Consumption. Sucrose preference test – Consumption data. Beverage consumption (addition of sucrose and water consumption) for the three sessions.
Sheet 13 Mvp_YM. . Y-Maze test. Poss (Poss=number of total visited arm - 2) is used as an alias of locomotor activity, Real (Real: three consecutive different arms visited) and percentage of alternation (%Alt) as an alias for working memory. %Alt = Real/Poss*100. Same Arm Return (SAR) and Alternance Arm Return (AAR) were recorded and percentage calculated. Test was conducted on a single trial and data over either 6 or 8 minutes were transcribed separately.

Sheet 14 Mvp_NOR. Novel-Object recognition test. Exploration time (s) of the different objects (ObjectA in 1st session, and ObjectFam and ObjectNew, in 2nd session). Discrimination ratio = (ObjectNew – ObjectFam)/(ObjectNew+ObjectFam). Percentage of exploration of the new object over total object exploration %Tnew=ObjectNew/(ObjectNew+ObjectFam). 50% was considered as no recognition of the familiar object.

Sheet 15-17 Mvp_SR_Phase1/2/3. Social Recognition – Phase 1/2/3. Total distance moved (cm) during each phase of SR test. Cumulative duration of the mouse localization in each of the three-chamber (Left, Center or Right) either Zone1 (the entire chamber), Zone2 (closer to the goal box).

Sheet 18 Mvp_Marbles. Marbles burying test. Number of marbles that were ¾ covered, completely covered or not covered after 15minutes of test period, and their relative percentage.

Sheet 19 Mvp_CA_All. Tracking of locomotor activity (in centimeters) every hour over the duration of the test (32 hours). Statistical analysis was performed over the combined Dark and Light phase (from 7pm to 7pm). Back and front activity, as well as rearing, licks and pellet distribution are also recorded.

Sheet 20 Mvp_CA_food consumption. Circadian consumption. Weight of pellets and water in grams consumed over 32 hours. The difference between pellet distribution and pellet lost is considered as pellet distribution. The weight of water bottles before and after the trial was also measured to quantify water consumption.

Sheet 21 Mvp_PPI_Avg. Pre-Pulse Inhibition test – Average value. Averaged amplitude of the acoustic startle reflex (arbitrary unit) following 60 dB (Background Noise = BN), 70 dB pulse (P70), 75 dB (P75), 85 dB (P85), 90 dB (P90) and 110 dB (ST110) acoustic stimulus. PPI: Percentage of Pre-Pulse (PP) Inhibition calculated as PPX %=(PPX-ST110)/ST110.

Sheet 22 Mvp_PPI_Max. Pre-Pulse Inhibition test – Max value. Maximum amplitude of the acoustic startle reflex (arbitrary unit) following 60 dB (Background Noise = BN), 70 dB pulse (P70), 75 dB (P75), 85 dB (P85), 90 dB (P90) and 110 dB (ST110) acoustic stimulus. PPI: Percentage of Pre-Pulse (PP) Inhibition calculated as PPX %=(PPX-ST110)/ST110.

Dataset 6 Raw behavioral data for *Mapk3*$^{+/-}$

Raw data sheets for the analysis of *Mapk3*$^{+/-}$ mutant mice (n=12) versus WT (n=12) mice in two different cohorts. See **5.14 Behavioral Analysis** for more detailed description of test procedures.

Sheet 1 Legend
Sheet 2 Mapk3_C1_OF. Open Field raw data. Measured parameters: Distance Traveled (cm), Number of rearing, % Time Spent in the Center, Distance (cm) in Center, Latency (s) in Center and % Distance in Center. Data are calculated over five minutes periods along the test (max 30min).
Sheet 3 Mapk3_C1_EPM. Elevated Plus Maze. Measured parameters: number of entries and time spent in closed or open arms, number of head dip, rearing and extension, and latency (s) before entering open arm.
Sheet 4 Mapk3_C1_FC_Cond. Fear Conditioning – Conditioning session. Immobility duration measured over three consecutive periods of two minutes (Hab-1, Hab-2 and Post-US1). Electrical foot shock, preceded by light and tone stimuli is given right before Post-US1.
Sheet 5 Mapk3_C1_FC_Context. Fear Conditioning – Context session. Duration of immobility (s) of the mouse during three consecutive periods of two minutes. Percentage of freezing over the total time (360s) is calculated and indicated.
Sheet 6 Mapk3_C1_FC_Cue. Fear Conditioning – Cue session. Immobility (s) of the mouse measured over four consecutive periods of two minutes (Pre-Cue1, Cue1, Pre-Cue2 and Cue2). Percentage of freezing during Cue1 and Cue2 is calculated on the duration of the session.
Sheet 7-8 Mapk3_C1/C2_TST. Tail Suspension Test for cohort 1/2. Immobility (s) of the mouse measured over three consecutive periods of two minutes (Block1, Block2 and Block3) after suspension by the tail. Latency (s) to immobility is also recorded.
Sheet 9 Mapk3_C2_PTZ. PTZ-susceptibility test. Latency and duration of the clonic and tonic phases of status epilepticus after PTZ intraperitoneal injection.
Sheet 10 Mapk3_C2_Rotarod_Learning. Rotarod test – Learning session. Time (in seconds) spent on the rotarod at increasing speed from 4 to 40 rotations per minute (rpm) over five minutes session, during four sessions on three consecutive days.
Sheet 11 Mapk3_C2_Rotarod_Test. Rotarod test – Test session. Latency before falling off the rotarod at rotation speeds (4, 10, 16, 22, 28, 34 and 40 rpm). Maximum per trial is set to 120 seconds. Two trials were done and the mean of both was calculated and gathered in the bottom panel.
Sheet 12 Mapk3_C2_SP_Data. Sucrose preference test – Test data. Weight of water and 0.8% sucrose bottles before and after each session. Consumption corresponds to the difference between the two weights. Sucrose preference was calculated as the percentage of sucrose consumption over the total consumption during habituation (Hab, 1h), and testing days (Test-Day1 and Test–Day2, 12 hours during dark cycle).
Sheet 13 Mapk3_C2_SP_Consumption. Sucrose preference test – Consumption data. Beverage consumption (addition of sucrose and water consumption) for the three sessions.
Sheet 14 Mapk3_C1_YM. Y-Maze test. Poss (Poss=number of total visited arm - 2) is used as an alias of locomotor activity, Real (Real: three consecutive different arms visited) and percentage of alternation (%Alt) as an alias for working memory. %Alt = Real/Poss*100. Same Arm Return (SAR) and Alternance Arm Return (AAR) were

recorded and percentage calculated. Test was conducted on a single trial and data over either 6 or 8 minutes were transcribed separately.

Sheet 15 Mapk3_C1_NOR. Novel-Object recognition test. Exploration time (s) of the different objects (ObjectA in 1st session, and ObjectFam and ObjectNew, in 2nd session). Discrimination ratio = (ObjectNew – ObjectFam)/ (ObjectNew+ObjectFam). Percentage of exploration of the new object over total object exploration %Tnew=ObjectNew/ (ObjectNew+ObjectFam). 50% was considered as no recognition of the familiar object.

Sheet 16-18 Mapk3_C1_SR_Phase1/2/3. Social Recognition – Phase 1/2/3. Total distance moved (cm) during each phase of SR test. Cumulative duration of the mouse localization in each of the three-chamber (Left, Center or Right) either Zone1 (the entire chamber), Zone2 (closer to the goal box).

Sheet 19 Mapk3_C1_Marbles. Marbles burying test. Number of marbles that were ¾ covered, completely covered or not covered after 15minutes of test period, and their relative percentage.

Sheet 20 Mapk3_C1_CA_All. Tracking of locomotor activity (in centimeters) every hour over the duration of the test (32 hours). Statistical analysis was performed over the combined Dark and Light phase (from 7pm to 7pm). Back and front activity, as well as rearing, licks and pellet distribution are also recorded.

Sheet 21 Mapk3_C1_CA_Consumption. Circadian consumption. Weight of pellets and water in grams consumed over 32 hours. The difference between pellet distribution and pellet lost is considered as pellet distribution. The weight of water bottles before and after the trial was also measured to quantify water consumption.

Sheet 22 Mapk3_C1_PPI_Avg. Pre-Pulse Inhibition test – Average value. Averaged amplitude of the acoustic startle reflex (arbitrary unit) following 60 dB (Background Noise = BN), 70 dB pulse (P70), 75 dB (P75), 85 dB (P85), 90 dB (P90) and 110 dB (ST110) acoustic stimulus. PPI: Percentage of Pre-Pulse (PP) Inhibition calculated as PPX %=(PPX-ST110)/ST110.

Sheet 23 Mapk3_C1_PPI_Max. Pre-Pulse Inhibition test – Max value. Maximum amplitude of the acoustic startle reflex (arbitrary unit) following 60 dB (Background Noise = BN), 70 dB pulse (P70), 75 dB (P75), 85 dB (P85), 90 dB (P90) and 110 dB (ST110) acoustic stimulus. PPI: Percentage of Pre-Pulse (PP) Inhibition calculated as PPX %=(PPX-ST110)/ST110.

4.8 Supplementary Results

4.8.1 Male NAP genes implicated in two or more brain categories

At the exception of *Mvp* that we described in the main text and **Figure 45**, seven other NAP genes (*Bola2*, *Qprt*, *Maz*, *Ppp4c*, *Slx1b*, *Taok2* and *Zg16*) also gave very significant results affecting two or more brain categories (**Figure 39C**, **Figure 43** and **Dataset 2**). These findings are described in details below.

Male *Bola2*$^{+/-}$ showed four affected brain parameters pertaining to three main brain categories (commissure, subcortical structures and ventricle). The size of the soma of the corpus callosum was reduced (-10%, p=0.0066) concomitant with thin height of the same structure (-11%, p=0.044). The size of the hippocampus was also smaller by 12% (p=0.018). Intriguingly, the size of the third ventricle was smaller by no less than 40% (p=0.0071).

Male *Qprt*$^{+/-}$ showed three affected brain parameters pertaining to two main brain categories (commissure and subcortical). Consistently, the size of the corpus callosum, its width and the size of the internal capsule was smaller by 18% (p=0.011), 15% (p=0.0032) and 14% (p=0.011), respectively. It is worth mentioning that the impact of *Qprt* deletion seems to be specific to white matter structures.

Male *Maz*$^{+/-}$ showed six affected brain parameters pertaining to four main brain categories (all except brain size). Five brain parameters were reduced in size when compared to colony matched WTs including the height of the cingulate cortex (-15%, p=0.010), the lateral ventricles (-37%, p=0.012), the height of the corpus callosum (-17%, p=0.012), the internal capsule (-10%, p=0.044) and the habenular nucleus (-14%, p=0.046). By contrast, one parameter, the somatosensory cortex at Bregma -1.34 mm was increased in size (+8%, p=0.0098).

Male *Ppp4c*$^{+/-}$ associated with ten affected parameters belonging to four main brain categories (all except ventricle). These parameters were all decreased in size and included the total brain area at Bregma +0.98 mm and Bregma -1.34 mm (-10%, p=0.0077 and -7%, p=0.035, respectively), the height of the cingulate cortex (-7%, p=0.0079), the genu of the corpus callosum (-16%, p=0.0088), the width of the genu of the corpus callosum (-9%, p=0.023), the anterior commissure (-12%, p=0.014), the caudate putamen (-10%, p=0.019), the somatosensory cortex at Bregma +0.98 mm

and Bregma -1.34 mm (-10%, *p*=0.0013 and -3%, *p*=0.014, respectively) and the retrosplenial granular cortex (-9%, *p*=0.029).

Male *Slx1b*$^{+/-}$ exhibited six affected parameters all increased in size related to three brain categories (brain size, commissure and subcortex). The genu of the corpus callosum was enlarged by 11% (*p*=0.031), the total brain area by 7% (*p*=0.031), the soma of the corpus callosum by 10% (*p*=0.018), the height of the corpus callosum b 14% (*p*=0.032), the dorsal hippocampal commissure by 20% (*p*=0.018) and the dentate gyrus by 10% (*p*=0.018).

Male *Taok2*$^{+/-}$ showed six affected parameters spanning three main categories (brain size, cortex and commissure) including three reduced in size and three enlarged in size. While the total brain area was enlarged by 6% (*p*=0.0088), the cingulate cortex by 9% (P=0.0029) and the height of the corpus callosum by 27% (*p*=0.014), the anterior commissure was smaller by 17% (*p*=0.0068), the dorsal hippocampal commissure decreased in size by 15% (*p*=0.009) and the motor cortex smaller by 8% (*p*=0.0065).

Finally, male *Zg16*$^{+/-}$ exhibited seven affected parameters all increased in size pertaining to all categories at the exception of the ventricles. The anterior commissure was increased in size by 26% (*p*=0.004), the total brain area by 7% (*p*=0.011), the anterior commissure by 44% (*p*=0.00083), the fimbria by 13% (*p*=0.015), the habenular nucleus by 10% (*p*=0.039), the arcuate nucleus by 37% (*p*=0.012) and the piriform cortex by 21% (*p*=0.016).

4.8.2 Male NAP genes implicated in one brain category

Five genes (*Doc2a*, *Fam57b*, *Hirip3*, *Spn* and *Tbx6*) presented specific phenotypes, affecting one brain category.

Male *Fam57b*$^{+/-}$ showed smaller cortices across several brain parameters: cingulate (-11%, *p*=0.032), piriform (-18%, *p*=0.018 at Bregma +0.98 mm and -20% *p*=0.004 at Bregma -1.34 mm), retrosplenial (-19%, *p*=0.0018) and motor (-9%, *p*=0.016). By contrast, male *Tbx6*$^{+/-}$ showed thicker cortices: retrosplenial (+11%, *p*=0.000019), piriform (+21%, *p*=0.00069) and somatosensory (+8%, *p*=0.0029). Male *Doc2a*$^{+/-}$ showed enlarged somatosensory cortex (+8%, *p*=0.025), whereas *Spn*$^{+/-}$ thick genu of the corpus callosum (+12%, *p*=0.030) and male *Hirip3*$^{+/-}$ small anterior commissure (-24%, *p*=0.025).

4.8.3 Female NAP genes involved in two or more brain categories

Six NAP genes (*Slx1b*, *Bola2*, *Gdpd3*, *Tbx6*, *Doc2a* and *Spn*) gave significant results affecting two or more brain categories in female (**Figure 44**). These findings are described below.

Female *Slx1b*$^{+/-}$ exhibited three affected parameters all increased in size spanning two brain categories (commissure and brain size). It included an increase of 6% (p=0.0084) for the total brain area, 10% (p=0.016) for the soma of the corpus callosum and 9% (p=0.0024) for the height of the corpus callosum.

Female *Bola2*$^{+/-}$ showed 6 affected parameters all decreased in size, pertaining to three brain categories (commissure, brain size and cortex). The brain size was reduced by 9% (p=0.040) concomitant with a thinner somatosensory cortex by 11% (p=0.0053). The height of the genu of the corpus callosum was reduced by 15% (p=0.029) consistently the soma of the corpus callosum was smaller by 14% (p=0.0019) for its area, 11% (p=0.043) for its width and 10% (p=0.0097) for its height.

Female *Gdpd3*$^{+/-}$ affected three categories (commissure, ventricle and cortex) with an equal number of affected brain parameters. The lateral ventricle was increased by 81% (p=0.012), the corpus callosum was thicker by 14% (p=0.036) however the motor cortex was smaller by 10% (p=0.012).

Female *Tbx6*$^{+/-}$ showed increased measurements across three categories (subcortex, cortex and brain size). The total brain was enlarged by 10% (p=0.000056), the fimbria by 13% (p=0.00027), the piriform cortex by 16% (p=0.045) and the somatosensory cortex by 12% (p=0.000045).

Female *Doc2a*$^{+/-}$ had three increased parameters across two categories (commissure and cortex). The cingulate cortex was enlarged by 16% (p=0.037), the motor cortex enlarged by 10% (p=0.015) and the width of the soma of the corpus callosum longer by 8% (p=0.027).

Finally, *Spn*$^{+/-}$ exhibited three increased parameters spanning two categories. The corpus callosum height was thicker by 19% (p=0.02), the internal capsule was enlarged

by 14% (p=0.019) and the fimbria enlarged b 20% (p=0.017). Of note, these three brain parameters pertain to white matter structures of the brain.

4.8.4 Female NAP genes involved in one brain category

Four genes (*Ppp4c*, *Fam57b*, *Maz* and *Zg16*) showed specific phenotypes, affecting one single brain category (see **Figure 44**).

Female *Ppp4c*$^{+/-}$ showed one associated parameter, the height of the genu of the corpus callosum was diminished by 8% (p=0.013). Female *Fam57b*$^{+/-}$ showed a very mild effect on the total brain area (-3%, p=0.041). Female *Maz*$^{+/-}$ showed decreased height of the cingulate cortex (-11%, p=0.016) and finally, female *Zg16*$^{+/-}$ showed an increased size of the brain (+7%, p=0.011).

5. Material and Methods

5.1 Study samples and mouse knockout constructions

5.1.1 Single-gene knockout

The mouse syntenic 16p11.2 region *Sult1a-Spn* encompasses 30 protein-coding genes on chromosome 7: *Sult1a* (new name *Sut1a1*), *Slx1b*, *Bola2*, *Coro1a*, *Mapk3*, *Gdpd3*, *Ypel3*, *Tbx6*, *Ppp4c*, *Aldoa*, *Fam57b*, *4930451I11Rik* (previous name *C16orf92*), *Doc2a*, *Ino80e*, *Hirip3*, *Taok2*, *Tmem219*, *Kctd13*, *Asphd1*, *Sez6l2*, *Cdipt*, *Mvp*, *Pagr1a* (previous name *2900092E17Rik*), *Prrt2*, *Maz*, *Kif22*, *Zg16*, *AI467606*, *Qprt* and *Spn*. A description of each gene is provided in **Table 3** (columns A-G) including unique accession number based on the mouse genome browser Ensembl (https://www.ensembl.org/Mus_musculus/) assembly number GRCm38.p6, chromosome 7 strand, start and end position of the gene of interest, genomic length in base pairs and a description of the gene.

Among these 30 protein-coding genes, 20 (see columns H-L in **Table 3**) were incorporated in our analysis of neuroanatomical phenotyping explained in **Figure 42**. The mutant mice were acquired through collaboration with various sources. Two third came from five mouse production centers of the International Mouse Phenotyping Consortium (IMPC): i) Mouse Clinical Institute (ICS), Illkirch, France (*Slx1b*, *Gdpd3*, *Ppp4c*, *Fam57b*, *Taok2*, *Kctd13*, *Sez6l2*, *Maz*, *Zg16* and *Spn*) ii) Wellcome Sanger Institute (WSI), Cambridge, UK (*Prrt2*), iii) Mouse Biology Program (MBP), UC Davis, USA (*Hirip3*), iv) Italian National Research Council (CNR), Monterotondo, Italy (*Ino80e*), and v) Texas A&M Institute for Genomic Medicine (TIGM) USA (*Mvp*). One third came from researchers worldwide who previously studied these mutant mice outside of the field of brain development including *Bola2*[200], *Coro1a*[201], *Mapk3*[87], *Tbx6*[202], *Doc2a*[203], *Kctd13*[161] and *Qprt*[204]. Each mouse transfer was approved by a Mouse Transfer Agreement with the corresponding laboratories. For four lines (*Bola2*, *Coro1a*, *Prrt2* and *Qprt*), for which animal transfer was not possible, brain samples were obtained instead by collaborators instructed to use our own standard operating procedures[1,158].

Twenty-one mouse mutants were generated using a gene targeting approach (exception *Mvp* used gene trapping, see below) noted by the superscript “tm” for

targeted mutation or “em” for endonuclease mediated mutation (see column I in **Table 3**). The targeted approach comprised three strategies: i) the “Knockout-first allele” strategy[205] producing tm1a (*Fam57b*, *Ino80e*, *Sez6l2* and *Prrt2*) and tm1b (*Taok2* and *Sez6l2*) alleles (**Figure 54A**), ii) the reporter-tagged deletion strategy producing tm1 (*Bola2*, *Mapk3*, *Tbx6*, *Doc2a* and *Hirip3*) and tm1.1 (*Coro1a*, *Kctd13* and *Qprt*) alleles (**Figure 54B**), and iii) the CRISPR/Cas9 strategy producing em1 (*Slx1b*, *Gdpd3*, *Ppp4c*, *Maz*, *Zg16* and *Spn*) and em2 (*Kctd13*) alleles, similar to a previous report[206].

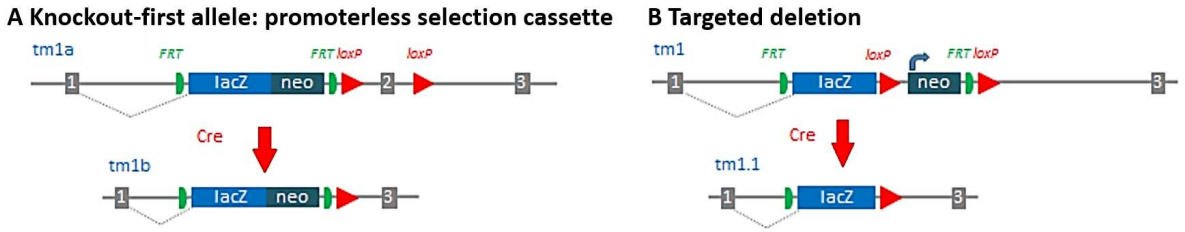


Figure 54 Mutant allele design.

(**A**) ‘Knockout-first’ strategy. (**B**) Deletion allele with the promoter-driven targeting cassette. Green boxes represent FRT sites cleaved in the presence of Flipase (Flp). Red triangles represent loxP sites cleaved by Cre-recombinase (Cre). The insertion of the LacZ and neo cassette is expected to disrupt gene function. Adapted from IMPC.

The “Knockout-first allele” strategy relies on the identification of a critical exon (CE) common to all transcript variants, upstream of which a LacZ cassette was inserted to make a constitutive knockout named tm1a. Unlike the tm1a allele, tm1b creates a frame-shift mutation upon deletion of the selected exon after constitutive Cre-recombination. Alleles annotated tm1 were generated by homologous recombination targeted to the CE and replacing it by LacZ and promoter-driven Neo cassettes, flanked by FRT and loxP sites. Tm1 alleles can generate a tm1.1 derivative allele through Cre excision, which will remove the Neo cassette from the locus. Alleles that were generated by CRISPR/Cas9 also relied on a CE that was deleted using four gRNAs (two gRNAs 5’ and two gRNAs 3’ to the CE region). A single genotype-confirmed F1 mouse was produced to establish the colony used to generate mice for phenotyping. For *Kctd13* both tm1.1 and em2 strategies were used. Mvp mouse model was generated using a gene trapping strategy at the Texas A&M Institute for Genomic Medicine (TIGM). The clone number utilized was IST10954E2. The strategy relied on an insertion of a β-galactosidase/neomycin phosphotransferase (β-Geo) cassette

within the first intron of the Mvp gene. The cassette was flanked by a splice acceptor and a polyadenylation site which triggers gene transcription premature arrest (**Figure 47A**). To ensure high comparability between mouse models, the mice used in this study came from the same genetic background strain C57BL/6 (see column J in **Table 3**). We maintained most of the mice on a pure inbred C57BL/6N background (Taconic Biosciences) representing more than 80% of the lines tested in this study to minimize variation in phenotyping results due to strain differences. One gene (*Kctd13*) was tested on both C57BL/6N (B6N) and C57BL/6J (B6J) genetic background.

Among the 30 protein-coding genes of the mouse 16p11.2 syntenic region, ten were not available during the course of the study for various reasons (see column H in **Table 3**). Two mutant lines (*Pagr1a* and *Kif22*) were at the stage of embryonic stem cell production therefore were not yet produced. Five (*Aldoa*, *4930451I11Rik*, *Tmem219*, *Cdipt* and *AI467606*) were cryopreserved at the Baylor College of Medicine (Texas USA) and one (*Ypel3*) at the Mouse Biology Program (MBP) UC Davis USA, but their rederivation was proven difficult. Two (*Sult1a* and *Asphd1*) were reassigned to a new partner of the IMPC: the Czech Centre for Phenogenomics (CCP) at the Institute of Molecular genetics of the Czech Academy of Sciences (IMG), Czech Republic, Europe.

5.1.2 *Mvp*::*Mapk3* double knockout model

To study genetic interaction between *Mapk3* and *Mvp*, we generated a double-knockout line by crossing single-gene mutant models.

Because the two genes are on the same locus, the gametes resulting from meiosis of each can only segregate for one of the mutation, with 50% chance to segregate as a heterozygous or WT haplotype. Our breeding strategy gives rise to four groups, equally distributed with 25% chance to get each genotype in the F1 generation. For this study, we bred *Mapk3* mice with *Mvp* mice. The sex of the parent bearing each allele deletion was not taken into account. Considering the core strain of each of our mouse model, F1 generation from *Mvp* and *Mapk3* breeding was characterized by a mixed genetic background (50%B6Nx50%B6J).

Double-heterozygotes, WTs as well as the two intermediate heterozygotes for each single-gene knockouts were present in the offspring segregating according to

Mendelian ratio. Four groups of mice were therefore assessed for neuroanatomical characterization.

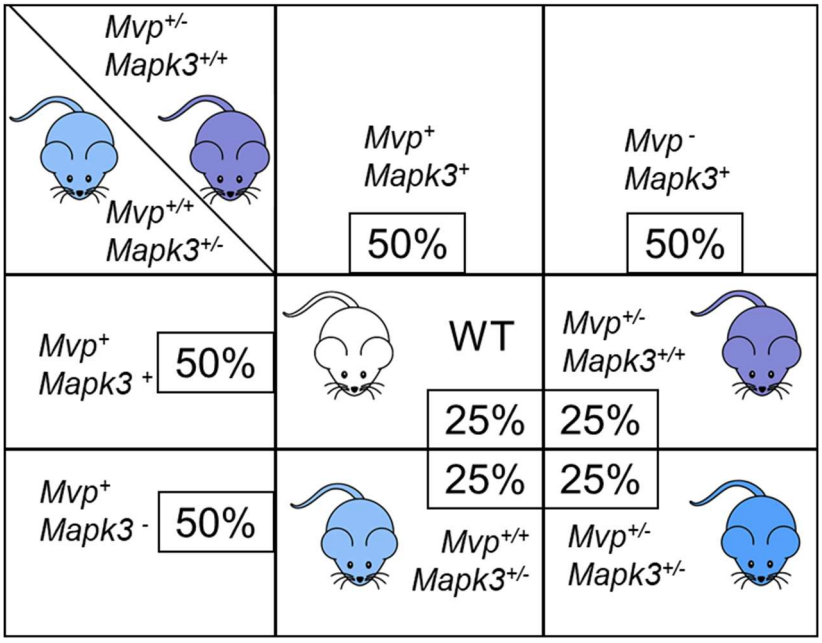


Figure 55 Double knockout breeding strategy.

Punnett square recapitulating the expected birth ratio for each genotype after the breeding of *Mvp* and *Mapk3* heterozygous mice. *Mvp*$^{+/+}$*::Mapk3*$^{+/+}$, *Mvp*$^{+/-}$*::Mapk3*$^{+/+}$, *Mvp*$^{+/+}$*::Mapk3*$^{+/-}$ and *Mvp*$^{+/-}$*::Mapk3*$^{+/-}$ are generated at equal ratio (25%).

5.1.3 *16p11.2*$^{+/Del}$ mouse models

Mice deleted for the entire 16p11.2 locus (*16p11.2*$^{+/Del}$) were generated and analyzed for NeuroAnatomical phenotype. The deletion encompassed the region from *Sult1a* to *Spn* and was previously described[74]. In this study, we decided to test the impact of this chromosomic rearrangement in two different strains (C57BL6/N and C3B), in order to see if the genetic background has an impact on cerebral morphology for this mutation.

5.2 Animal welfare

Seventeen of the 21 single-gene mouse lines were housed and maintained within our own animal facility at the Mouse Clinical Institute, Illkirch, France (see column L in **Table 3**). A typical stocking density of 3-5 mice was used per cage in individually ventilated caging under a controlled 12-hours light/dark cycles. Animals were provided with chow food and water *ad libitum*. The ambient temperature was 21±2°C and the humidity was 55±10%. All animals were regularly monitored for health and welfare concerns and were additionally checked prior to and after procedures. In addition to

bedding substrate, standard environmental enrichment of a couple of nestlets and a cardboard tunnel were provided. The care and use of mice in the study was carried out in accordance with the local ethics committee (Com'Eth) under the reference number 2016010717527861. The housing conditions described for single-gene models were the same than the ones for double-knockout models, as well as *16p11.2*$^{+/Del}$ mice.

5.3 Genotyping and sexing of study samples

Genotyping primers were designed for each of the seventeen mouse lines housed in our own animal facility (Mouse Clinical Institute, Illkirch, France). A list of sequences is provided in

along with respective annealing temperature and expected product size in both WT and KO conditions. Sexing was determined visually except for embryos and young postnatal age where we combined visual assessment with genetic testing using the SRY reactions.

Mice were genotyped according to the following instructions. Genomic DNA was extracted from mouse tail clip based on sample lysis by detergents and purified by precipitation using isopropanol. Polymerase chain reactions (PCR) were carried out with dNTPs (Roche), 10x PCR Buffer (Roche) and TaqPolymerase (Roche) in a MasterCycler (Eppendorf) under the following conditions: 95°C for 4minutes (1 cycle), 94°C for 30 seconds followed by another 30 seconds at primer-specific annealing temperature and 72°C for 1minute (34 cycles), 72°C for 7minutes (1 cycle) and then 4°C until use. PCR products were then loaded on 3% agarose gel with Ethidium Bromide (BET) (0.3µg/ml, Euromedex) and run for 25 minutes at 130V in an electrophoresis setup. The gel was then revealed using UV light.

The genotyping strategy to assess double-knockout lines consisted of a combination of the genotyping of corresponding single-gene models. DNA extracts were used with specific primers and genotyped separately. The combined genotypes were then merged in the same file to obtain final reading. The genotyping of the *16p11.2* mice was conducted as previously described[74].

Genotyping					
Gene	**Target**	**Primer Name**	**Sequence**	**Annealing Temperature (°C)**	**Product size (bp)**
Ino80e	Ino80e_WT Ino80e_KO	CCDC95FOR	CAGGGATGCACATACCTTGAGAGT	72	115 239
		CCDC95REV	GCTCAGTGTTCTTCCTAGGAGACT		
		FTR REV	CCTTCCTCCTACATAGTTGGCAGT		
Tbx6	Tbx6_WT	Tbx6_wtF	GGGAGAATGAGGATCCAGG	62	200
		Tbx6_wtR	TACCATCCACGAGAGTTGTAC		
	Tbx6_KO	Tbx6_mutF	ATTGCACGCAGGTTCTCCGG	62	550
		Tbx6_mutR	GTCACGACGAGATCCTCGCC		
Doc2a	Doc2a_WT	Doc2a_wtF	TGGGTGCAGATTACAGCTTC	60	325
		Doc2a_wtR	TGAGGATTCTACAGTGCAGC		
	Doc2a_KO	Doc2a_mutF	GGGCGCCCGGTTCTTTTTGTC	60	224
		Doc2a_mutR	GCCATGATGGATACTTTCTCG		
Kctd13	Kctd13_WT Kctd13_KO	Kctd13_wtF	GGGGAAGGGCTTAACATAGA	60	209
		Kctd13_wtR	AGATCCCTCCCAAATCTGCT		
	Kctd13_KO	Kctd13_koF	TCCATGAATGCGAAGTGAAG	60	521
		Kctd13_koR	GGCCAAACTTGATGACCAGT		
Mapk3	Mapk3_WT	Mapk3_for	GCTGAGCAATGACCACATCTGCTA	58	438
		Mapk3_rev	GGGAAACTCTCCTACCTTGGAATTA		
	Mapk3_KO	Mapk3_f3	GCCTCTGCAACTCTGACCACAGGA	58	650
		Mapk3_Neo9	GGACATAGCGTTGGCTACCCGTGA		
Mvp	Mvp_WT	MVP Ef	AGGACAGAGCCTGGAAGTCA	62	512
		MVP Er	TTTTGCTTGGGATGGCTAAG		
	Mvp_KO	MVP Ef	AGGACAGAGCCTGGAAGTCA	62	308
		MVP Kf	CTTGCAAAATGGCGTTACTTAAGC		
Sexing	SRY	SRY-F	TTGTCTAGAGAGCATGGAGGGCCATGTCAA	62	272
		SRY-R	CCACTCCTCTGTGACACTTTAGCCCTCCGA		

Table 9 Genotyping and sexing primers.

List of the different primers used in this study to determine the genotype and sex of each mouse. Annealing temperature for PCR and expected product size are indicated.

5.4 Brain histological screen

5.4.1 Sample processing

Brain samples were harvested from mice using a high-throughput approach, where each mouse was characterized by a series of standardized operating procedures[8,9]. The collection of brain samples was performed blind with experimenters not knowing the genotype of the mouse. Factors thought to affect the variables were standardized as far as possible. Where standardization was not possible, steps were taken to reduce potential bias. For example, as recommended in the ARRIVE guidelines[207], we used a “minimized operator” defined as “The process by which steps are taken to minimize the potential differences in the effector by training and monitoring of operator”.

Brain samples from at least three mice per genotype and per gender were collected, which was estimated based on power calculation, limited throughput of both the necropsy collections and histological workflows and 3Rs in regards to animal use.

Control brains were systematically collected within each of the mutant lines on the same genetic background. Six cohorts representing a total of 458 mice were analyzed with the main difference being the number of genes targeted by the mutation. The first was for single-gene mutation, consisted of 347 mice made of 180 mutants and 167 matched WTs. The second dataset was for the analysis of homozygous *Kctd13* mice, composed of 14 individuals. The third dataset was for the analysis of *Mvp* homozygous mice, composed of 10 individuals. The fourth was for *Mvp* analysis at different time points, composed of 42 mice of 13 mutants and 29 matched WTs, The fifth was for *Mvp::Mapk3* double-gene mutations composed of 20 mice with 9 mutants and 11 matched WTs, and the sixth for deletion of the entire 16p11.2 interval comprising 25 mice (13%) of 12 mutants and 13 matched WTs.

For validation purposes, a set of mutants was studied multiple times through the production of independent cohorts or at different ages. Typically, adult mice aged 16 weeks old were studied (exception of *Ino80e* and *Ppp4c* aged 24-25 weeks old). *Kctd13*$^{<tm1.1Nk>}$ mice were studied at 8 and 16 weeks of age.

An overview of sample processing is shown in **Figure 42**. Standard operating procedures are described in more details elsewhere[8,9]. Animals were euthanized in a CO_2 chamber, and brains were carefully collected and fixed in 10% neutral buffered formalin. The samples were transferred to 70% ethanol 48 hours after drop fixation. The cerebellar part was discarded and the remaining brain was trimmed at coronal plane around Bregma -0.10mm to produce two blocks and then embedded in paraffin using an automated embedding machine (Sakura Tissue-Tek VIP). Brains were cut at a thickness of 5µm on a sliding microtome (HM 450 Microm France) on symmetrical and stereotaxic planes such that we obtained sections matching well defined planes at Bregma +0.98mm and Bregma -1.34 mm. Our precision was estimated to be no more than 30µm, anterior or posterior, to the histological section of interest. The sections were doubled-stained with 0.1% Luxol Fast Blue (Solvent Blue 38; Sigma-Aldrich) and 0.1% Cresyl violet acetate (Sigma-Aldrich), in order to label myelin and neurons, respectively. After mounting on slides, the sections were scanned at cell-level resolution using the Nanozoomer whole-slide scanner 2.0HT C9600 series (Hamamatsu Photonics, Shizuoka, Japan).

Every aspect of the procedure was managed through a relational database using the FileMaker (FM) Pro database management system (detailed elsewhere[165]), for example generating image scan names and directory architecture for 2TB of image data, image quality control, measurements and statistical pipelines using FM build-in scripting capabilities and an interface with R (version 3.4.0, https://www.r-project.org/) and ImageJ (Fiji, version 1.51e) scripts. Within this database, co-variate and measurement data for about 1000 brain images were collected and QCed entirely blind to the genotype.

5.4.2 Histomorphometric analyses and quality control

Figure 42 gives the overall brain histomorphometric procedure. Careful and precise sectioning is a prerequisite in this protocol; sections must be symmetrical and match well-defined anatomical features (explained above). By contrast to more conventional histopathological screens that often rely on qualitative assessment, we used a quantitative approach where each section had to pass well-defined stereotaxic coordinates defined accordingly to the Mouse Brain Atlas before image analysis. Each image was quality controlled to assess whether (i) the section is at the correct position, (ii) the section is symmetrical, (iii) the staining is of good quality, and (iv) the image is good quality. Only images that fulfilled all of the quality control checks were processed. These quality control steps are essential for the detection of small to moderate neuroanatomical phenotypes (NAPs) and without which the large majority of NAPs would be missed. This is explained in great details elsewhere[165].

A total of 35 co-variates, for example sample processing dates and usernames were collected at every step of the procedure (**Dataset 1**), as well as 67 brain morphological parameters of 39 area and 28 length measurements (**Table 4**). Out of the 67 coronal brain parameters analyzed, 50 were measured on both the left and the right hemispheres including 20 in the cortices and 8 in the hippocampus. Since our histomorphological procedure does not allow interhemispheric comparisons for sure (sometimes sections can turn upside down during the histological procedure), we combined measurements of left and right hemispheres, reducing the number of coronal parameters to 42.

These parameters encompass five main categories: brain size, commissures (callosal, anterior and hippocampal), ventricles (lateral and third), cortex (motor, somatosensory,

cingulate, piriform and retrosplenial) and subcortex (hippocampus, amygdala, tracts, caudate putamen, internal capsule, habenula, fimbria and hypothalamus). The total number of neuroanatomical measures taken throughout the study was 19,278. All samples were also systematically assessed for cellular ectopia (misplaced neurons). Of note, amongst all the mouse mutant lines assessed in the study, we found no occurrence of ectopia.

5.4.3 Statistics

We began by a quality control step of the collected data and checked for the presence of erroneous measures, typographical errors and outliers, and corrected the measurement if necessary. At this step, we removed from the study female *Qprt*$^{+/-}$ data, since the data quality checks for this group did not fulfil our expertise criteria. To assess whether assumptions for statistical tests were met, quantile plots were drawn for each of the brain parameters in matched WTs where normality was met for most parameters at the exception of the size of the ventricles. Data were analyzed using a two-tailed Student t-test of equal variance to determine whether a brain region is associated with a neuroanatomical phenotype or not. We critically evaluated any results with p-values lower than 0.05. The mutant samples were analyzed using their colony controls (the same mouse line, the same genetic background). The percentage change of affected structures relative to the WTs (set to 100%) was also calculated as following: %change= (average mutant/average WT)*100.

Dataset 2 provides association and percentage change data for the twenty-one assessed single-gene alleles across 42 left and right combined parameters. In order to determine the level of confidence of each of the NAP, they were ranked within six different categories (NeuroAnatomical Threshold (NAT), defined after their p-value (below 0.05, 0.01, 0.005, 0.001, 0.0005 or 0.0001). The definition of a NAP is only possible according to a specific NAT.

5.5 Follow-up histology studies of MVP/vault

5.5.1 Embryonic stage

For embryos at E18.5, the whole embryo was fixed in Bouin solution for 48 hours. The embryonic brains were then harvested, transferred to 70% ethanol, and manually embedded in paraffin using the following steps: three incubation baths in 70% ethanol

for 30 minutes each, two baths in 95% ethanol for 30 minutes each, two baths in 100% ethanol for 45 minutes each, three baths in Histosol Plus for 1 hour each, and five baths in warm paraffin (60°C) for 30 minutes each, followed by incubation in warm paraffin overnight before casting in a mold. Brains were cut at a thickness of 5µm on a microtome (HM 450, Microm Microtech, France), such that we obtain sections matching planes displayed in **Figure 56** for embryonic coronal sections. The sections were stained with 0.1% Cresyl violet acetate (Sigma-Aldrich) and scanned using Nanozommer 2.0HT, C9600 series at 20× resolution.

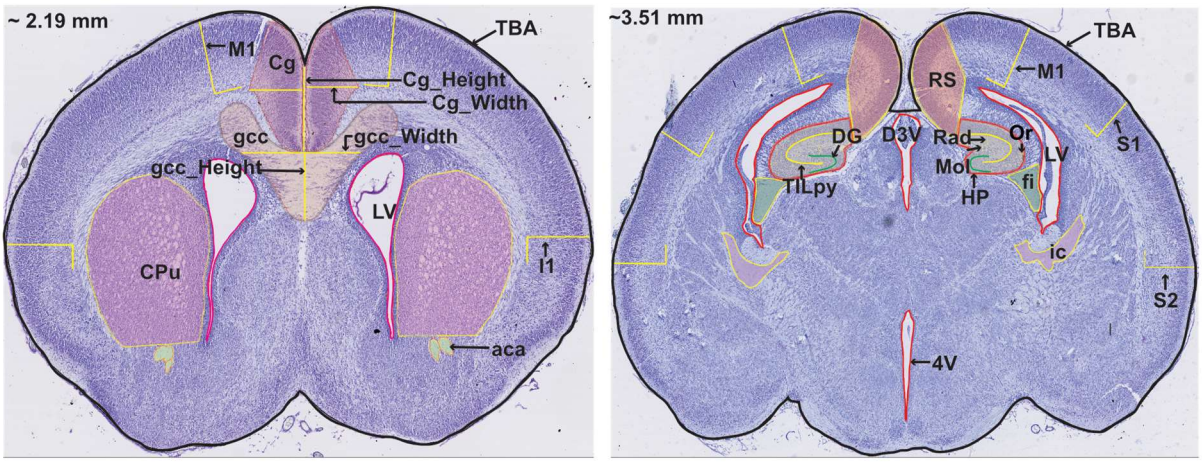


Figure 56 Coronal section of interest at E18.5.

Measured regions are delineated and associated with their parameter names on representative images of the two coronal sections of interest. See **Table 7** for the description of acronym used.

5.5.2 Sagittal orientation

After dissection and fixation, the brains were cut into halves along the sagittal midline, and embedded separately in paraffin. The section of interest was determined as corresponding to Figure 106 of the Mouse Brain Atlas (Lateral +0.60 mm)[158]. Brains were sectioned to match the defined section at a thickness of 5µm. The staining, scanning, and quality control steps were identical to the coronal procedure. 40 brain morphological parameters, made of 25 areas, 14 lengths and 1 number, were measured for each brain. This is described in more details in **Table 8** and in[158].

5.5.3 Determination of layers in somatosensory and motor cortices

Cortical layers are easily identifiable on histological sections. One of our question was to determine if the neuroanatomical defaults observed in male cortices were restricted to specific layers or if the whole cortex was impacted. To do this, we set the line used to measure the length of these cortices (1_S2 and 1_M1 (in red in **Figure 57**)) as the base to create a rectangle, with 0.22mm (green) or 0.65mm (yellow) wideness, respectively, which allows the measurement of the area of these cortices.

Then, we manually delineated each of the layers according to their cytoarchitecture features: Layer I: external layer with very few cells; Layer II/III: densely packed neurons; Layer IV (not present in motor cortex): more densely packed neurons; Layer V: loosely packed with bigger neurons; Layer VI: medial, densely packed neurons. The area as well as the cell count and averaged cellular area were measured and calculated in these regions.

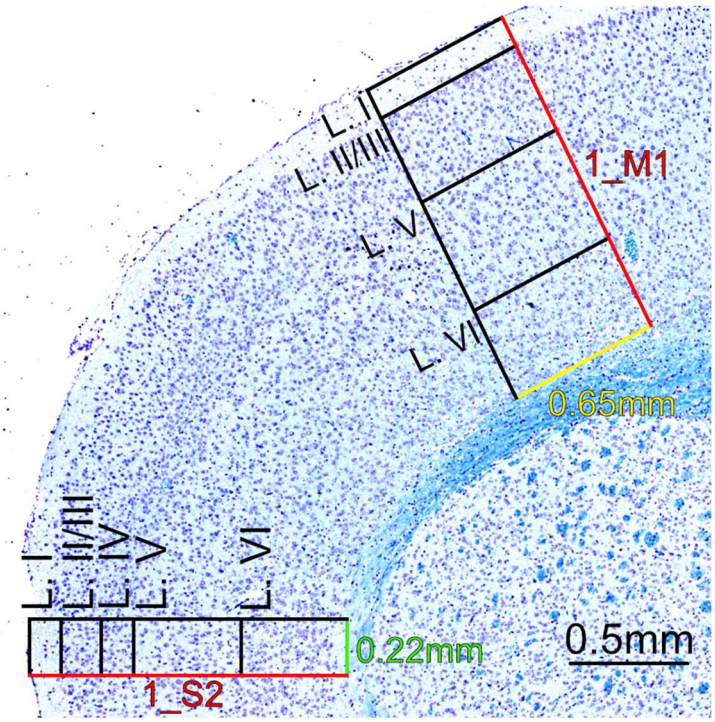


Figure 57 Cortical layer determination.

5.5.4 Measurement of cellular features

We developed an automated plugin in Fiji to measure the number of cells as well as the averaged cell area of each cell. These measurements are done on areas already

delineated when measuring the area of the structure of interest. For example, cingulate gyrus cell count (1_Cg_cellcount) and cell area (1_Cg_avgcellarea) are measured in the area determined as being the cingulate gyrus (1_Cg) in the automated protocol. The total area of all cells within the region (cellarea) is divided by the number of cells in order to determine the averaged area of each cell (avgcellarea): avgcellarea = cellarea/cellcount.

5.6 Evaluation of MVP/vault transcript level

5.6.1 Tissue dissection

Whole brain from WT and *Mvp* mutant mice were harvested at different developmental (E16.5, E18.5, P0, P10, P30) and adulthood (10, 16 and 30 weeks old) stages. Until P10, mice were decapitated with sharpened scissors and P30 and adult cohorts by increased CO_2 inhalation in gas chamber. For some adult cohorts, brains underwent microdissection in order to separate various brain structures of interest (such as cortex (CTX), hippocampus (HP), striatum (STR) and cerebellum (CRB)). Anterior cingulate cortex and primary somatosensory cortex were very carefully harvested using a mouse brain matrix. Additional peripheral tissues such as liver, lung, kidney and heart were also extracted from the same mice. Tissue samples were stored in cryotubes (Nunc) and deeply freezed in liquid nitrogen to prevent RNA degradation. Excess of tissue samples were kept at -80°C for long-term storage.

5.6.2 RNA extraction

Total RNA was extracted from samples by using a phenol-chloroform technique. Samples were transferred and homogenized with TRI Reagent (T9424, Sigma Aldrich) in tubes containing ceramic beads (Precellys Lysing Kit). The phase separation was done by addition of chloroform followed by centrifugation at 12000g for 5 minutes. Supernatant was collected and RNA extraction was conducted with RNeasy Plus Mini Kit (74134, Qiagen) as told by the manufacturer. RNA concentration and quality were assessed by spectrophotometry (Nanodrop 2000, Thermo Scientific). Only RNA samples with a ratio A_{260}/A_{280} close to 1.8 and a ratio A_{260}/A_{230} above 2.0, to exclude guanidine thiocyanate-containing buffer contamination, were further processed. RNA integrity number (RIN) was assessed using the Bioanalyser (Agilent 2100). Samples with RIN < 8 were not used in downstream experimental settings.

1µg of total RNA was reverse-transcribed to complementary DNA (cDNA) using SupercriptIII First-Strand Synthesis Supermix (11752–050, Invitrogen) for qualitative and quantitative reverse transcription PCR (RT-PCR). cDNA was stored at -20°C for future usage.

5.6.3 Qualitative RT-PCR

RT-PCR reactions were utilized in order to visually assess transcript expression level. This was done in four independent experiments: i) compare differences in *Mvp* expression between brain structures in WT, ii) compare *Mvp* expression patterns between male and female WT mice, iii) validate the *Mvp* mutant mouse model, iv) assess the expression level of the overlapping gene *Pagr1* in the *Mvp* mutant mouse model. PCR reactions were carried as described in section 4.3 of the present document. Primers were specially designed in order to amplify a region spanning two adjacent exons in order not to amplify genomic DNA (**Table 10**). PCR products were then loaded on 3% agarose gel containing BET and migrated by electrophoresis at 130V.

Expression					
Gene	**Target**	**Primer Name**	**Sequence**	**Annealing Temperature (°C)**	**Product size (bp)**
Mvp	Mvp_202	MVP qPCR F	TCCCTCTGGACCAAAATGAG	65	114
		MVP qPCR R	CTGACCACTCGGGTCTTGTT		
Mvp	Mvp_Com	Mvp_rtPCR_Com_L	TCACCATGGCAACTGAAGAG	65	202
		Mvp_rtPCR_Com_R	GACACAGGGTTGGCCACTAT		
Pagr1b	Pagr1b	Pagr1b_L	CTCCGATCCTCAGGAAACAG	65	234
		Pagr1b_R	GGGCAAACAGTACCCTCTCA		
Mvp	Intron	MVP Ef	AGGACAGAGCCTGGAAGTCA	65	542
		MVP Er	TTTTGCTTGGGATGGCTAAG		
Pagr1a/b	Pagr1ab	Pagr1ab_Com2_F	AGCCAATGACACCAAAGGAC	65	218
		Pagr1ab_Com2_R	GTTTCCTGAGGATCGGAGTG		
Mvp	Mvp_201	Mvp_rtPCR_201_L	GGAAGGTGGGATCAAGGAAG	65	172
		Mvp_rtPCR_201_3_R	CCAACCTCTACACGGGACAC		

Table 10 Sequences of RT-PCR primers.

5.6.4 Quantitative RT-PCR

Efficiencies of the TaqMan assays were checked using a cDNA dilution series from the extracts of cortex (CTX), hippocampus (HPC), cerebellum (CRB) and liver (LIV). PCR reactions were performed using Fast Blue qPCR Mastermix Plus (RT-QP2X-03+FB; Eurogentec) with 300nM of each primer and 100nM of FAM-labelled TaqMan probes in a final reaction of 15µl with a standard amplification procedure (T°A=60°C, 40 cycles). Normalization was performed by carrying out in parallel the amplification of

two other housekeeping genes (*Hprt* and *Gnas*). All the tested samples were performed in triplicate, and the results were reported as the mean ± SEM.

Primer Name	Sequence
Mvp_qPCR_F	TCCCTCTGGACCAAAATGAG
Mvp_qPCR_R	CCTTTTCCCACAGGACTTCA
Mvp_qPCR_Probe	ATGTCAAGACGGGAAAGGTG
Gnas_qPCR_F	AGAACATCCGCCGTGTCTTC
Gnas_qPCR_R	CCTTCTTAGAGCAGCTCGTATTGG
Gnas_qPCR_Probe	CGTGACATCATCCAGCGCATGCAT
Hprt_qPCR_L	GCAAACTTTGCTTTCCCTGG
Hprt_qPCR_R	CTTCGAGAGGTCCTTTTCACC

Table 11 Sequences of qRT-PCR primers and probes

5.7 Immunohistofluorescence

5.7.1 Sample processing

Adult mice were anesthetized using Ketamine (150mg/kg) and Xylazine (400ug/kg) in NaCl, and intracardially perfused with PBS, to remove blood within the vessels, and then fixed with 4% Paraformaldehyde (PFA) solution in PBS. Brains were harvested, post-fixed in 4% PFA for 24 hours at room temperature (RT), and then transferred in 30% sucrose solution for 24 hours at 4°C to ensure cryoprotection of the tissues. Each brain was then cut into either three (for coronal sectioning) or two (for sagittal sectioning) and embedded in Cryomatrix (Thermo Scientific) with a fast freeze device PrestoChill (Millestone).

5.7.2 MVP/Vault immunostaining

Brains were sectioned throughout the entire tissue blocks with a cryo-microtome (K400 station with dry ice on HM 450 microtome, Microm Microtech France) at 30µm in thickness. Floating sections were then stored at 4°C in 0.1% sodium aside in PBS.

Immunostainings for MVP were realized without detergent. The permeabilisation step was done by incubation of the brain histological sections in methanol at -20°C for 10 minutes. Sections were rinsed in PBS, blocked in a solution with 10% Normal Donkey Serum (NDS) and 1% Bovine Serum Albumin (BSA) in PBS and incubated with rabbit polyclonal vault antibodies (N2-B15, provided by Leonard Rome, dilution 1:1000) and mouse anti-NeuN antibodies (MAB377, Millipore, dilution 1:1000) in 0.1%NDS/PBS at

4 °C overnight under agitation. After washing with PBS, the sections were incubated for three hours RT with fluorescence-conjugated secondary antibodies coupled to anti-rabbit-Alexa-488 and anti-mouse-Alexa-647 (1:1000, Thermo Fischer Scientific). Nuclei were counterstained with Hoechst 33342 (1:10,000 dilution; Sigma-Aldrich), and sections were mounted on Superfrost blades using FluorSave mounting solution (Calbiochem). The slides were stored in the dark at 4 °C. Images were acquired using confocal microscope (TCS SP5; Leica) at 20× or 80x magnification and analyzed using ImageJ software. We carefully assessed MVP/vault expression as mild (+), moderate (++) or strong (+++) across all the positive brain regions (**Table 6**).

5.7.3 phospho-ERK immunohistofluorescence

ERK activity was studied in *Mvp*$^{+/-}$*::Mapk3*$^{+/-}$ double-heterozygous, their matching singe-heterozygous (*Mvp*$^{+/+}$*::Mapk3*$^{+/-}$ and *Mvp*$^{+/-}$*::Mapk3*$^{+/+}$) and WT mice in both male and female. The brains were sectioned (10µm), then rehydrated in PBS for 10 min. Antigen retrieval using 1× Reveal Decloaker (Biocare) was performed for 10 min at 95°C. Sections were blocked in 2% (postnatal) normal goat or donkey serum for 1 h at room temperature with 0.1% (v/v) Triton X-100 in PBS. Slides were incubated with primary antibodies (polyclonal rabbit anti-pERK (Cell Signaling Technology; 1:100)) overnight at 4°C, rinsed with PBS and incubated with corresponding secondary antibodies (AlexaFluor 488 conjugated to goat anti-rabbit (Invitrogen) (1:1000)) for 1–2 h at room temperature. DNA was stained with 4′,6′-diamidino-2-phenylindole for 5 min (Invitrogen).

We counted at least three consecutive tissue slices per animal. In postnatal brains, coronal sections were used to count cells in 400µm boxes in prefrontal cortex. Slides were picked at random and the investigator was blinded to genotypes. The number of mice evaluated is indicated in figure legends, with three to seven mouse brains per group.

5.8 Western Blot Analysis

Livers from adult male mice were dissected and homogenized with 300µL of lysis buffer containing 1× RIPA buffer (ThermoFischer), phenylmethylsulfonyl fluoride 1%, sodium orthovanadate 1%, and protease inhibitor 1% in tubes containing ceramic beads (Precellys Lysing Kit). The tubes were incubated for 30 minutes at 4°C and centrifuged for 20 minutes at 17,000 × g at 4 °C, and the supernatant was isolated for

Western blotting. 60µg of protein was separated on 10% SDS/PAGE (Mini-PROTEAN TGX Gels, 12%, 10-well, 4561043, BIO-RAD) and transferred onto nitrocellulose membrane (#1620115, BIO-RAD). Membranes were blocked with 1% BSA diluted in tris buffered saline with Tween 20 (50mM Tris, 150mM NaCl, 0.05% Tween 20) and probed using either rabbit MVP-antibodies (named "George", provided by Leonard Rome), ERK-antibody (4370S, CellSignalling) or Phospho-ERK (4695S, CellSignallig) overnight at 4°C. Membranes were then incubated with horseradish peroxidase (HRP)-conjugated anti-rabbit (GE Healthcare) secondary antibody. Antibody–protein interactions were revealed using chemiluminesence (RPN2108; GE Healthcare), and relative protein expression was quantified using ImageJ (https://imagej.nih.gov/ij).

5.9 Immunocytofluorescence

5.9.1 Primary neuronal cultures

Female mice on gestation were culled in increasing CO_2 gas chamber, and embryos harvested at embryonic day 18.5 (E18.5) then placed in dissection solution S1 (Phosphate Buffer Saline (PBS) 1x, Bovine Serum Albumin 3mg/ml, D-glucose 3mg/ml, MgSO4 1mM). The skin and skull of each embryo were removed and the two brain hemispheres separated. The hippocampus and the cortex were harvested from each hemisphere and collected in their corresponding tubes with 500µl of S1 solution. The samples were then centrifuged at 10,000g for 1 minute, and the supernatant removed. 500µl of enzymatic dissociation solution S2 (S1 + Trypsin 2.5mg/ml and DNase I 2.5mg/ml) was then added to each sample and incubated for 20 minutes at 37°C by intermittently mixing the tubes. 500µl of Trypsin inhibition solution S3 (S1+ Soybean Trypsin Inhibitor (SBTI) 80µg/ml, DNase 130µg/ml and $MgSO_4$ 250µM) was added to each sample in order to stop dissociation and centrifuged at 10,000g for 5 minutes. The supernatant was removed without disturbing the pellet, and 300µl of mechanical dissociation solution S4 (S1+ SBTI 0.5mg/ml, DNase 0.08mg/ml and $MgSO_4$ 1.5mM) was added. The cells were then dissociated by pipetting up and down and centrifuged at 10,000g for five minutes. The supernatant was then replaced by seeding medium (DMEM, HEPES 10mM, Horse serum 10%, Gentamycin 50µg/ml and Insulin 1µg/ml), and cells were counted using the Hemocytometer (Malassez cell). Between 20,000 and 30,000 cells were plated on Poly-L lysine coated coverslips in 24-well plates and incubated at 37°C with 5% CO_2. After 24 hours, the medium was

replaced using filtered GIBCO Neurobasal Medium1x (Invitrogen) supplemented with GIBCO B27 medium 1x (Invitrogen), Penicillin, Streptomycin and L-Glutamine (0.5mM). After 4 days of *in vitro* culture (DIV4), the cells were fixed with 4% PFA in 6% sucrose for 15 minutes, and then stored in EtOH at 4°C until use.

5.9.2 Immunocytostaining

Fixed cells were incubated O/N at 4°C with primary antibodies (rabbit polyclonal anti-MAP2 (AB5622, Millipore) and mouse monoclonal antibody against pan-axonal neurofilaments (SMI-312R, Covance) both diluted at 1:1000 in saturation solution (0.2% Triton, 1% BSA, 10% Normal Donkey Serum (S2170, Dutscher) in Tris-Buffered Saline). Donkey anti-rabbit coupled to Alexa647 (ab150075, Abcam) and donkey anti-mouse coupled to Alexa488 (ab150105, Abcam) were used as secondary antibodies in saturation solution without Triton. Nuclei were stained with Hoechst 3342 (1:10,000 dilution; Sigma-Aldrich). The coverslips were then mounted on slides by using AquaPolyMount (PolyScience, Inc.) and stored at 4°C until imaging.

5.9.3 Cell morphological assessment

Images were acquired using a regular epifluorescence microscope 100x (Leica) at magnification of 0.55x and analyzed using ImageJ (https://imagej.nih.gov/ij). The area of the soma was determined following the MAP2 staining around the cytoplasm, often characterized by an ovoid-like shape. The processes and lamellipodia were entirely excluded in the measurement. Axonal length was analyzed by measuring the length from the start of the axon until the tip of the primary axonal process, stopping when the growth cone begins. The growth cone area was defined and measured from the beginning of the terminal widening of the axon and included all visible lameli- and filopodia (see **Figure 48I** for schematic explanation). Shapiro-Wilk normality test revealed that these datasets followed a lognormal distribution. We then converted all of our data into log(10) to perform the statistical analysis using a two-tailed Student's *t* test. 0.05 was set as the significance threshold. Graphs represent the mean + SEM of the raw data.

5.10 RNA sequencing

RNA-Seq libraries were generated from 200 ng of total RNA using TruSeq Stranded mRNA LT Sample Preparation Kit (Illumina, San Diego, CA), according to

manufacturer's instructions. Briefly, following purification with poly-T oligo attached magnetic beads, the mRNA was fragmented using divalent cations at 94°C for 2 minutes. The cleaved RNA fragments were copied into first strand cDNA using reverse transcriptase and random primers. Strand specificity was achieved by replacing dTTP with dUTP during second strand cDNA synthesis using DNA Polymerase I and RNase H. Following addition of a single 'A' base and subsequent ligation of the adapter on double stranded cDNA fragments, the products were purified and enriched with PCR (30 seconds at 98°C; [10 seconds at 98°C, 30 seconds at 60°C, 30 seconds at 72°C] x 12 cycles; 5 minutes at 72°C) to create the cDNA library. Surplus PCR primers were further removed by purification using AMPure XP beads (Beckman-Coulter, Villepinte, France) and the final cDNA libraries were checked for quality and quantified using capillary electrophoresis. These libraries were sequenced on the Illumina Hiseq 4000 as Single-end 50 base reads following Illumina's instructions.

RNA sequencing analysis and base calling were performed using RTA 2.7.3 and bcl2fastq 2.17.1.14. Reads were preprocessed using cutadapt 1.10[208]in order to remove adapter, polyA and low-quality sequences (Phred quality score below 20), reads shorter than 40 bases were discarded for further analysis. Reads mapping to rRNA and spike-in sequences were also discarded (this mapping was performed using bowtie 2.2.8[209]). Reads were then mapped onto the mm10 assembly of *Mus musculus* genome using STAR[210] version 2.5.3a (--twopassMode Basic). Gene expression was quantified using htseq-count 0.6.1p1[211] and gene annotations from Ensembl release 91. Statistical analysis was performed using R 3.3.2 and DESeq2 1.16.1 Bioconductor library[212].

5.11 Golgi staining

GolgiCox staining was performed using the FD Rapid GolgiStain Kit (FD NeuroTechnologies, Ellicott City, MD) on entire fresh brains, right after being removed from the skull (4WT, 6 *Mvp*$^{+/-}$, 4 *Mvp*$^{-/-}$), and processed as indicated by the manufacturer. After impregnation, brains were embedded in 3% low-melting agarose and cut with vibratome. 100µm thick sections were mounted on gelatine-coated slides and allow drying for two days before staining (following notice). Images were acquired with slide scanner Hamamatsu at 40x resolution with multi-layer mode (41 focal plan every 0.2µm), thus allowing the analysis of the dendrites over 8µm in the tissue. Typical

pyramidal neurons from somatosensory cortex were identified, and 25-100 µm-long secondary apical dendrites were analyzed. The length of the dendrite section as well as the number of visible tangential dendritic spines were recorded to calculate the spine density (spine density= # of dendritic spine counted/length of the dendritic section measured). Data were collected and are presented in **Dataset 3**. All the measurements were taken by the same operator, blind to the genotype and perform manually on NDPviewer2.0.

5.12 Electrophysiology

5.12.1 Preparation of acute slices

A total of 14 mice (8 *Mvp*$^{+/+}$ and 6 *Mvp*$^{-/-}$), 16-19 weeks old, were anesthetized (urethane i.p., 1.9 g/Kg) and killed by decapitation. The brain was removed and immediately immersed in cold (0-4°C) sucrose-based artificial cerebrospinal fluid containing (in mM): 248 sucrose, 11 glucose, 26 $NaHCO_3$, 2 KCl, 1.25 KH_2PO_4, 2 $CaCl_2$ and 1.3 $MgSO_4$ (bubbled with 95% O_2 and 5% CO_2). Transverse slices (400µm thick) were performed with a vibratome (VT1000S, Leica, Nussloch, Germany). Slices were maintained at room temperature in a chamber filled with artificial cerebrospinal fluid containing (in mM): 126 NaCl, 26 $NaHCO_3$, 2.5 KCl, 1.25 NaH_2PO_4, 2 $CaCl_2$, 2 $MgCl_2$ and 10 glucose (bubbled with 95% O_2 and 5% CO_2; pH 7.3; 310 mOsm measured).

5.12.2 Electrophysiological recordings

Slices were transferred to a recording chamber and continuously superfused with oxygenated artificial cerebrospinal fluid. Pyramidal cingulate neurons were recorded in the whole-cell configuration. Patch pipettes were pulled from borosilicate glass capillaries (Harvard Apparatus, Edenbridge, UK) using a P-2000 puller (Sutter Instruments, Novato, CA, USA). They were filled with a solution containing the following (in mM): 145 KCl, 10 HEPES and 2 $MgCl_2$ (pH 7.3, adjusted with KOH; osmolarity 310 mOsm adjusted with sucrose) (3.5–4.5 MΩ). All recordings were performed in presence of bicuculline (10µM) and tetrodotoxine (0.5µM. Voltage-clamp recordings were performed with a Multiclamp 700A amplifier (Molecular Devices, Union City, CA, USA) at a holding potential fixed at -60 mV. Recordings were acquired and analyzed with WinWCP 4.3.5 (courtesy of Dr. J. Dempster, University of Strathclyde, Glasgow, United Kingdom). All recordings were performed at 34°C.

5.13 Electronic microscopy

Adult (16 weeks) *Mvp* mutant mice and corresponding WT were deeply anesthetized with Ketamine (150mg/kg) and Xylazine (400ug/kg) in NaCl, and intracardially perfused with PBS, and then fixed with 2.5% glutaraldehyde and 2.5% PFA in cacodylate buffer (0.1 M [pH 7.4]). Portion of cingulate gyrus were then washed 30 min in cacodylate buffer, post-fixed with 1% osmium tetroxide in 0.1 M cacodylate buffer for 1 h at 4°C, and dehydrated through graded alcohol (50%, 70%, 90%, and 100%) and propylene oxide for 30 min each. Samples were oriented and embedded in Epon 812. Semithin sections were cut at 2µm with the Leica Ultracut UCT ultramicrotome and stained with 1% toluidine blue and 1% sodium borate. Ultrathin sections were cut at 70nm and contrasted with uranyl acetate and lead citrate. Electron microscopy observation and image acquisition were performed at 70kv with the Morgagni 268D electron microscope (FEI Electron Optics, Eindhoven, the Netherlands) and equipped with the Mega View III camera (Soft Imaging System).

5.14 Behavioral Analysis

Behavioral experiments were conducted at the Mouse Clinical Institute (Illkirch, France), by the same experimenter, blinded to the genotype to avoid any potential biases. All mice tested were given several days of rest between experiments and were age matched. Four cohorts of male mice were analyzed: A) 12 mice *Mvp*$^{+/-}$::*Mapk3*$^{+/-}$ *versus* 12 matched WT for the double heterozygous knockout mice of *Mvp* and *Mapk3*, B) 12 mice *Mvp*$^{+/-}$ *versus* 10 matched WT for the single-gene mutation of *Mvp*, C) 12 mice *Mapk3*$^{+/-}$ *versus* 12 matched WT for the first cohort of the single-gene mutation of *Mapk3* and D) 12 mice *Mapk3*$^{+/-}$ *versus* 12 matched WT for the second cohort.

We used a comprehensive pipeline of sixteen behavioral tests, including: (i) open field, (ii) elevated plus maze, (iii) fear conditioning, (iv) forced swim, (v) tail suspension test, (vi) pentylenetetrazol (i.v.PTZ) seizure test, (vii) grip strength test, (viii) rotarod coordination test, (ix) sucrose preference, (x) Y maze, (xi) novel object recognition 3 h, (xii) social recognition, (xiii) social interaction, (xiv) marble burying, (xv) circadian activity and (xvi) pre-pulse inhibition (PPI) test.

5.14.1 Behavioral pipeline

Behavioral analyses were done between 9 and 26 weeks of age to model phenotypes found in young adults with the 16p11.2 deletion. Only male mice were utilized for behavioral testing to avoid any estrous cycle interference. **Table 12** gives the details of tests and the age at which they were carried out.

			Age (in weeks)			
Function	**Test**	**Acronym**	***Mvp*** $^{+/-}$ ***:: Mapk3*** $^{+/-}$	***Mvp*** $^{-/-}$	***Mapk3*** $^{+/-}$ ***Cohort 1***	***Mapk3*** $^{+/-}$ ***Cohort 2***
Anxiety	Open Field	OF	12-13	11-15	13-15	15-17
	Elevated Plus Maze	EPM	12-13	13-17	13-15	
Memory	Fear conditioning	FC	19-20	20-24	21-23	
	Y-maze	YM	12-13	11-15	13-15	
	Novel Object Recognition	NOR	14-15	15-19	14-16	
	Social recognition	SR	16-17	16-20	18-20	
Locomotion	Circadian activity	CA	15-16	14-18	17-19	
Epilepsy	Pentylenetetrazol test	PTZ	25-26	21-25		13-17
Coordination	Rotarod	Rot	17-18	17-21		11-15
Motor function	Grip strength	GP	23-24			
Depression	Tail suspension	TST	18-19	20-24	19-21	12-16
	Forced swim	FS	23-24			
Anhedonia	Sucrose preference	SP	13-14	13-17		09-13
Sociability	Three-chamber	SI	22-23			
Autism	Marbles	Marbles	12-13	12-16	14-16	
Schizophrenia	Pre-Pulse Inhibition	PPI	18-19	16-20	19-21	

Table 12 Behavioral pipeline.

Summary of the sixteen behavioral tests realized on four different cohorts (*Mvp*::*Mapk3*, *Mvp* and two distinct *Mapk3*). Each test is associated with a core function. The acronym used throughout this study is indicated in the third column. The age range of the mice for each test is referred in weeks.

5.14.2 Open field

The open-field test allows studying basic locomotor activity, hyperactivity, exploratory behavior and anxiety in mice. The mouse was put in an arena of dimensions 44.3 × 44.3 × 16.8 cm made of PVC (Panlab) fitted with two frames of 32 infrared (IR) beams and illuminated at 150 Lux, and their activity was recorded for 30 minutes using a video tracking system (Ethovision; Noldus); data were analyzed using Acti-Track software.

Speed and distance covered by the mice were then quantified as well as the percentage of time spent in defined zones of the field (periphery, intermediate, or center) and the latency before entering the center for the first time. The natural avoidance of mice for highly luminous places was used as a proxy of anxiety level.

5.14.3 Elevated plus maze

We used the elevated plus maze (Imetronic) as an additional test to evaluate anxiety. The setup consisted of a plus-shaped arena illuminated at 50 Lux at an elevation of 66cm, with opposite arms being either open (without walls, 30 × 5 cm) or closed (with walls 30 × 5 × 15 cm). The apparatus was equipped with IR captors, allowing the detection of mouse movements in the enclosed arms and different areas of the opens arms. It was also possible to evaluate head dips over the open arm. The mouse was placed in the center and allowed to freely explore the arms for a period of 5 minutes. The index of anxiety was calculated against the time spent in the open arms as opposed to the total time spent exploring arms.

5.14.4 Fear conditioning

Learning and memory of aversive stimuli is a key function to assure survival for mice. To test for this, fear conditioning was used as a three-phase test: conditioning, contextual testing and cued testing. This was measured in a operant chamber (Coulbourn, Bilaney, Dusseldorf, Germany, www.blaney.com) to associate light/tone stimuli with a mild electrical foot shock. Freezing was measured as a proxy of learning/memory performance during each phase by using infrared sensor. During the conditioning phase, the mouse was placed in the chamber for six minutes. There were four minutes of habituation (Hab-1 and Hab-2, two minutes each) where freezing time was measured. This was followed by tone/light stimuli (Cue: blue light and 80 dB/10kHz tone) for 20 seconds and a mild electrical foot shock (Unconditional stimulus=US) of 0.4 mA for one second. Freezing time was recorded for two additional minutes (Post-US). On the following day, the mouse was put back in the operant chamber in which it had been trained to associate a tone/light with an electrical foot shock. Freezing to the background context (Cont) was measured over a period of six minutes (Cont-1, Cont-2 and Cont-3, two minutes each). Five hours later, the mouse was placed in a new chamber of distinct appearance (i.e. different wall color, floor texture odor) and amount of time freezing was measured for eight minutes. At the end of the first two minutes

(Pre-cue1), the tone/light stimuli were presented for a period of two minutes (Cue1) and freezing measured again. This sequence was repeated once again (Pre-Cue2 and Cue2).

5.14.5 Forced swim

The forced swim test (also known as the despair test) was used to assess depression-like behaviors. The mouse was placed in a glass becher filled with water (20-21°C) with no possible escape and let for six minutes. The duration of immobility as well as the latency before the first immobilization were recorded.

5.14.6 Tail suspension

The tail suspension test is another way to evaluate depression-like behaviors during which the animal cannot escape. The mouse was suspended above the ground by their tail with adhesive and observed for 6 minutes. The duration (in seconds) of immobility as well as the latency before the first immobility were evaluated.

5.14.7 Pentylenetetrazol

Pentylenetetrazol (PTZ) is a convulsing drug used at high dose to model status epilepticus in animal models. After intraperitoneal injection of PTZ (50mg/kg in NaCl 0.9%), the mouse is susceptible to enter clonic (shaking of the body due to skeletal muscles contractions) and/or tonic phase (elongation of the limbs). The duration of each phase and the latency before entering the seizure were manually recorded.

5.14.8 Grip strength

Neuromuscular functions were assessed by recording the maximal force (in grams) that a mouse can exert while grasping at a metal grid attached to a dynamometer (Bioseb). The mouse was placed on the testing apparatus to grasp the grid and gently pulled back by their tail until the release. Resulting grip strength force was recorded and normalized as percentage of bodyweight. This test was used to evaluate forelimb muscular strength as well as a combination of the four paws and repeated for four successive trials.

5.14.9 Rotarod

The rotarod test was used to assess sensorimotor coordination. The mouse was placed facing the direction of rotation in a rotating bar of 5cm diameter (Bioseb) at several inches from the table. The mouse underwent three days of training during

which the mouse was trained to stay on the rod for two minutes with increasing speed (from 4 to 40rpm) for maximum five minutes. The test was stopped when the mouse felt down. Each mouse underwent four trials every day and the latency before falling off the rotating bar was recorded. On the test day, the mouse was placed on the rod at determined speed (4, 10, 16, 22, 28, 34 or 40 rpm) and let for maximal two minutes (120 seconds). The test was stopped when the mouse felt down and the latency before falling off the rotating bar recorded.

5.14.10 Sucrose preference

Mice have a natural preference for sweets, which is associated with pleasure-seeking behavior. The mouse was exposed to two different drinking bottles, one with water and one with 0.8% sucrose, and recorded for its drinking consumption. In the habituation phase, the bottles were exposed for one hour for training purposes. On the following day, the bottles were left for 15 hours (from 6am to 9pm), and the weight of the bottles were measured. Sucrose preference was based on the percentage of sucrose beverage consumption over the total consumption (sucrose + water). Sucrose preference superior to 50% was considered as normal pleasure-seeking behavior and below 50% as anhedonia, associated with depression-like behavior.

5.14.11 Y-maze

The Y-maze test was used to evaluate short-term working memory based on the innate curiosity of the mice to explore an arm that has not been previously explored; this preferential behavior, when occurring at a frequency greater than 50%, is called spontaneous alternation. The testing apparatus consisted of a Y-shaped maze with three white, opaque, Plexiglas arms of equal length (40 × 9 × 16 cm) at 120° angles to each other, each identified with walls designed with unique motifs, illuminated at 100 Lux. Animals were placed at the center of the maze and allowed to explore freely the three arms for 8 minutes. The numbers of arm entries were recorded and quantified to yield percentage of alternation.

5.14.12 Novel object recognition

The novel object recognition is similar to the Y-maze test, as it is based on the innate tendency of mice to explore novel objects over familiar ones. The animals were habituated (100-Lux illumination) and tested (70-Lux illumination) in the open-field arena. They were first presented with two Object A, glass marble or plastic dice, and

the exploration time was recorded. After a retention period (3 hours to test short-term memory), the mice were presented again with one object A (ObjectFam) together with a novel object B (ObjectNew), and distinct exploration time for each objects were manually recorded. The calculation was based on the percentage of time spent exploring the new object over total time spent exploring both objects.

5.14.13 Three-chamber

The social recognition test is quite similar to the novel object recognition test. Instead of objects, the test mouse was left to explore an unfamiliar mouse placed in its proximity. The arena was made up of three identical chambers of identical size (illuminated at 90 Lux), a central chamber with two adjacent chambers connected internally with doors. In each adjacent chamber, a goal box was positioned delimited by a sliding grid. The test mouse was first allowed to freely explore the three chambers for 10 minutes. The test mouse was then placed in the central chamber, and an unfamiliar mouse was briefly introduced into one of the adjacent wire boxes kept closed. The doors of the central chamber were opened; the test mouse explored the three chambers, and its exploratory behavior was recorded for 10 minutes, specifically its preference for each of the wire boxes (with mouse or without mouse). Again, the test mouse was placed in the central chamber, and another unfamiliar mouse (novel mouse) was introduced briefly into the other wire box kept closed. The doors of the central chamber were opened; the test mouse explored the three chambers, and its exploratory behavior was recorded for 10 minutes. Its preference for exploring the wire box of the novel as opposed to the wire box of the familiar mouse was evaluated.

5.14.14 Social interaction

Social interaction was evaluated by putting two mice of the same genotype, sex, and weight but from different cages in an open field (dimensions 44.3 × 44.3 × 16.8 cm) illuminated at 50 Lux for a duration of 10 minutes. Several parameters were recorded including the number of sniffing, pawing, and following.

5.14.15 Marbles burying

The marble burying test was used to assess repetitive and perseverative behaviors of the mouse. The chamber consisted of a housing cage filled with five centimeters of bedding on which 24 marbles were equidistantly laid down. The mouse was left in the chamber for 15 minutes. The marbles were categorized as fully covered (100%),

partially covered (75%) or not covered. An increase in the number of covered marbles indicate the presence of repetitive and pervasive behaviors.

5.14.16 Circadian activity

The circadian activity test was used to assess spontaneous activity in a full light/dark cycle. The mouse was put in individual cages (11 × 21 × 18 cm) fitted with IR captors linked to an electronic interface (Imetronic). Locomotor activity was recorded for a period of 32 hours for *Mvp::Mapk3* and *Mvp,* or 35 hours for *Mapk3* single gene mutation. It included a habituation period of 8 or 11 hours, followed by testing for 12 hours each in the dark and light phases. Feeding behavior was also evaluated using a lickometer and pellet feeder (test Diet; Hoffman La-Roche).

5.14.17 Acoustic startle response and pre-pulse inhibition

In mouse, as in humans, unexpected strong acoustic stimuli (pulse) cause startle response. This response can be attenuated when a weaker stimulus is applied right before the pulse (pre-pulse). This principle results from the integration of sensorimotor information, and is impaired in numerous disease such as autism or schizophrenia. In order to measure mouse ability to integrate these signals, the pre-pulse inhibition (PPI) paradigm consisted of exposing the mouse to randomize and repeated acoustic items and of calculating the mouse's flinching response in a chamber measuring baseline movements over a period of 50 minutes. Portion of the test in shown as example in **Figure 58**. Item A consisted of a single acoustic pulse P (60, 70, 75, 85 or 90 dB) presented during 20 milliseconds and the startle response recorded right at the end of the stimulus during 100 milliseconds. Item B comprised a combination of a pre-pulse PP (70, 75, 85 or 90 dB) 80 milliseconds before a loud pulse of 110 dB (lasting 40 milliseconds). In item B, the startle response was recorded only during the 110 dB pulse and the percentage of PPI was calculated by dividing it by the response obtained in item C, at single acoustic pulse 110 dB (ST110 for startle 110), in order to determine the inhibition of the effect that the pre-pulse stimulus exerted over the single startle response at 110 dB: %PPI(X) = PP(X)/ST110*100. Item D consisted of no stimulus (NOSTIM) trials in which only background noise (BN=60 dB) was presented to measure baseline movement of the animal in the chamber. The average response of repeated acoustic items was chosen over the maximum response and was given in an arbitrary unit by the software SR-LAB (San Diego Instruments).

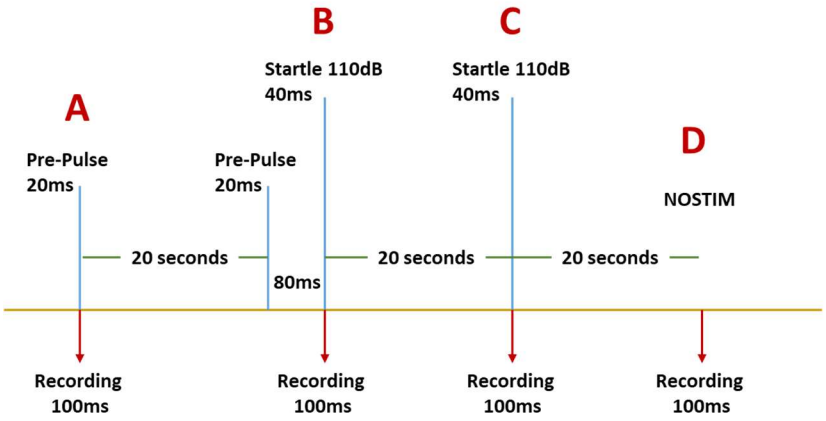


Figure 58 Acoustic startle response and pre-pulse inhibition pipeline design.

5.14.18 Statistics for behavior study

We assessed statistical significance using GraphPad Prism v8.0.2 (GraphPad Software, Inc., San Diego, CA, USA). All acquired behavioral data were tested for normality (Shapiro-Wilk normality test) and checked for outliers according to the ROUT method (with Q=1%). Outliers are labelled in red in **Datasets 3-5**. They were discarded from statistical analysis.

In cohorts consisted of two groups, data were analyzed using Student's t-test (OF, EPM, TST, PTZ, YM, NOR, SR and CA) or two-way ANOVA for repeated measures followed by Sidak's post-hoc (OF, FC, Rot, SP, Mar, CA and PPI). Student's t-tests were used to compare recognition index values to the chance level (50%) (SP, YM, NOR or SR).

The data are represented as the mean ± standard error of the mean (SEM) and the significance threshold was $p < 0.05$ unless otherwise indicated. Significant results were indicated by stars (*:p-value<0.05, ** p-value<0.01, *** p-value<0.001).

5.15 Statistics

Statistical analyses were performed with GraphPad Prism 8.0.2, using two-tailed Student's *t*-tests, Welch's *t*-tests, one-way or two-way ANOVA followed by post doc tests. Tests performed are indicated in the figure legends. Results are reported as box plots with individual data points overlaid as mean ± sample standard error of the mean, except for **Figure 2**, **Figure 45**, **Figure 48B** and **Figure 49E**, which show average for simplicity purposes. Figures were prepared using Affinity Designer. Significance was reported as follows: *$p<0.05$, **$p<0.01$ and ***$p<0.001$. All replicates are biological replicates. For qPCR data, delta C_T values were obtained by normalizing C_T values of *Mvp* against two references genes, *Gnas* and *Hprt*.

5.16 Summary of key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples	Brain samples of mutant mouse lines	**Table 3**
Chemicals, Peptides, and Recombinant Proteins		
Cresyl violet acetate	Sigma-Aldrich	Cat#C5042-106
Solvent Blue 38	Sigma-Aldrich	Cat#S3382-25G
Lithium carbonate	Sigma-Aldrich	Cat#13010-100G-R
Oxalic acid	Sigma-Aldrich	Cat#241172-50G
Deposited Data		
Assessed brain parameters	This paper	Table 4 Table 7 Table 8
Comprehensive heat map of neuroanatomical defects	This paper	Dataset 2
Software and Algorithms		
FileMakerPro version 14.0.6	FileMaker Inc.	https://www.filemaker.com/
ImageJ/Fiji version 1.51e	*Schneider* et al., 2012	http://imagej.nih.gov/ij/
Ensembl (Human GRCh38.p13 ; Mouse: GRCm38.p6)	Ensemble.org	https://www.ensembl.org/
NPDview2.0	Hamamatsu	https://www.hamamatsu.com/
Other		
Expression datasets		Table 6
Mouse whole body phenotypes	The Jackson Laboratory	http://www.informatics.jax.org/downloads/reports/index.html#pheno

6. Concluding Remarks

My doctoral work makes a significant contribution in mouse neuroanatomical phenotyping to unravel the complex genetic interactions that are at play in CNV disorders such as the 16p11.2 microdeletion syndrome (OMIM #611913). Indeed, studying endophenotypes such as neuroanatomy enabled to decipher the implication of multiple genes (*Mvp, Tbx6, Ppp4c* and *Taok2*) likely undergoing additive and synergistic epistasis.

The characterization of the double heterozygotes for *Mvp* and *Mapk3* and their single-gene intermediates, both at the neuroanatomical and behavioral level is a good example of such complex interactions. The decreased expression of *Mapk3* rescues the impact of *Mvp*-deficiency on brain size and anatomy while giving rise to modification of anxiogenic behaviors and epileptic resistance, suggesting that these two distinct core phenotypes may in fact rely on common pathways.

Previous MVP publications reported either no defects in *Mvp* mutants, or challenged the animals to see an impact on brain physiology. We were the first to observe an impact of the loss of *Mvp* in physiological conditions in the nervous system. The role of MVP in neuronal growth and brain connectivity in multiple brain regions (such as the cingulate gyrus, somatosensory cortex and hippocampus) as well as its specific expression profiling within structures associated with the limbic system raises exciting perspectives for its implication in emotion, memory and attention management.

Moreover, all these defects being only observed in males gives also a key towards the insights of sexual dimorphism present is ASD patients.

I hope the discoveries I made on the role of the major vault protein in brain morphology and physiology will be a milestone in the understanding of the function of this enigmatic organelle, which has not yet deliver its entire potential.

7. References

1. Mikhaleva, A., Kannan, M., Wagner, C. & Yalcin, B. Histomorphological Phenotyping of the Adult Mouse Brain. *Curr. Protoc. Mouse Biol.* **6**, 307–332 (2016).
2. Navarro, L. *et al.* Alteration of major vault protein in human glioblastoma and its relation with EGFR and PTEN status. *Neuroscience* **297**, 243–251 (2015).
3. Chen, Y.-L. *et al.* Mechanisms underlying lung resistance-related protein (LRP)-mediated doxorubicin resistance of non-small cell lung cancer cells. *Chin. J. Physiol.* **59**, 331–347 (2016).
4. Kolli, S., Zito, C. I., Mossink, M. H., Wiemer, E. A. C. & Bennett, A. M. The major vault protein is a novel substrate for the tyrosine phosphatase SHP-2 and scaffold protein in epidermal growth factor signaling. *J. Biol. Chem.* **279**, 29374–29385 (2004).
5. Golzio, C. *et al.* KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. *Nature* **485**, 363–367 (2012).
6. Chiurazzi, P. & Pirozzi, F. Advances in understanding - genetic basis of intellectual disability. *F1000Research* **5**, (2016).
7. Barkovich, A. J., Guerrini, R., Kuzniecky, R. I., Jackson, G. D. & Dobyns, W. B. A developmental and genetic classification for malformations of cortical development: update 2012. *Brain J. Neurol.* **135**, 1348–1369 (2012).
8. Asperger, H. Die ,,Autistisehen Psychopathen" im Kindesalter. (1943).
9. Kanner, L. Autistic disturbances of affective contact. *Nerv. Child* (1943).
10. Autism and Developmental Disabilities Monitoring Network Surveillance Year 2006 Principal Investigators, Centers for Disease Control and Prevention (CDC) & King. Prevalence of autism spectrum disorders - Autism and Developmental

Disabilities Monitoring Network, United States, 2006. *Morb. Mortal. Wkly. Rep. Surveill. Summ. Wash. DC 2002* **58**, 1–20 (2009).

11. Cluzel, S. & Compagnon, C. Stratégie nationale pour l'autisme au sein des troubles du neuro-développement. at (2018).
12. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®), Fifth Edition*. (2013).
13. Doshi-Velez, F., Ge, Y. & Kohane, I. Comorbidity clusters in autism spectrum disorders: an electronic health record time-series analysis. *Pediatrics* **133**, e54-63 (2014).
14. Volkmar, F. R. & McPartland, J. C. From Kanner to DSM-5: autism as an evolving diagnostic concept. *Annu. Rev. Clin. Psychol.* **10**, 193–212 (2014).
15. Geschwind, D. H. Advances in autism. *Annu. Rev. Med.* **60**, 367–380 (2009).
16. Baron-Cohen, S. & Belmonte, M. K. Autism: a window onto the development of the social and the analytic brain. *Annu Rev Neurosci* **28**, 109–26 (2005).
17. Baron-Cohen, S., Knickmeyer, R. C. & Belmonte, M. K. Sex differences in the brain: implications for explaining autism. *Science* **310**, 819–23 (2005).
18. Greenberg, D. M., Warrier, V., Allison, C. & Baron-Cohen, S. Testing the Empathizing-Systemizing theory of sex differences and the Extreme Male Brain theory of autism in half a million people. *Proc Natl Acad Sci U A* **115**, 12152–12157 (2018).
19. Lombardo, M. V. *et al.* Sex-specific impact of prenatal androgens on social brain default mode subsystems. *Mol. Psychiatry* (2018) doi:10.1038/s41380-018-0198-y.
20. Crider, A. & Pillai, A. Estrogen Signaling as a Therapeutic Target in Neurodevelopmental Disorders. *J. Pharmacol. Exp. Ther.* **360**, 48–58 (2017).

21. Quartier, A. *et al.* Genes and Pathways Regulated by Androgens in Human Neural Cells, Potential Candidates for the Male Excess in Autism Spectrum Disorder. *Biol. Psychiatry* **84**, 239–252 (2018).
22. Wu, H. *et al.* Phenotype-to-genotype approach reveals head-circumference-associated genes in an autism spectrum disorder cohort. *Clin. Genet.* (2019) doi:10.1111/cge.13665.
23. Broca, P. Sur la circonvolution limbique et la scissure limbique. *Bull. Mém. Société Anthropol. Paris* **12**, 646–657 (1877).
24. Papez, J. W. A proposed mechanism of emotion. 1937. *J. Neuropsychiatry Clin. Neurosci.* **7**, 103–112 (1937).
25. Botvinick, M., Nystrom, L. E., Fissell, K., Carter, C. S. & Cohen, J. D. Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature* **402**, 179–181 (1999).
26. Bush, G. *et al.* Dorsal anterior cingulate cortex: a role in reward-based decision making. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 523–528 (2002).
27. Rolls, E. T. The cingulate cortex and limbic systems for emotion, action, and memory. *Brain Struct. Funct.* **224**, 3001–3018 (2019).
28. Kohli, J. S., Kinnear, M. K., Martindale, I. A., Carper, R. A. & Muller, R.-A. Regionally decreased gyrification in middle-aged adults with autism spectrum disorders. *Neurology* **93**, e1900–e1905 (2019).
29. Kohli, J. S. *et al.* Local Cortical Gyrification is Increased in Children With Autism Spectrum Disorders, but Decreases Rapidly in Adolescents. *Cereb. Cortex N. Y. N 1991* **29**, 2412–2423 (2019).
30. Postema, M. C. *et al.* Altered structural brain asymmetry in autism spectrum disorder in a study of 54 datasets. *Nat. Commun.* **10**, 4958 (2019).

31. Edwards, T. J., Sherr, E. H., Barkovich, A. J. & Richards, L. J. Clinical, genetic and imaging findings identify new causes for corpus callosum development syndromes. *Brain J. Neurol.* **137**, 1579–1613 (2014).

32. Paul, L. K., Corsello, C., Kennedy, D. P. & Adolphs, R. Agenesis of the corpus callosum and autism: a comprehensive comparison. *Brain J. Neurol.* **137**, 1813–1829 (2014).

33. Wolff, J. J. *et al.* Altered corpus callosum morphology associated with autism over the first 2 years of life. *Brain J. Neurol.* **138**, 2046–2058 (2015).

34. Alexander, A. L. *et al.* Diffusion tensor imaging of the corpus callosum in Autism. *NeuroImage* **34**, 61–73 (2007).

35. Just, M. A., Cherkassky, V. L., Keller, T. A., Kana, R. K. & Minshew, N. J. Functional and anatomical cortical underconnectivity in autism: evidence from an FMRI study of an executive function task and corpus callosum morphometry. *Cereb. Cortex N. Y. N 1991* **17**, 951–961 (2007).

36. Lefebvre, A., Beggiato, A., Bourgeron, T. & Toro, R. Neuroanatomical Diversity of Corpus Callosum and Brain Volume in Autism: Meta-analysis, Analysis of the Autism Brain Imaging Data Exchange Project, and Simulation. *Biol. Psychiatry* **78**, 126–134 (2015).

37. Li, Q. *et al.* Decreased interhemispheric functional connectivity rather than corpus callosum volume as a potential biomarker for autism spectrum disorder. *Cortex J. Devoted Study Nerv. Syst. Behav.* **119**, 258–266 (2019).

38. Supekar, K. *et al.* Brain hyperconnectivity in children with autism and its links to social deficits. *Cell Rep.* **5**, 738–747 (2013).

39. Courchesne, E. & Pierce, K. Why the frontal cortex in autism might be talking only to itself: local over-connectivity but long-distance disconnection. *Curr. Opin. Neurobiol.* **15**, 225–230 (2005).

40. Harris, K. P. & Littleton, J. T. Transmission, Development, and Plasticity of Synapses. *Genetics* **201**, 345–375 (2015).

41. Reichelt, A. C., Rodgers, R. J. & Clapcote, S. J. The role of neurexins in schizophrenia and autistic spectrum disorder. *Neuropharmacology* **62**, 1519–1526 (2012).

42. Monteiro, P. & Feng, G. SHANK proteins: roles at the synapse and in autism spectrum disorder. *Nat. Rev. Neurosci.* **18**, 147–157 (2017).

43. Bourgeron, T. From the genetic architecture to synaptic plasticity in autism spectrum disorder. *Nat. Rev. Neurosci.* **16**, 551–563 (2015).

44. Folstein, S. & Rutter, M. Infantile autism: a genetic study of 21 twin pairs. *J. Child Psychol. Psychiatry* **18**, 297–321 (1977).

45. Wisniowiecka-Kowalnik, B. & Nowakowska, B. A. Genetics and epigenetics of autism spectrum disorder-current evidence in the field. *J. Appl. Genet.* **60**, 37–47 (2019).

46. Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019).

47. State, M. W. & Levitt, P. The conundrums of understanding genetic risks for autism spectrum disorders. *Nat. Neurosci.* **14**, 1499–1506 (2011).

48. de la Torre-Ubieta, L., Won, H., Stein, J. L. & Geschwind, D. H. Advancing the understanding of autism disease mechanisms through genetics. *Nat. Med.* **22**, 345–361 (2016).

49. Betancur, C. Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. *Brain Res.* **1380**, 42–77 (2011).

50. Iossifov, I. *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216–221 (2014).

51. Sanders, S. J. *et al.* Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* **87**, 1215–1233 (2015).

52. Pinto, D. *et al.* Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *Am. J. Hum. Genet.* **94**, 677–694 (2014).

53. Kalsner, L. & Chamberlain, S. J. Prader-Willi, Angelman, and 15q11-q13 Duplication Syndromes. *Pediatr. Clin. North Am.* **62**, 587–606 (2015).

54. Iakoucheva, L. M., Muotri, A. R. & Sebat, J. Getting to the Cores of Autism. *Cell* **178**, 1287–1298 (2019).

55. Krishnan, A. *et al.* Genome-wide prediction and functional characterization of the genetic basis of autism spectrum disorder. *Nat. Neurosci.* **19**, 1454–1462 (2016).

56. De Rubeis, S. *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209–215 (2014).

57. Marshall, C. R. *et al.* Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet* **82**, 477–88 (2008).

58. Weiss, L. A. *et al.* Association between microdeletion and microduplication at 16p11.2 and autism. *N Engl J Med* **358**, 667–75 (2008).

59. Zufferey, F. *et al.* A 600 kb deletion syndrome at 16p11.2 leads to energy imbalance and neuropsychiatric disorders. *J. Med. Genet.* **49**, 660–668 (2012).

60. Walsh, K. M. & Bracken, M. B. Copy number variation in the dosage-sensitive 16p11.2 interval accounts for only a small proportion of autism incidence: a

systematic review and meta-analysis. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **13**, 377–384 (2011).

61. Goldenberg, P. An Update on Common Chromosome Microdeletion and Microduplication Syndromes. *Pediatr. Ann.* **47**, e198–e203 (2018).
62. Rosenfeld, J. A., Coe, B. P., Eichler, E. E., Cuckle, H. & Shaffer, L. G. Estimates of penetrance for recurrent pathogenic copy-number variations. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **15**, 478–481 (2013).
63. Walters, R. G. *et al.* A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. *Nature* **463**, 671–675 (2010).
64. Maillard, A. M. *et al.* 16p11.2 Locus modulates response to satiety before the onset of obesity. *Int. J. Obes. 2005* **40**, 870–876 (2016).
65. Niarchou, M. *et al.* Psychiatric disorders in children with 16p11.2 deletion and duplication. *Transl. Psychiatry* **9**, 8 (2019).
66. Qureshi, A. Y. *et al.* Opposing brain differences in 16p11.2 deletion and duplication carriers. *J. Neurosci. Off. J. Soc. Neurosci.* **34**, 11199–11211 (2014).
67. Owen, J. P. *et al.* Brain MR Imaging Findings and Associated Outcomes in Carriers of the Reciprocal Copy Number Variation at 16p11.2. *Radiology* **286**, 217–226 (2017).
68. Martin-Brevet, S. *et al.* Quantifying the Effects of 16p11.2 Copy Number Variants on Brain Structure: A Multisite Genetic-First Study. *Biol. Psychiatry* **84**, 253–264 (2018).
69. Cárdenas-de-la-Parra, A. *et al.* Developmental trajectories of neuroanatomical alterations associated with the 16p11.2 Copy Number Variations. *NeuroImage* **203**, 116155 (2019).

70. Blumenthal, I. *et al.* Transcriptional consequences of 16p11.2 deletion and duplication in mouse cortex and multiplex autism families. *Am. J. Hum. Genet.* **94**, 870–883 (2014).

71. Luo, R. *et al.* Genome-wide transcriptome profiling reveals the functional impact of rare de novo and recurrent CNVs in autism spectrum disorders. *Am. J. Hum. Genet.* **91**, 38–55 (2012).

72. Portmann, T. *et al.* Behavioral abnormalities and circuit defects in the basal ganglia of a mouse model of 16p11.2 deletion syndrome. *Cell Rep.* **7**, 1077–1092 (2014).

73. Horev, G. *et al.* Dosage-dependent phenotypes in models of 16p11.2 lesions found in autism. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 17076–17081 (2011).

74. Arbogast, T. *et al.* Reciprocal Effects on Neurocognitive and Metabolic Phenotypes in Mouse Models of 16p11.2 Deletion and Duplication Syndromes. *PLoS Genet.* **12**, e1005709 (2016).

75. Pucilowska, J. *et al.* The 16p11.2 deletion mouse model of autism exhibits altered cortical progenitor proliferation and brain cytoarchitecture linked to the ERK MAPK pathway. *J. Neurosci. Off. J. Soc. Neurosci.* **35**, 3190–3200 (2015).

76. Yang, M. *et al.* 16p11.2 Deletion Syndrome Mice Display Sensory and Ultrasonic Vocalization Deficits During Social Interactions. *Autism Res. Off. J. Int. Soc. Autism Res.* **8**, 507–521 (2015).

77. Li, M., Liu, J. & Zhang, C. Evolutionary history of the vertebrate mitogen activated protein kinases family. *PloS One* **6**, e26999 (2011).

78. Roskoski, R. ERK1/2 MAP kinases: structure, function, and regulation. *Pharmacol. Res.* **66**, 105–143 (2012).

79. Satoh, Y. *et al.* Extracellular signal-regulated kinase 2 (ERK2) knockdown mice show deficits in long-term memory; ERK2 has a specific function in learning and memory. *J. Neurosci. Off. J. Soc. Neurosci.* **27**, 10765–10776 (2007).

80. Satoh, Y. *et al.* ERK2 contributes to the control of social behaviors in mice. *J. Neurosci. Off. J. Soc. Neurosci.* **31**, 11953–11967 (2011).

81. Samuels, I. S. *et al.* Deletion of ERK2 mitogen-activated protein kinase identifies its key roles in cortical neurogenesis and cognitive function. *J. Neurosci. Off. J. Soc. Neurosci.* **28**, 6983–6995 (2008).

82. Newbern, J. *et al.* Mouse and human phenotypes indicate a critical conserved role for ERK2 signaling in neural crest development. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 17115–17120 (2008).

83. Mazzucchelli, C. *et al.* Knockout of ERK1 MAP kinase enhances synaptic plasticity in the striatum and facilitates striatal-mediated learning and memory. *Neuron* **34**, 807–820 (2002).

84. Silingardi, D. *et al.* ERK pathway activation bidirectionally affects visual recognition memory and synaptic plasticity in the perirhinal cortex. *Front. Behav. Neurosci.* **5**, 84 (2011).

85. Selcher, J. C., Nekrasova, T., Paylor, R., Landreth, G. E. & Sweatt, J. D. Mice lacking the ERK1 isoform of MAP kinase are unimpaired in emotional learning. *Learn. Mem. Cold Spring Harb. N* **8**, 11–19 (2001).

86. Bost, F. *et al.* The extracellular signal-regulated kinase isoform ERK1 is specifically required for in vitro and in vivo adipogenesis. *Diabetes* **54**, 402–411 (2005).

87. Pagès, G. *et al.* Defective thymocyte maturation in p44 MAP kinase (Erk 1) knockout mice. *Science* **286**, 1374–1377 (1999).

88. Buscà, R., Pouysségur, J. & Lenormand, P. ERK1 and ERK2 Map Kinases: Specific Roles or Functional Redundancy? *Front. Cell Dev. Biol.* **4**, 53 (2016).

89. Lefloch, R., Pouysségur, J. & Lenormand, P. Total ERK1/2 activity regulates cell proliferation. *Cell Cycle Georget. Tex* **8**, 705–711 (2009).

90. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).

91. Dérijard, B. *et al.* Independent human MAP-kinase signal transduction pathways defined by MEK and MKK isoforms. *Science* **267**, 682–685 (1995).

92. von Kriegsheim, A. *et al.* Cell fate decisions are specified by the dynamic ERK interactome. *Nat. Cell Biol.* **11**, 1458–1464 (2009).

93. Samuels, I. S., Saitta, S. C. & Landreth, G. E. MAP'ing CNS development and cognition: an ERKsome process. *Neuron* **61**, 160–167 (2009).

94. Rauen, K. A. The RASopathies. *Annu. Rev. Genomics Hum. Genet.* **14**, 355–369 (2013).

95. Nateri, A. S. *et al.* ERK activation causes epilepsy by stimulating NMDA receptor activity. *EMBO J.* **26**, 4891–4901 (2007).

96. Lawrence, M. C. *et al.* Chromatin-bound mitogen-activated protein kinases transmit dynamic signals in transcription complexes in beta-cells. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 13315–13320 (2008).

97. Frémin, C., Saba-El-Leil, M. K., Lévesque, K., Ang, S.-L. & Meloche, S. Functional Redundancy of ERK1 and ERK2 MAP Kinases during Development. *Cell Rep.* **12**, 913–921 (2015).

98. Paul, R., Cohen, D. J. & Volkmar, F. R. Autistic behaviors in a boy with Noonan syndrome. *J. Autism Dev. Disord.* **13**, 433–434 (1983).

99. Vithayathil, J., Pucilowska, J. & Landreth, G. E. ERK/MAPK signaling and autism spectrum disorders. *Prog. Brain Res.* **241**, 63–112 (2018).

100. Komiyama, N. H. *et al.* SynGAP regulates ERK/MAPK signaling, synaptic plasticity, and learning in the complex with postsynaptic density 95 and NMDA receptor. *J. Neurosci. Off. J. Soc. Neurosci.* **22**, 9721–9732 (2002).

101. Davis, S. & Laroche, S. Mitogen-activated protein kinase/extracellular regulated kinase signalling and memory stabilization: a review. *Genes Brain Behav.* **5 Suppl 2**, 61–72 (2006).

102. Adviento, B. *et al.* Autism traits in the RASopathies. *J. Med. Genet.* **51**, 10–20 (2014).

103. Thomas, G. M. & Huganir, R. L. MAPK cascade signalling and synaptic plasticity. *Nat. Rev. Neurosci.* **5**, 173–183 (2004).

104. Fasano, S. & Brambilla, R. Ras-ERK Signaling in Behavior: Old Questions and New Perspectives. *Front. Behav. Neurosci.* **5**, 79 (2011).

105. Cui, Y. *et al.* Neurofibromin regulation of ERK signaling modulates GABA release and learning. *Cell* **135**, 549–560 (2008).

106. Mainberger, F. *et al.* Lovastatin improves impaired synaptic plasticity and phasic alertness in patients with neurofibromatosis type 1. *BMC Neurol.* **13**, 131 (2013).

107. Pucilowska, J. *et al.* Pharmacological Inhibition of ERK Signaling Rescues Pathophysiology and Behavioral Phenotype Associated with 16p11.2 Chromosomal Deletion in Mice. *J. Neurosci. Off. J. Soc. Neurosci.* **38**, 6640–6652 (2018).

108. Morè, L., Lauterborn, J. C., Papaleo, F. & Brambilla, R. Enhancing cognition through pharmacological and environmental interventions: Examples from

preclinical models of neurodevelopmental disorders. *Neurosci. Biobehav. Rev.* **110**, 28–45 (2020).

109. Kedersha, N. & Rome, L. H. Isolation and characterization of a novel ribonucleoprotein particle: large structures contain a single species of small RNA. *J. Cell Biol.* **103**, 699–709 (1986).

110. Scheper, R. J. *et al.* Overexpression of a M(r) 110,000 vesicular protein in non-P-glycoprotein-mediated multidrug resistance. *Cancer Res.* **53**, 1475–1479 (1993).

111. Scheffer, G. L. *et al.* The drug resistance-related protein LRP is the human major vault protein. *Nat. Med.* **1**, 578–582 (1995).

112. Kickhoefer, V. A., Stephen, A. G., Harrington, L., Robinson, M. O. & Rome, L. H. Vaults and Telomerase Share a Common Subunit, TEP1. *J. Biol. Chem.* **274**, 32712–32717 (1999).

113. Kickhoefer, V. A. *et al.* The 193-kD vault protein, VPARP, is a novel poly(ADP-ribose) polymerase. *J. Cell Biol.* **146**, 917–928 (1999).

114. Kickhoefer, V. A. *et al.* Vault ribonucleoprotein particles from rat and bullfrog contain a related small RNA that is transcribed by RNA polymerase III. *J. Biol. Chem.* **268**, 7868–7873 (1993).

115. Mossink, M. H. *et al.* Disruption of the murine major vault protein (MVP/LRP) gene does not induce hypersensitivity to cytostatics. *Cancer Res.* **62**, 7298–7304 (2002).

116. Paspalas, C. D. *et al.* Major vault protein is expressed along the nucleus-neurite axis and associates with mRNAs in cortical neurons. *Cereb. Cortex N. Y. N 1991* **19**, 1666–1677 (2009).

117. Rome, L. H. & Kickhoefer, V. A. Development of the Vault Particle as a Platform Technology. *ACS Nano* **7**, 889–902 (2013).

118. Kedersha, N. L., Miquel, M. C., Bittner, D. & Rome, L. H. Vaults. II. Ribonucleoprotein structures are highly conserved among higher and lower eukaryotes. *J. Cell Biol.* **110**, 895–901 (1990).

119. Daly, T. K., Sutherland-Smith, A. J. & Penny, D. In silico resurrection of the major vault protein suggests it is ancestral in modern eukaryotes. *Genome Biol. Evol.* **5**, 1567–1583 (2013).

120. Tanaka, H. *et al.* The structure of rat liver vault at 3.5 angstrom resolution. *Science* **323**, 384–388 (2009).

121. Kedersha, N. L., Heuser, J. E., Chugani, D. C. & Rome, L. H. Vaults. III. Vault ribonucleoprotein particles open into flower-like structures with octagonal symmetry. *J. Cell Biol.* **112**, 225–235 (1991).

122. Esfandiary, R., Kickhoefer, V. A., Rome, L. H., Joshi, S. B. & Middaugh, C. R. Structural stability of vault particles. *J. Pharm. Sci.* **98**, 1376–1386 (2009).

123. Poderycki, M. J. *et al.* The vault exterior shell is a dynamic structure that allows incorporation of vault-associated proteins into its interior. *Biochemistry* **45**, 12184–12193 (2006).

124. Llauró, A. *et al.* Mechanical stability and reversible fracture of vault particles. *Biophys. J.* **106**, 687–695 (2014).

125. Stephen, A. G. *et al.* Assembly of vault-like particles in insect cells expressing only the major vault protein. *J. Biol. Chem.* **276**, 23217–23220 (2001).

126. van Zon, A. *et al.* Multiple human vault RNAs. Expression and association with the vault complex. *J. Biol. Chem.* **276**, 37715–37721 (2001).

127. Tanaka, H. & Tsukihara, T. Structural studies of large nucleoprotein particles, vaults. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* **88**, 416–433 (2012).

128. Lange, C., Walther, W., Schwabe, H. & Stein, U. Cloning and Initial Analysis of the Human Multidrug Resistance-Related MVP/LRP Gene Promoter. *Biochem. Biophys. Res. Commun.* **278**, 125–133 (2000).

129. Asakuno, K. *et al.* Involvement of a DNA binding protein, MDR-NF1/YB-1, in human MDR1 gene expression by actinomycin D. *Biochem. Biophys. Res. Commun.* **199**, 1428–1435 (1994).

130. Mrazek, J. *et al.* Polyribosomes are molecular 3D nanoprinters that orchestrate the assembly of vault particles. *ACS Nano* **8**, 11552–11559 (2014).

131. Bateman, A. & Kickhoefer, V. The TROVE module: a common element in Telomerase, Ro and Vault ribonucleoproteins. *BMC Bioinformatics* **4**, 49 (2003).

132. Zheng, C.-L. *et al.* Characterization of MVP and VPARP assembly into vault ribonucleoprotein complexes. *Biochem. Biophys. Res. Commun.* **326**, 100–107 (2005).

133. Sutovsky, P. *et al.* Expression and proteasomal degradation of the major vault protein (MVP) in mammalian oocytes and zygotes. *Reprod. Camb. Engl.* **129**, 269–282 (2005).

134. Scheffer, G. L., Schroeijers, A. B., Izquierdo, M. A., Wiemer, E. A. C. & Scheper, R. J. Lung resistance-related protein/major vault protein and vaults in multidrug-resistant cancer. *Curr. Opin. Oncol.* **12**, 550–556 (2000).

135. Chugani, D. C., Kedersha, N. L. & Rome, L. H. Vault immunofluorescence in the brain: new insights regarding the origin of microglia. *J. Neurosci. Off. J. Soc. Neurosci.* **11**, 256–268 (1991).

136. Suprenant, K. A., Bloom, N., Fang, J. & Lushington, G. The major vault protein is related to the toxic anion resistance protein (TelA) family. *J. Exp. Biol.* **210**, 946–955 (2007).

137. Herrmann, C., Golkaramnay, E., Inman, E., Rome, L. & Volknandt, W. Recombinant major vault protein is targeted to neuritic tips of PC12 cells. *J. Cell Biol.* **144**, 1163–1172 (1999).

138. Eichenmüller, B. *et al.* Vaults bind directly to microtubules via their caps and not their barrels. *Cell Motil. Cytoskeleton* **56**, 225–236 (2003).

139. Li, J. Y. *et al.* Axonal transport of ribonucleoprotein particles (vaults). *Neuroscience* **91**, 1055–1065 (1999).

140. Slesina, M., Inman, E. M., Rome, L. H. & Volknandt, W. Nuclear localization of the major vault protein in U373 cells. *Cell Tissue Res.* **321**, 97–104 (2005).

141. Kickhoefer, V. A. *et al.* The Telomerase/vault-associated protein TEP1 is required for vault RNA stability and its association with the vault particle. *J. Cell Biol.* **152**, 157–164 (2001).

142. Liu, Y. *et al.* Vault poly(ADP-ribose) polymerase is associated with mammalian telomerase and is dispensable for telomerase function and vault structure in vivo. *Mol. Cell. Biol.* **24**, 5314–5323 (2004).

143. Berger, W. *et al.* Overexpression of the human major vault protein in astrocytic brain tumor cells. *Int. J. Cancer* **94**, 377–382 (2001).

144. Ryu, S. J. & Park, S. C. Targeting major vault protein in senescence-associated apoptosis resistance. *Expert Opin. Ther. Targets* **13**, 479–484 (2009).

145. Nandy, C. *et al.* Epstein-barr virus-induced expression of a novel human vault RNA. *J. Mol. Biol.* **388**, 776–784 (2009).

146. Mossink, M. H., van Zon, A., Scheper, R. J., Sonneveld, P. & Wiemer, E. A. C. Vaults: a ribonucleoprotein particle involved in drug resistance? *Oncogene* **22**, 7458–7467 (2003).

147. Ip, J. P. K. *et al.* Major Vault Protein, a Candidate Gene in 16p11.2 Microdeletion Syndrome, Is Required for the Homeostatic Regulation of Visual Cortical Plasticity. *J. Neurosci. Off. J. Soc. Neurosci.* **38**, 3890–3900 (2018).

148. Berger, W., Steiner, E., Grusch, M., Elbling, L. & Micksche, M. Vaults and the major vault protein: novel roles in signal pathway regulation and immunity. *Cell. Mol. Life Sci. CMLS* **66**, 43–61 (2009).

149. Abbondanza, C. *et al.* Interaction of vault particles with estrogen receptor in the MCF-7 breast cancer cell. *J. Cell Biol.* **141**, 1301–1310 (1998).

150. Xiao, Y.-S. *et al.* Major vault protein is a direct target of Notch1 signaling and contributes to chemoresistance in triple-negative breast cancer cells. *Cancer Lett.* **440–441**, 156–167 (2019).

151. Losert, A. *et al.* The major vault protein mediates resistance to epidermal growth factor receptor inhibition in human hepatoma cells. *Cancer Lett.* **319**, 164–172 (2012).

152. Herlevsen, M., Oxford, G., Owens, C. R., Conaway, M. & Theodorescu, D. Depletion of major vault protein increases doxorubicin sensitivity and nuclear accumulation and disrupts its sequestration in lysosomes. *Mol. Cancer Ther.* **6**, 1804–1813 (2007).

153. Suprenant, K. A. Vault ribonucleoprotein particles: sarcophagi, gondolas, or safety deposit boxes? *Biochemistry* **41**, 14447–14454 (2002).

154. Noguchi, T., Matozaki, T., Horita, K., Fujioka, Y. & Kasuga, M. Role of SH-PTP2, a protein-tyrosine phosphatase with Src homology 2 domains, in insulin-stimulated Ras activation. *Mol. Cell. Biol.* **14**, 6674–6682 (1994).

155. Liang, P. *et al.* MVP interacts with YPEL4 and inhibits YPEL4-mediated activities of the ERK signal pathway. *Biochem. Cell Biol. Biochim. Biol. Cell.* **88**, 445–450 (2010).

156. Vasu, S. K. & Rome, L. H. Dictyostelium vaults: disruption of the major proteins reveals growth and morphological defects and uncovers a new associated protein. *J. Biol. Chem.* **270**, 16588–16594 (1995).

157. Northoff, G., Wainio-Theberge, S. & Evers, K. Is temporo-spatial dynamics the 'common currency' of brain and mind? In Quest of 'Spatiotemporal Neuroscience'. *Phys. Life Rev.* (2019) doi:10.1016/j.plrev.2019.05.002.

158. Collins, S. C. *et al.* A Method for Parasagittal Sectioning for Neuroanatomical Quantification of Brain Structures in the Adult Mouse. *Curr. Protoc. Mouse Biol.* **8**, e48 (2018).

159. Kannan, M. *et al.* WD40-repeat 47, a microtubule-associated protein, is essential for brain development and autophagy. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E9308–E9317 (2017).

160. Loomes, R., Hull, L. & Mandy, W. P. L. What Is the Male-to-Female Ratio in Autism Spectrum Disorder? A Systematic Review and Meta-Analysis. *J. Am. Acad. Child Adolesc. Psychiatry* **56**, 466–474 (2017).

161. Arbogast, T. *et al.* Kctd13-deficient mice display short-term memory impairment and sex-dependent genetic interactions. *Hum. Mol. Genet.* **28**, 1474–1486 (2019).

162. Escamilla, C. O. *et al.* Kctd13 deletion reduces synaptic transmission via increased RhoA. *Nature* **551**, 227–231 (2017).

163. Buehler, D. C. *et al.* Bioengineered vaults: self-assembling protein shell-lipophilic core nanoparticles for drug delivery. *ACS Nano* **8**, 7723–7732 (2014).

164. Kim, E. *et al.* Crosstalk between Src and major vault protein in epidermal growth factor-dependent cell signalling. *FEBS J.* **273**, 793–804 (2006).

165. Collins, S. C. *et al.* Large-scale neuroanatomical study uncovers 198 gene associations in mouse brain morphogenesis. *Nat. Commun.* **10**, 3465 (2019).

166. Iyer, J. *et al.* Pervasive genetic interactions modulate neurodevelopmental defects of the autism-associated 16p11.2 deletion in Drosophila melanogaster. *Nat. Commun.* **9**, 2548 (2018).

167. McCammon, J. M., Blaker-Lee, A., Chen, X. & Sive, H. The 16p11.2 homologs fam57ba and doc2a generate certain brain and body phenotypes. *Hum. Mol. Genet.* **26**, 3699–3712 (2017).

168. White, J. K. *et al.* Genome-wide Generation and Systematic Phenotyping of Knockout Mice Reveals New Roles for Many Genes. *Cell* **154**, 452–464 (2013).

169. Du, Q., de la Morena, M. T. & van Oers, N. S. C. The Genetics and Epigenetics of 22q11.2 Deletion Syndrome. *Front. Genet.* **10**, 1365 (2019).

170. Golzio, C. & Katsanis, N. Genetic Architecture of Reciprocal CNVs. *Curr. Opin. Genet. Dev.* **23**, 240–248 (2013).

171. Shinawi, M. *et al.* Recurrent reciprocal 16p11.2 rearrangements associated with global developmental delay, behavioural problems, dysmorphism, epilepsy, and abnormal head size. *J. Med. Genet.* **47**, 332–341 (2010).

172. Richter, M. *et al.* Altered TAOK2 activity causes autism-related neurodevelopmental and cognitive abnormalities through RhoA signaling. *Mol. Psychiatry* **24**, 1329–1350 (2019).

173. Michetti, C. *et al.* The PRRT2 knockout mouse recapitulates the neurological diseases associated with PRRT2 mutations. *Neurobiol. Dis.* **99**, 66–83 (2017).

174. Pessoa, L. A Network Model of the Emotional Brain. *Trends Cogn. Sci.* **21**, 357–371 (2017).

175. Herrmann, C., Volknandt, W., Wittich, B., Kellner, R. & Zimmermann, H. The major vault protein (MVP100) is contained in cholinergic nerve terminals of electric ray electric organ. *J. Biol. Chem.* **271**, 13908–13915 (1996).

176. Goebel-Goody, S. M. *et al.* Therapeutic implications for striatal-enriched protein tyrosine phosphatase (STEP) in neuropsychiatric disorders. *Pharmacol. Rev.* **64**, 65–87 (2012).

177. Jacquemont, S. *et al.* A higher mutational burden in females supports a 'female protective model' in neurodevelopmental disorders. *Am. J. Hum. Genet.* **94**, 415–425 (2014).

178. Baron-Cohen, S. *et al.* Why are autism spectrum conditions more prevalent in males? *PLoS Biol.* **9**, e1001081 (2011).

179. Kumar, V. J. *et al.* Linking spatial gene expression patterns to sex-specific brain structural changes on a mouse model of 16p11.2 hemideletion. *Transl. Psychiatry* **8**, 109 (2018).

180. Ransome, M. I. & Boon, W. C. Testosterone-induced adult neurosphere growth is mediated by sexually-dimorphic aromatase expression. *Front Cell Neurosci* **9**, 253 (2015).

181. Grissom, N. M. *et al.* Male-specific deficits in natural reward learning in a mouse model of neurodevelopmental disorders. *Mol. Psychiatry* **23**, 544–555 (2018).

182. Crepel, A. *et al.* Narrowing the critical deletion region for autism spectrum disorders on 16p11.2. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet. Off. Publ. Int. Soc. Psychiatr. Genet.* **156**, 243–245 (2011).

183. Mulkey, S. B. & du Plessis, A. J. Autonomic nervous system development and its impact on neuropsychiatric outcome. *Pediatr. Res.* **85**, 120–126 (2019).

184. Fillinger, C., Yalcin, I., Barrot, M. & Veinante, P. Afferents to anterior cingulate areas 24a and 24b and midcingulate areas 24a' and 24b' in the mouse. *Brain Struct. Funct.* **222**, 1509–1532 (2017).

185. Fillinger, C., Yalcin, I., Barrot, M. & Veinante, P. Efferents of anterior cingulate areas 24a and 24b and midcingulate areas 24a' and 24b' in the mouse. *Brain Struct. Funct.* **223**, 1747–1778 (2018).

186. Tsanov, M. Speed and Oscillations: Medial Septum Integration of Attention and Navigation. *Front. Syst. Neurosci.* **11**, 67 (2017).

187. Onodera, S. & Hicks, T. P. Projections from substantia nigra and zona incerta to the cat's nucleus of Darkschewitsch. *J. Comp. Neurol.* **396**, 461–482 (1998).

188. Gordon-Fennell, A. G. *et al.* The Lateral Preoptic Area: A Novel Regulator of Reward Seeking and Neuronal Activity in the Ventral Tegmental Area. *Front. Neurosci.* **13**, 1433 (2019).

189. Balaban, C. D. Neurotransmitters in the vestibular system. *Handb. Clin. Neurol.* **137**, 41–55 (2016).

190. Jiang, Z., Rajamanickam, S. & Justice, N. J. CRF signaling between neurons in the paraventricular nucleus of the hypothalamus (PVN) coordinates stress responses. *Neurobiol. Stress* **11**, 100192 (2019).

191. Teissier, A., Soiza-Reilly, M. & Gaspar, P. Refining the Role of 5-HT in Postnatal Development of Brain Circuits. *Front. Cell. Neurosci.* **11**, 139 (2017).

192. Milczarek, M. M. & Vann, S. D. The retrosplenial cortex and long-term spatial memory: from the cell to the network. *Curr. Opin. Behav. Sci.* **32**, 50–56 (2020).

193. Sperlágh, B., Maglóczky, Z., Vizi, E. S. & Freund, T. F. The triangular septal nucleus as the major source of ATP release in the rat habenula: a combined neurochemical and morphological study. *Neuroscience* **86**, 1195–1207 (1998).

194. Yamaguchi, T., Danjo, T., Pastan, I., Hikida, T. & Nakanishi, S. Distinct roles of segregated transmission of the septo-habenular pathway in anxiety and fear. *Neuron* **78**, 537–544 (2013).

195. Jung, M. W., Wiener, S. I. & McNaughton, B. L. Comparison of spatial firing characteristics of units in dorsal and ventral hippocampus of the rat. *J. Neurosci. Off. J. Soc. Neurosci.* **14**, 7347–7356 (1994).

196. Salay, L. D., Ishiko, N. & Huberman, A. D. A midline thalamic circuit determines reactions to visual threat. *Nature* **557**, 183–189 (2018).

197. Zhou, M. *et al.* A central amygdala to zona incerta projection is required for acquisition and remote recall of conditioned fear memory. *Nat. Neurosci.* **21**, 1515–1519 (2018).

198. Wang, X., Chou, X.-L., Zhang, L. I. & Tao, H. W. Zona Incerta: An Integrative Node for Global Behavioral Modulation. *Trends Neurosci.* **43**, 82–87 (2020).

199. Paxinos, G. & Franklin. *The Mouse Brain in Stereotaxic Coordinates. 3rd ed.* (Academic Press, San Diego, 2007).

200. Giannuzzi, G. *et al.* The Human-Specific BOLA2 Duplication Modifies Iron Homeostasis and Anemia Predisposition in Chromosome 16p11.2 Autism Individuals. *Am. J. Hum. Genet.* **105**, 947–958 (2019).

201. Mueller, P. *et al.* Regulation of T cell survival through coronin-1-mediated generation of inositol-1,4,5-trisphosphate and calcium mobilization after T cell receptor triggering. *Nat. Immunol.* **9**, 424–431 (2008).

202. Chapman, D. L. & Papaioannou, V. E. Three neural tubes in mouse embryos with mutations in the T-box gene Tbx6. *Nature* **391**, 695–697 (1998).

203. Sakaguchi, G. *et al.* Doc2alpha is an activity-dependent modulator of excitatory synaptic transmission. *Eur. J. Neurosci.* **11**, 4262–4268 (1999).

204. Terakata, M. *et al.* Establishment of true niacin deficiency in quinolinic acid phosphoribosyltransferase knockout mice. *J. Nutr.* **142**, 2148–2153 (2012).

205. Skarnes, W. C. *et al.* A conditional knockout resource for the genome-wide study of mouse gene function. *Nature* **474**, 337–342 (2011).

206. Boroviak, K., Doe, B., Banerjee, R., Yang, F. & Bradley, A. Chromosome engineering in zygotes with CRISPR/Cas9. *Genes. N. Y. N 2000* **54**, 78–85 (2016).

207. Karp, N. A. *et al.* Applying the ARRIVE Guidelines to an In Vivo Database. *PLoS Biol.* **13**, e1002151 (2015).

208. Kechin, A., Boyarskikh, U., Kel, A. & Filipenko, M. cutPrimers: A New Tool for Accurate Cutting of Primers from Reads of Targeted Next Generation Sequencing. *J. Comput. Biol. J. Comput. Mol. Cell Biol.* **24**, 1138–1143 (2017).

209. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

210. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinforma. Oxf. Engl.* **29**, 15–21 (2013).

211. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinforma. Oxf. Engl.* **31**, 166–169 (2015).

212. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

Université de Strasbourg

Perrine KRETZ

École doctorale Sciences de la vie et de la santé | ED 414
Université de Strasbourg

Analyses des causes génétiques du syndrome de micro-délétion du 16p11.2 et de l'impact de la protéine majeure de la voute (MVP) et de son interaction avec MAPK3 dans la physiologie cérébrale

Résumé

En utilisant des modèles génétiques murins, nous avons cherché à identifier lequel des 30 gènes du locus 16p11.2, associé à l'autisme, provoque des phénotypes neuroanatomiques. Nous montrons, contrairement aux études précédentes, que plusieurs gènes cartographiés dans cette région interagissent pour réguler la taille du cerveau et que les souris femelles présentent beaucoup moins de phénotypes. La protéine majeure de la voûte (MVP) est une protéine hautement conservée présente dans les cellules eucaryotes dont la fonction n'est toujours pas comprise. Dans cette étude, nous montrons que Mvp est le principal gène responsable des phénotypes neuroanatomiques, et qu'il régule la morphologie des neurones, après la naissance et spécifiquement chez les mâles. Nous démontrons également que la double délétion Mvp::Mapk3 restore le phénotype, suggérant que MVP et ERK sont impliqués dans la même voie de signalisation, par un potentiel rétrocontrôle négatif exercé par MVP sur ERK. Nos résultats fournissent la première preuve de l'implication de la voûte dans la régulation de la taille du cerveau des mammifères et des structures limbiques.

Mots clés: Génétique de la souris, troubles du spectre autistique, anatomie cérébrale, protéine majeure de la voute

Résumé en anglais

Using mouse genetic studies, we set out to identify which of the 30 genes causes brain size and neuroanatomical phenotypes at the autism-associated 16p11.2 locus. We show that multiple genes mapping to this region interact to regulate brain size in contrast to previous studies, with female mice exhibiting far fewer neuroanatomical phenotypes. The major vault protein (MVP), the main component of the vault organelle, is a highly conserved protein found in eukaryotic cells, yet its function is not understood. Here, we find MVP expression specific to the limbic system and show that Mvp is the top driver of neuroanatomical phenotypes, regulating the morphology of neurons, postnatally and specifically in male. We also demonstrate that the double deletion Mvp::Mapk3 rescues the brain size phenotype of Mvp-deficient male mice, suggesting that MVP and ERK are involved in the same signaling pathway in vivo with MVP possibly acting as an upstream regulator for ERK signaling controlling brain size. Our results provide the first evidence for the involvement of the vault organelle in the regulation of the mammalian brain size and limbic structures.

Keywords: Mouse genetic studies, autism spectrum disorders, brain anatomy, major vault protein