UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE DES SCENCES DE LA VIE ET DE LA SANTE

Institut de Génétique et de Biologie Moléculaire et Cellulaire

THÈSE présentée par :

Alexandre HALLER

soutenue le : **02 Septembre 2024**

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

Discipline : Sciences de la Vie et de la Santé

Spécialité : Aspects moléculaires et cellulaires de la biologie

Characterization of long non-coding RNA LENT, a potential therapeutic target in cutaneous melanoma

THÈSE dirigée par :

Mr. DAVIDSON Irwin	Directeur de Recherche CNRS, IGBMC, Illkirch

RAPPORTEURS :

Mme. BERTOLOTTO Corine	Directrice de Recherche INSERM, C3M, Nice
Mr. LE CAM Laurent	Directeur de Recherche INSERM, IRCM, Montpellier

AUTRES MEMBRES DU JURY :

Mme. LEUCCI Eleonora	Associate Professor, Katholieke Universiteit, Leuven
Mr. COIN Frédéric	Directeur de Recherche INSERM, IGBMC, Illkirch

« *Chaque enfant qu'on enseigne est un homme qu'on gagne* »

Victor Hugo

« *A vaincre sans péril, on triomphe sans gloire* »

Pierre Corneille

Remerciements

Je voudrais prendre le temps ici de remercier l'intégralité des personnes sans qui je n'aurais pas pu accomplir tout ce travail. La thèse a été une période de ma vie enrichissante, de par les rencontres que j'ai pu faire au fil des années, autant du point de vue scientifique qu'humain.

Je tiens tout d'abord à remercier les membres de mon jury d'avoir accepté d'évaluer mon travail de thèse : Dr **Laurent Le Cam**, Dr **Corine Bertolotto**, Dr **Eleonora Leucci** et Dr **Frédéric Coin**. J'ai également une pensée pour Pr **Marie-Dominique Galibert** et Dr **Christophe Romier**, les membres de mon comité de suivi de thèse, pour m'avoir aiguillé sur mon projet chaque année et pour leur bienveillance.

Evidemment, je tiens à sincèrement remercier mon directeur de thèse, **Irwin**, pour sa confiance et son aide tout au long de ma thèse. Tu as toujours été disponible pour moi lorsque j'en avais besoin tout en me laissant ma part d'indépendance. C'est un honneur pour moi d'avoir travaillé à tes côtés, ton niveau scientifique reste l'un des plus impressionnant que j'ai pu apprécier durant ma jeune carrière.

Je souhaite également remercier mon encadrant de master 2, **Partyk Ngondo**, qui m'a inculqué une rigueur de travail scientifique et m'a poussé à poursuivre mes études en thèse. Même si nos chemins se sont séparés après mon master, tu restes l'une des meilleures rencontres que j'ai pu faire dans le monde scientifique.

Un grand merci à l'ensemble des membres de l'équipe avec qui j'ai traversé ces quatre années, votre aide a toujours été précieuse : **Isabelle, Gabrielle, Guillaume, Igor, Justine, Philippe, Léane.** Et évidemment **Antonin**, ma thèse aurait été bien différente sans toi, tu sais à quel point je t'apprécie (mais pas trop quand même) et je te souhaite le meilleur pour la fin de ta thèse, tu es capable d'accomplir de belles choses. Je me demande toujours qui a mis le mot sur la fontaine à eau… Une pensée pour les anciens du labo, ce sont eux qui m'ont appris le plus de choses : **Bujamin,**

Pietro, Alexandra et plus particulièrement **Giovanni** pour m'avoir encadré et continué à m'apporter son soutien même en étant un post-doctorant très occupé. Un grand merci pour tous les compagnons de galère de l'institut, c'était toujours un plaisir de vous voir au cours d'un café dans les canapés du quatrième étage, surtout la team **Julian**, **Anthony** et **Julie** que j'avais tendance à étrangement croiser plus souvent que les autres. Ne lâchez rien, vous êtes bientôt au bout (sauf Anthony).

Il est aussi important de mentionner les services communs de l'institut sans qui je n'aurais pas été capable d'aller aussi loin dans ce beau projet, nous avons la chance de les avoir et il est essentiel de la reconnaitre. Merci également à nos collaborateurs de l'IBMC pour leur expertise et leur sympathie : Dr **Eric Ennifar**, Dr **Angelita Simonetti** et Dr **Mattia D'Agostino.** Un grand merci aussi aux organismes et aux donateurs qui nous permettent de réaliser notre travail : IMCbio pour mon financement de thèse, la Fondation pour la Recherche Médicale pour ma quatrième année, la ligue contre le cancer qui labélise notre équipe et la SATT Connectus pour leur investissement dans notre projet.

Et enfin un grand remerciement à l'ensemble de mes proches. Les amis pour les sorties qui m'ont redonné du baume au cœur quand j'en avais besoin. Mes parents pour leur éternel soutien, vous m'avez toujours encouragé à faire de mon mieux et m'avez toujours épaulé, je vous suis infiniment reconnaissant pour votre présence. Le reste de ma famille également, toujours là en cas de besoin. Et pour finir, évidemment, **Marjorie**. Tu es la personne la plus importante pour moi, ton soutient durant ces années a été essentiel. C'est toujours le même plaisir pour moi de rentrer les soirs pour te retrouver, j'espère que ça durera encore très longtemps.

Table of contents

Table of figures

Abbreviations

ASO: AntiSense Oligonucleotide
asRNA: antisense RNA
ATP: Adenosine TriPhosphate
CAF: Cancer Associated Fibroblast
cAMP: cyclic adenosine monophosphate
CDS: Coding Sequence
CDX: Cell Derived Xenograft
ChIP: Chromatin ImmunoPrecipitation
ChiRP-MS: comprehensive identification of RNA binding proteins by mass spectrometry
circRNA: circular RNA
CSC: Cancer Stem Cell
DNA: DeoxyriboNucleic Acid
dsRNA: double stranded RNA
E x.x: Embryonic day
elncRNA: enchancer lncRNA
EMT: Epithelial to Mesenchymal Transition
ERV: Endogenous RetroVirus
FDA: Federal Drug Agency
G4: G-quadruplex
IFN: Interferon
IL: InterLeukine
IRES: Internal Ribosome Entry Site
ISR: Integrated Stress Response
lincRNA: Long intergenic noncoding RNA
LINE: Long INterspersed Elements
lncRNA: long noncoding RNA
LTR: Long Terminal Repeat
MET: Mesenchymal to Epithelial Transition
miRNA: microRNA
mPOS: mitochondria Precursor Over-accumulation Stress
mRNA: messenger RNA
MSC: Melanocyte Stem Cell
NCSC: Neural Crest Stem Cell
ORF: Open Reading Frame
OXPHOS: OXidative PHOSphorylation
PDX: Patient Derived Xenograft
PFS: Progression Free Survival
Pol: Polymerase
PRE: Pumilio Response Element
RAP-MS: RNA Antisense Purification coupled with Mass Spectrometry
RMC: Renal Medullary Carcinoma
RNA: RiboNucleic Acid
RNAi: RNA interference
rRNA: ribosomal RNA
RT-qPCR: retro-transcription followed by quantitive polymerase chain reaction
scRNA: single cell RNA
seq: sequencing
SINE: Short INterspersed Elements
siRNA: small interfering RNA
SKCM: Skin Cutaneous Melanoma
snoRNA: small nucleolar RNA
sRNA: small RNA
ssDNA: single stranded DNA
TE: Transposable Element
TF: Transcription Factor
TNF: Tumor Necrosis Factor
tRNA: transfer RNA
TSS: Transcription Start Site
UTR: UnTranslated Region
UVM: Uveal Melanoma
UVR: UltraViolet Radiation
XCI: X-Chromosome Inactivation

Résumé en Français

Introduction

Mon travail de thèse concerne le mélanome, un cancer particulièrement agressif, responsable de la grande majorité des décès liés à un cancer de la peau. Le mélanome apparait suite à des mutations notamment dans les protéines BRAF ou NRAS qui activent la voie MAP Kinase au sein des mélanocytes, entrainant une prolifération anormale des cellules et une résistance à la mort cellulaire. La prise en charge des mélanomes métastatiques de stade avancé fait appel principalement à l'immunothérapie mais aussi à des inhibiteurs de la voie MAP Kinase. Cependant une partie importante des patients ne répond pas aux traitements ou développe une résistance provoquant des rechutes. Les causes précises responsables de ces cas réfractaires aux traitements ne sont pas encore bien mises en évidence mais certaines pistes émergent.

De nombreux efforts ont été déployés pour comprendre les mécanismes moléculaires, tant génétiques qu'épigénétiques, à l'origine de la résistance aux thérapies. L'hétérogénéité cellulaire du mélanome est un obstacle majeur à une thérapie efficace. Les tumeurs de mélanome sont composées d'une variété de types de cellules caractérisés par des signatures d'expression génique spécifiques et des propriétés prolifératives, invasives et de résistance aux drogues différentes. Alors que les tumeurs naïves comprennent une grande majorité de cellules de mélanome de type mélanocytiques, d'autres états cellulaires tels que les cellules souches de la crête neurale (NCSC) et l'état indifférencié/mésenchymateux ont été caractérisés. Ces différents types cellulaires semblent capables de changer d'état de façon dynamique en réponse à leur environnement et notamment aux traitements, permettant une adaptation continue. Il est alors essentiel de mettre en évidence de nouvelles cibles et des nouvelles thérapies afin de lutter contre ce cancer.

Dans ce contexte, mon équipe d'accueil a identifié plusieurs ARN longs non-codants (lncRNAs) spécifiques et essentiels dans le mélanome. Les lncRNAs sont des molécules d'ARN d'au

moins 200 nucléotides ne possédant pas de cadre de lecture ouvert. Les lncRNAs sont impliqués dans une multitude de processus (transcription, traduction, réplication de l'ADN…) en interagissant avec des protéines ou acides nucléiques cibles. Les lncRNAs ont souvent des profils d'expression très spécifiques, dont certains uniquement présents dans des cellules tumorales. Ils sont ainsi des cibles thérapeutiques potentielles car leur dégradation peut être induite à l'aide d'oligonucléotides de synthèse de séquence complémentaire (antisense oligonucleotides, ASOs). L'équipe a contribué à l'identification de SAMMSON, un lncRNA impliqué dans la régulation de la traduction, et plus récemment, a identifié LENOX un régulateur du métabolisme et de l'homéostasie mitochondriale. LENOX est exprimé dans tous les types cellulaires de mélanome et est essentiel pour leur viabilité et pour la résistance aux inhibiteurs des MAP Kinases.

Mon travail de thèse se concentre sur un lncRNA nommé LENT (LncRNA ENhancer of Translation), exprimé principalement dans les cellules de mélanome de type mélanocytique sous le contrôle direct du facteur de transcription MITF. L'expression de LENT est associée à un mauvais pronostique des patients et sa déplétion par des ASO entraine l'apoptose des cellules de mélanome. Mon travail de thèse vise à caractériser le mécanisme d'action de LENT, ses partenaires protéiques et sa fonction afin d'évaluer son potentiel comme nouvelle cible thérapeutique dans le mélanome.

Résultats

J'ai montré que LENT est bien plus exprimé dans les cellules de mélanome cutané par rapport aux autres cancers ou tissus sains en explorant des bases de données mais aussi *in vitro* dans des lignées de cellules de mélanome. LENT est exprimé sous le contrôle direct du facteur de transcription MITF et une analyse des données de scRNA-seq a montré qu'il est exprimé dans plusieurs types de cellules de mélanome dont les cellules mélanocytiques (cellules exprimant des marqueurs de la ligné mélanocytaire désignées aussi comme pigmentées ou prolifératives), mais absent des cellules de mélanome de type NCSC et mésenchymateux. Des expériences de fractionnement cellulaire et de RNAscope sur les cellules en culture ou sur des sections de mélanome humains ont montré que LENT est localisé principalement dans le cytosol. Le silencing de LENT par des ASOs entraine une diminution de la prolifération et une induction de l'apoptose, tandis que son expression ectopique dans les cellules de mélanome ou HEK293T entraine une augmentation de leur prolifération. Cet effet est également observable sur des xénogreffes de cellules de mélanome *in vivo* chez la souris. En effet, l'injection d'ASOs par voie sous-cutanée diminue de façon significative la taille des tumeurs sur les souris en comparaison avec un groupe contrôle non-injecté. De plus, le silencing de LENT coopère avec celui des lncRNAs SAMMSON ou LENOX pour induire l'apoptose des cellules de mélanome. La coopérativité LENT-LENOX a aussi été observée *in vivo* avec une expérience de xénogreffe sur souris, comme mentionné ci-dessus. LENT est donc un nouveau lncRNA fortement exprimé dans les cellules de mélanome et essentiel à leur prolifération et leur viabilité, ciblable par ASO dans les cellules en culture et en xénogreffes *in vivo*.

Afin de comprendre son mode d'action, j'ai purifié LENT avec des oligonucléotides biotinylés et des billes de streptavidine à partir des extraits de cellules de mélanome dans des conditions natives ou suite au pontage par UV. Les protéines retenues avec LENT ont été identifiées par spectrométrie de masse. J'ai ainsi identifié l'hélicase DHX36 comme une protéine interagissant sélectivement avec LENT et non avec PCA3, un autre lncRNA utilisé comme contrôle. DHX36 est une résolvase de structures de type G-quadruplex (G4), requise pour la traduction des ARNm avec une structure G4. Cette interaction a été validé par immunoblot et par immunoprécipitation avec un anticorps spécifique de DHX36 et analyse des ARNs associés par RT-qPCR. Une expérience de

retard sur gel avec LENT transcrit *in vitro* et DHX36 purifiée a montré que cette interaction est bien directe entre les deux molécules.

J'ai ensuite montré par immunofluorescence et par fractionnement cellulaire que DHX36 et LENT sont enrichis au niveau des mitochondries dans les cellules mélanocytiques. Des expériences de fractionnement cellulaire en présence de digitonine ou de trypsine sur des mitochondries purifiées ont indiqué une association particulièrement forte de DHX36 avec ces organelles. J'ai donc pu montrer que LENT et DHX36 sont fortement associés aux mitochondries dans les cellules de mélanome. En revanche, le silencing de LENT n'altère ni la localisation ni la quantité de DHX36 dans les cellules. Nous avons donc supposé que LENT pourrait moduler l'activité de DHX36. Pour cela, un pulldown de DHX36 fut réalisé en présence ou en absence de LENT et les ARNs liés à l'hélicase furent récupérés et analysés par séquençage. Cette expérience nous a permis de déterminer l'étendue des ARNs liés par DHX36 dans une cellule de mélanome et de les comparer avec d'autres études menées dans différents types cellulaire. Il apparaît que dans les cellules de mélanome, DHX36 lie particulièrement des ARNs impliqués dans des processus mitochondriaux, en lien avec sa localisation particulière et sont enrichis dans des motifs GC-riches impliqués dans la formation des structures G4. D'autres processus sont communs entre les différentes études, notamment ceux liés à la réponse au stress et à la transcription.

Par la suite, nous avons cherché à comparer les profils d'ARNs liés à DHX36 avec ou sans LENT. Plusieurs centaines d'ARNs sont différentiellement liés par DHX36 après silencing de LENT, certains étant plus fortement liés à l'hélicase et d'autres, au contraire, moins liés. Les ARNs plus liés sont impliqués principalement dans des processus de mitophagie et de dommages à l'ADN. A l'inverse les ARNs moins liés sont corrélés avec des processus de synthèse du cholestérol et de la régulation des lysosomes. Une autre observation faite lors de la préparation des librairies pour cette expérience est la présence d'ARNs ribosomiques (18S et 28S) liés à DHX36. Nous avons donc pensé que DHX36 pourrait interagir avec la machinerie traductionnelle pour moduler la traduction des ARNs liés par l'hélicase. Pour vérifier cela, nous avons réalisé des expériences de fractionnement de polysomes suivi d'une analyse par western blot, montrant la présence de DHX36 dans la majorité d'entre elles, avec un enrichissement fort dans la fraction 80S. La présence de LENT dans ces fractions fut également déterminée par RT-qPCR, montrant une association forte de LENT avec la

fraction 80S par rapport aux autres fractions. La présence de DHX36 et LENT dans ces fractions suggère que ces molécules impactent la traduction cellulaire.

Nous avons examiné cette possibilité en réalisant des expériences de fractionnement de polysomes en présence ou absence de LENT, suivi d'un séquençage ARN sur les différentes fractions. La délétion de LENT semble modifier la quantité de centaines d'ARNs dans les ces fractions, en particulier dans les fractions de polysomes légères où de nombreux ARNs se trouvent moins abondants par rapport à la condition contrôle. Une analyse d'ontologie a montré que ces ARNs moins associés aux polysomes en absence de LENT sont impliqués dans la régulation de la dégradation et de la modification des protéines au niveau du réticulum endoplasmique.

Nous avons voulu savoir si une baisse de l'association de ces ARNm avec DHX36 et/ou avec les polysomes avait un effet visible sur la quantité des protéines codées par ces ARNs. Ainsi, le niveau des protéines NOX4 et RBPJ est augmenté après silencing de LENT, correspondant à une augmentation de la liaison entre DHX36 et les ARNm codant ces protéines. A l'inverse, les niveaux protéiques de UBE4A et de CTSD sont diminués et leurs ARNm sont moins liés à DHX36 après silencing de LENT. De façon similaire, les niveaux protéines de WFS1 et de HSPA5 sont réduits après délétion de LENT, leurs ARNm étant moins associés aux polysomes. La liaison d'un ARNm avec DHX36 semble donc moduler sa traduction, LENT régulant les ARNs liés par DHX36 et leur engagement dans le processus de traduction.

Afin de comprendre l'effet de la déplétion de LENT dans les cellules de mélanome, nous avons réalisé des expériences de microscopie électronique à transmission avant ou après déplétion de LENT. De manière évidente, la déplétion de LENT fait apparaitre de nombreux autophagosomes dans le cytoplasme des cellules, dont beaucoup contiennent des fragments de mitochondrie. Une induction de l'auto/mitophagie a été confirmé par western blot par l'accumulation de LC3B sous sa forme couplée aux lipides, suite au silencing de LENT. Cette accumulation a également été constaté dans des extraits de tumeurs xénogreffes des souris injectées avec l'ASO ciblant LENT. Cette observation est en lien avec le fait que les ARNs plus liés à DHX36 en absence de LENT sont impliqués dans le processus de mitophagie.

Cette mitophagie impacte la fonction et l'homéostasie des mitochondries. Des expériences permettant de mesurer la capacité respiratoire des cellules montrent que l'absence de LENT réduit la capacité respiratoire maximale des cellules et supprime leur capacité de réserve. De plus, nous

avons observé que la traduction et le niveau d'expression de plusieurs protéines de la chaine respiratoire mitochondriale étaient fortement induites après dépletion de LENT. Des expériences de fractionnement cellulaire et d'immunofluorescence ont montré que ces protéines tendent à s'accumuler dans le cytosol et même dans le noyau. Ce phénomène de mPOS (accumulation de précurseurs de protéines mitochondriales) entraine un stress cellulaire provoquant des dommages à l'ADN et la mort des cellules. En effet, j'ai pu montrer une accumulation de gH2AX, un marqueur de dommages à l'ADN, par western blot ainsi que par immunofluorescence quelques heures après l'accumulation de LC3B et donc de la mitophagie.

Conclusion

Mes travaux de thèse ont permis de caractériser LENT un nouveau lncRNA fortement exprimé dans les cellules de mélanome cutané. Cet ARN peut être ciblé par des ASOs, *in cellulo* et *in vivo*, entraînant sa déplétion et provoquant l'arrêt de la prolifération puis la mort cellulaire de mélanome. LENT interagit directement avec l'hélicase DHX36, responsable du déroulement des structures G4 sur l'ADN et l'ARN. En interagissant avec cette protéine, LENT module la liaison de DHX36 avec plusieurs centaines d'ARNm augmentant ou diminuant leur traduction. Cette modulation impacte la traduction des ARNm cibles, se répercutant sur différents processus cellulaires, notamment ceux codant des protéines impliquées dans la mitophagie et l'adressage des protéines au niveau du réticulum endoplasmique. Ainsi, la déplétion de LENT entraîne une mitophagie suivi d'une activation de la voie des dommages à l'ADN. En conséquence de la mitophagie, des protéines mitochondriales s'accumulent dans le cytosol et le noyau, entrainant une réponse mPOS qui combinée avec les dommages à l'ADN, conduit à la mort par apoptose. Cet effet est renforcé par la dérégulation des processus de réticulum endoplasmique, empêchant une prise en charge correcte des protéines et une dégradation normale de ces dernières.

Ces résultats soulignent l'importance capitale de la fonction mitochondriale dans les cellules de mélanome de type mélanocytique. Ces cellules ont donc développé un réseau de lncRNAs pour promouvoir leur survie avec SAMMSON, LENOX et LENT agissant tous pour optimiser la fonction mitochondriale par des mécanismes différents. En l'occurrence, LENT est nécessaire dans ces cellules de par la localisation de DHX36, fortement associée aux mitochondries indépendamment de LENT. DHX36 ne possédant pas une forte sélectivité de liaison, cette protéine nécessiterai donc un cofacteur lui permettant de discriminer les ARNs liés. Les raisons pour lesquelles ces cellules sont très dépendantes de leur fonction mitochondriale restent à éclaircir. Il a notamment déjà été montré que d'autres lncRNAs sont capables de moduler l'activité de DHX36 par liaison directe en indirecte.

En perspective, il serait important de poursuivre l'étude de la fonction de LENT d'un point de vue moléculaire pour comprendre précisément son mécanisme d'action. La détermination de la structure de LENT ainsi que son interface d'interaction avec DHX36 sont essentielles pour comprendre comment LENT modifie soit l'interaction de DHX36 avec des ARNs, soit sa sélectivité,

soit son activité résolvase. Malgré les expériences réalisées, le lien entre la traduction de protéines cibles et la mitophagie nécessite une étude plus approfondie. Parmi les protéines les plus affectées, se trouve NOX4 déjà connue pour réguler le ROS mitochondriale et la mitophagie ainsi que FUNDC1, une protéine liée à la mitophagie dans des conditions d'hypoxie. Il serait donc intéressant de tester si des inhibiteurs de ROS ou de NOX4 permettent d'atténuer les effets du silencing de LENT. Pour conclure, LENT est une nouvelle cible thérapeutique intéressante contre le mélanome cutané, pouvant être combiné avec d'autres ASO ciblant des lncRNAs tels que LENOX ou SAMMSON mais également avec des traitements déjà utilisés en clinique comme les inhibiteurs de la voie de signalisation MAPK.

Preface

This thesis aims to decipher the role of a novel long non-coding RNA (lncRNA) named LENT (LncRNA ENhancer of Translation) in skin cutaneous melanoma. LncRNAs represent a class of molecules largely understudied which display different roles in genetic diseases but also more recently shown, in cancer. Melanoma represents almost 80 % of deaths related to skin cancer and while therapies are available, several mechanisms of resistance challenges researchers to study new treatment options. In this context, it will be demonstrated in this work how lncRNAs could represent a new class of cancer therapy targets by exploring the example of LENT in melanoma.

The present manuscript is divided in several parts, beginning with an introduction separated in two sections. The first section will focus specifically on lncRNA, on their molecular aspects, but also regarding their role in diseases and how to target them. The second section focuses on skin cutaneous melanoma; its cell of origin, the onset of the disease as well as the characteristics of melanoma cells and the current treatments available. The introduction ends with the role of lncRNAs in melanoma, linking the two sections.

Results are presented in the form of a research article, containing its own introduction, material and methods, results, discussion and also references. To better understand the context of the results, I present some background concerning G-quadruplexes and the resolvase DHX36. The thesis ends with a discussion and gives perspectives about the next steps of the project.

Annexes present the different projects I have worked on, in parallel to the main project about LENT.

Introduction

Section 1: Long non-coding RNAs

A. LncRNAs characteristics

1. Definition

To begin with, it is necessary to state precisely what is a lncRNA. Historically, this class of molecules was distinguished from other types of RNA by mainly two criteria: size and coding ability. Indeed, it has been proposed that a lncRNA must be at least 200 nucleotides long [1], as a way to separate them from small non-coding RNAs such as tRNAs, snoRNAs, siRNAs or miRNAs for instance. This cut-off was apparently chosen as it represented a common exclusion size for RNA-seq library preparation for small RNAs. Additionally, as their name suggests, a lncRNA must not contain an open reading frame (ORF) composed of a start and stop codons, distinguishing them from mRNAs. By using this definition, the human genome could contain more than 100 000 lncRNAs [2,3], which includes rRNAs (5.8S, 28S and 18S) as a specific class of lncRNAs. However, annotation of lncRNA loci can be challenging due to their overall low expression and poor evolutionary conservation [2], generating disparities between databases [4,5], and making it difficult to assess precisely the total number of lncRNAs present in the human genome. Nevertheless, in terms of expression, lncRNAs represent only a very small fraction of the total RNA population in eukaryotic cells, representing 0.03 % to 0.2 % of the total RNA mass, much less than the abundant rRNAs or tRNAs [6].

However, a recent article gave a consensus statement about lncRNA research, revisiting the historical definition [7]. Indeed, the cut-off of 200 nucleotides was mostly arbitrary and left a “grey zone” of transcripts with sizes close to this limit, a little higher or lower. The consensus proposes to fit non-coding RNAs into three classes: first, small RNAs as transcripts shorter than 50 nucleotides; second, a class of transcripts between 50 and 500 nucleotides comprising RNA Pol III transcripts,

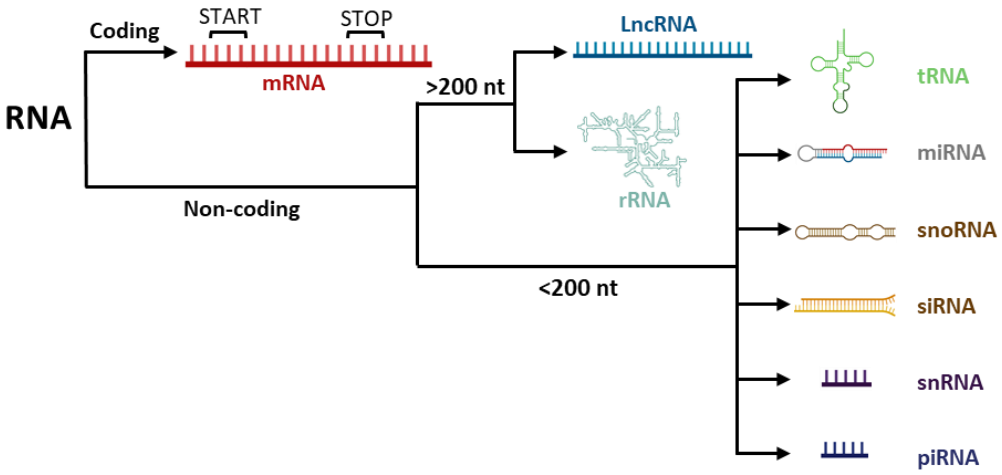


Figure 1: RNA classes.

The transcriptome is composed of different classes of RNA based on their coding ability and length. RNAs composed of an ORF that encodes a peptide are classified as messenger RNA (mRNA). If a noncoding RNA is longer than 200 nucleotides, it is designated as a long noncoding RNA (lncRNA), based on the historical definition. Ribosomal RNAs (rRNA) are also long noncoding RNAs, except for the 5,8S which is 150 nucleotides long. If a noncoding RNA is shorter than 200 nucleotides, it's a small RNA (sRNA) which is a class composed of different types of RNAs displaying various regulatory functions such as: transfer RNA (tRNA), micro-RNA (miRNA), small nucleolar RNA (snoRNA), small interfering RNA (siRNA), small nuclear RNA (snRNA) and piwi-interacting RNA (piRNA).

Pol V transcripts in plants and small Pol II transcripts; and third, lncRNAs longer than 500 nucleotides and mostly transcribed by RNA Pol II [7].

Whether we use the historical or more recent definition, it is clear that lncRNAs are defined more by what they are not, than by what they are: not mRNAs by their absence of coding potential and not small RNAs by their size (**Figure 1**). In this manuscript, we will mainly refer to the historical definition, as most lncRNAs studies were based following those criteria, but it is important to keep in mind that this definition is dynamic and prone to evolve as research in lncRNAs progresses.

2. Subtypes

Respecting the above definition, it is possible to classify lncRNAs into different categories, based mostly on their genomic location and proposed by the GENCODE annotation catalog [8]. The majority of lncRNAs are considered intergenic (lincRNAs), having their own promoter and their locus not overlapping with any nearby coding genes. They are mostly transcribed by RNA Pol II and are often processed and polyadenylated [9] , even though they do not encode any peptide. They can resemble mRNA by their structure, but are often way less expressed as it has been shown by RNA-seq in HEK293 cells for instance [10]. They can contain less introns compared to mRNA, as only 2.6 % of protein coding genes are intron-less compared to 32 % of lncRNAs. This feature potentially explains their low expression as introns are known to promote gene expression [10].

On the opposite, lncRNAs can be intragenic, with their sequences or part of their sequences overlapping with a coding gene. For instance, sense intronic lncRNAs are transcribed from introns of coding genes and do not possess their own promoter. They are less represented than lincRNAs in the literature and less studied, moreover they seem to be regulated through different processes [8,11]. Yet, it seems that they could represent the majority of ncRNAs present in mammalian cells, as observed in mice [12]. LncRNAs can also be considered antisense (asRNA), meaning that they are transcribed from the opposite strand of a coding gene, overlapping completely or partially with its sequence [8] . Their name is often composed of the name of the coding sequence followed by the suffix “AS”. Finally, some lncRNAs can be considered « bidirectional » as their promoter is shared with a coding gene but are transcribed in the opposite direction. If an RNA does not fall into any of these categories, it is considered as a processed transcript [8]. To add a layer of complexity, it has also been

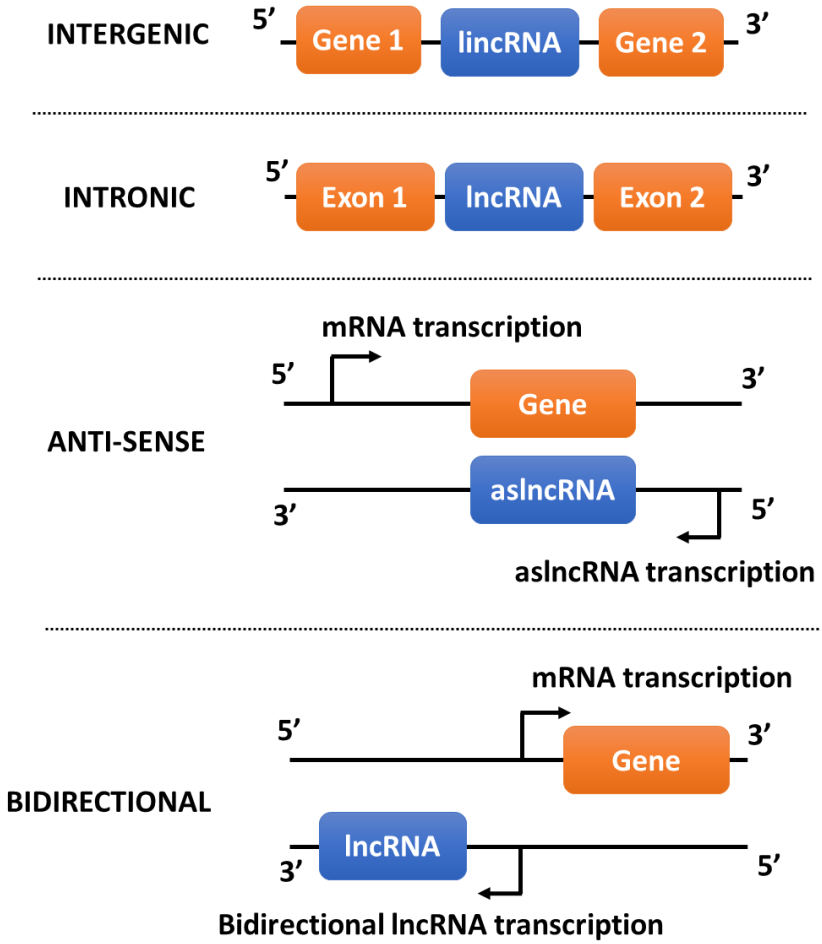


Figure 2: LncRNA subtypes.

The different categories of lncRNAs are essentially based on their genomic localization. Intergenic lncRNA (LincRNA) possess their own promoter and their locus is located between other genes, not overlapping with them. In contrast, intronic or genic lncRNA are transcribed from introns of coding genes. lncRNA can also be antisense (aslncRNA), transcribed from the opposite strand of a coding gene, partially or completely. Bidirectional lncRNA share their promoter with a coding gene but are transcribed from the opposite strand.

shown that lncRNAs may be processed and give rise to smaller RNAs, from their exons or introns, especially snoRNAs [13]. The different lncRNA subtypes are represented on **Figure 2.**

As stated above, it is also worth mentioning that rRNAs can be considered as lncRNAs, especially the 18S and 28S based on the recent definition, even though they display regulation and function very specific to their class. Their role in ribosome architecture and translation has been widely discussed in the literature [14] and will not be further elaborated in this manuscript as they are mostly not annotated or considered as lncRNAs.

3. Origins and conservation in evolution

Human lncRNAs are known to be poorly conserved, even between primates [15], and with well described orthologs present in mice known for only few and even fewer traced back to the common ancestor of vertebrates [16]. Overall, it has been estimated that almost 400 lncRNA genes could have originated more than 300 million years ago as orthologs of those genes have been characterized in chicken (**Figure 3**) [17,18]. Indeed, most of human lncRNAs are primate specific and display a low, specific expression limited to certain tissues or cancers [19,20]. A minority of lncRNAs are well conserved and display a strong expression such as XIST [21], NEAT1 [22] or MALAT1 [23], which will be described in coming parts. This lack of conservation suggests that most lncRNAs present on the human genome appeared recently during evolution, but also potentially that annotating orthologs can be difficult as RNA function is carried mainly through its 3D structure [24]. This particularity can hinder the ability of algorithms to identify lncRNAs carrying similar function in different organisms as they may not share similar sequences but can display similar 3D structures once processed.

Several hypotheses were suggested to explain the appearance of lncRNAs on the human genomes and are still discussed today. Those propositions encompass: loss of protein coding potential due to codon mutations, duplication events of repeated sequences from sRNA, de novo appearance of previously un-transcribed sequences due to changes in the genomic landscape, and insertion of transposable elements (**Figure 4**) [20]. The latter hypothesis seems responsible for the emergence of the majority of current human lncRNAs, as 75 % of those transcripts contain at least one exon of transposable element origin [20]. This number is much higher than for any other type of RNAs, suggesting a strong link between TEs and lncRNA origin. Moreover, lincRNAs are even richer

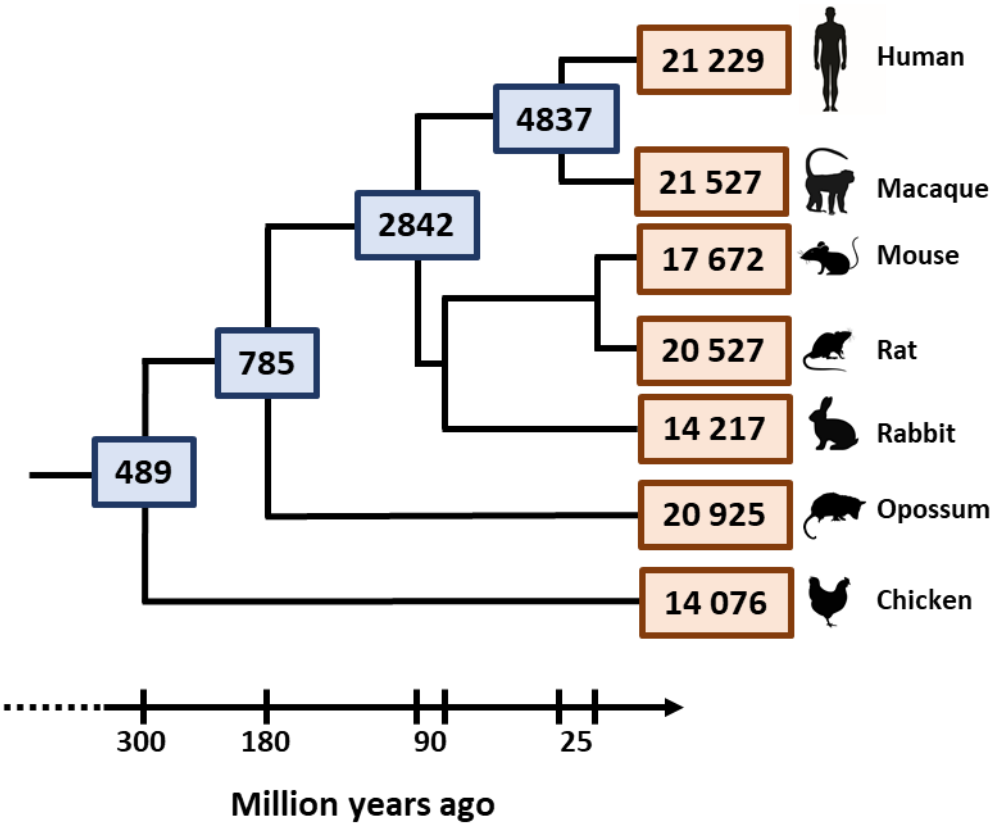


Figure 3: Phylogenetic tree of lncRNA orthologs.

The blue boxes indicate the number of lncRNAs orthologs between two branches. Brown boxes indicate the number of lncRNAs specific to the indicated species. Among the 31 678 human lncRNAs mapped in this study, only 489 are conserved between humans and chickens (approximately 1,5 %). Most lncRNAs are annotated as species-specific. Adapted from Sarropoulos et al., 2019.

in TE sequences compared to genic lncRNAs. A compelling example is the lncRNA-RoR, composed of 6 different TEs [25]. More specifically, different classes of TEs exist, but do all classes contribute to lncRNA origin? Among them are short interspersed elements (SINE), long interspersed elements (LINE) and LTR/ERV elements, which represent class I, but also DNA transposons which represent class II [26]. Class I elements are also known as retrotransposons as they need to be reverse transcribed to be inserted into the genome whereas class II are inserted through a cut and paste mechanism without being reverse transcribed. Even though the human genome is composed mainly of SINEs and LINEs of retroviral origin, they seem under-represented in lncRNAs sequences [20]. On the other hand, LTR retrotransposons and ERVs are depleted near coding genes, but highly represented in the vicinity of lncRNAs loci, an observation made in humans and mice.

In addition, TEs overlap with lncRNAs loci notably at the TSS (22.5 %) and at polyA sites (29.9 %), which represent a greater proportion compared to coding genes with TEs overlapping at 1.7 % of TSS and 7.9 % of polyA sites in this case [20]. Moreover, it was also observed that many human lncRNAs possess proximal promoters and TSS composed of long terminal repeats (LTRs) of primate specific TEs. Thus, LTRs often provide consensus TATA-elements and transcription factor binding sites [27], potentially explaining why so many lncRNAs show very tissue-specific expression. The reason why this class of retrotransposons is overrepresented in lncRNA loci and adjacent regions remain unclear, but a potential explanation may reside in the property of retroviruses to hijack transcriptional activators as a way to boost their own expression [28]. Indeed, the presence of LTR/ERVs would then be correlated with an increased transcription of their loci, giving rise to novel lncRNA composed of those elements or regulated by the binding of those TEs in their vicinity [29].

However, it is also important to understand how those novel lncRNAs can be selected once they appear during evolution. Globally, less than 10 % of the human genome is estimated to evolve under measurable selection, the rest being dominated by genetic drifting [30]. Indeed, the majority of lncRNAs accumulate mutations at a rate consistent with most of them lacking a function [31], supporting that the genome generates what can be considered as “junk transcripts” through cryptic TSS apparition, as seen above, or by pervasive transcription [32]. Among this pool of junk transcripts, some acquire beneficial functions by inhibition of nearby loci transcription rates [33] or by representing enhancer elements for other loci [34]. This process is known as neutral evolution and

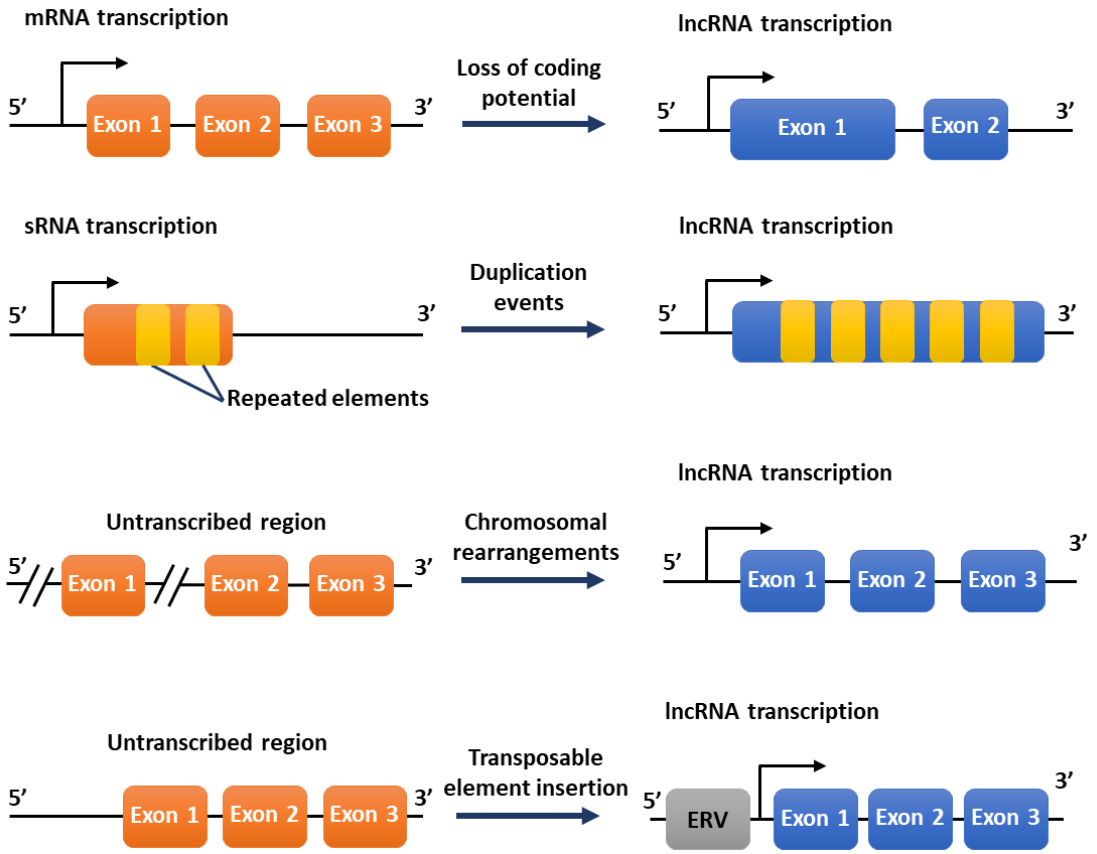


Figure 4: LncRNA origins.

Several genomic events can give rise to novel lncRNA. Genes can lose their coding ability through the removal of introns, exons or other regulatory elements but also by mutation of the START codon. They can also arise from duplication of repeated sequences, increasing the size of a previously small noncoding RNA locus. Some chromosomal rearrangements can also bring different regions of the genome close to each other, producing an environment favorable for the transcription of a novel locus. Previously untranscribed regions can eventually give rise to lncRNA through the insertion of a transposable element such as ERVs, providing transcription factor binding sites and other promoter sequences (TATA, Inr).

support an emergence of currently studied lncRNAs from a large pool of transcripts lacking a function and prone to disappear during evolution [35]. Selected lncRNAs can then gain additional functional domains trough subsequent evolution and selection, promoting their diversity.

In summary, lncRNAs show a wide diversity of structures, genomic locations, expression, and conservation. Many factors may contribute to their origins, but retroviral elements seem to be the major drivers of lncRNA expression and composition. They are mainly selected by neutral evolution from transcripts displaying regulatory functions of nearby genes. However, lncRNAs are very heterogenous and, thus, seem able to carry a plethora of functions in cells and to display a variety of mechanisms [36], which will be described in the following parts.

B. Molecular mechanisms displayed by LncRNAs

In this section, we will cover the main mechanisms by which LncRNAs exert their functions in cells, focusing on the molecular aspects of their mode of action by exploring different examples. We will separate those mechanisms by distinguishing nuclear and cytoplasmic lncRNAs. Indeed, many known lncRNAs seem present in the nucleus of human cells [8], consistent with them being first selected as the presence of their loci influence the transcription of nearby genes. Nevertheless, several lncRNAs were recently found acting in the cytoplasm, carrying out different functions and working through different mechanisms [37].

1. In the nucleus

First and foremost, many lncRNAs present in the nucleus were found to be associated with chromatin [38], suggesting that they may influence gene expression. It is the case for HOTTIP, a lncRNA known to control the expression of different *HOXA* genes by maintaining active chromatin marks near its target locus, through chromosome looping and recruitment of WDR5, a chromatin modifier [39]. In this case, HOTTIP acts as a scaffolder, interacting with a protein to maintain it at a certain localization to promote its specific action on a particular locus. This kind of mechanism is common to many lncRNAs, as their length enables them to fold into complex structures (loops, hairpins, long-distance interactions…) which represent potential interaction interfaces with other molecules, as shown for instance for the structured lncRNAs RepA [40] or Braveheart [41]. Another chromatin-associated example is lncPRESS1, a lncRNA regulated by p53 which interacts directly with the deacetylase SIRT6 [42]. By interacting with this protein, lncPRESS1 acts as a decoy to prevent SIRT6 access to open chromatin. Thus, these two chromatin-associated lncRNAs show similar protein scaffolding proprieties, but act in opposite manner, HOTTIP promoting protein localization to a locus and lncPRESS1 acting as a decoy, preventing a protein from accessing a locus. Both modulate epigenetic marks and chromatin accessibility, providing a first example of how lncRNAs act on gene expression.

In addition to interacting with chromatin modifiers, it has been shown that lncRNAs are also able to directly interact with ssDNA, forming DNA-RNA hybrids known as R-loops, acting as modulators of transcription and replication [43]. This type of interaction involves base-pair recognition

trough accessible regions of lncRNAs with complementary DNA sequence. A good example is the antisense lncRNA TARID, which was shown to bind the *TCF21* promoter and form an R-loop [44]. This R-loop promotes local DNA demethylation and thus triggers TCF21 expression. Moreover, R-loops are not the only structures formed by lncRNA interaction with DNA. This can be illustrated by a study on the antisense lncRNA KHPS1 which is able to form a triplex structure (RNA-DNA-DNA) by interacting near the TSS of the oncogene *SPHK1* [45]. This triplex serves as a platform to recruit the p300/CBP complex, promoting the transcription of the *SPHK1* locus.

It has been shown additionally that some lncRNAs can be transcribed from enhancer elements, known to promote expression of loci, even at long distance on the genome by forming chromatin loops [46]. An example of those elncRNAs are the RNAs ELEANORS which represent a set of lncRNAs able to promote several loci interactions around the *ESR1* gene, activating its transcription by forming a positive chromatin environment through the recruitment of activator proteins and transcription factors [47]. Indeed, they are often used as bridges to connect the enhancer and the target loci and can serve as a platform for protein complexes.

The previously described mechanisms involve regulation of gene transcription, by modifying the chromatin environment, either by protein or DNA binding. Nevertheless, nuclear lncRNAs can also harbor phase separation proprieties, a recent topic gaining interest in the last decade [48]. A very well conducted study described recently the implication of the lncRNA NORAD on phase separation in human cells [49]. NORAD is necessary to avoid aberrant mitosis in cells by negatively regulating Pumilio proteins PUM1 and PUM2, which act as post-transcriptional repressors that can bind pumilio response elements (PREs). The challenge to explain NORAD's mode of action resides in the fact that only 400 NORAD molecules are present in HCT116 cells, each NORAD containing 18 PREs, whereas more than 130 000 PREs are present on expressed mRNAs. Based on this affirmation, NORAD would seem unable to efficiently buffer Pumilio proteins. Interestingly, the authors described that NORAD and PUMs are found together inside nuclear bodies, disrupted by the loss of NORAD. Indeed, *in vitro,* the addition of NORAD to PUMs triggers phase separation, enabled by the presence of intrinsic disordered regions on PUMs. Thus, a very restricted number of lncRNAs could impact a large number of proteins by forming liquid-liquid phase separated structures, sequestering them away from their targets [50].

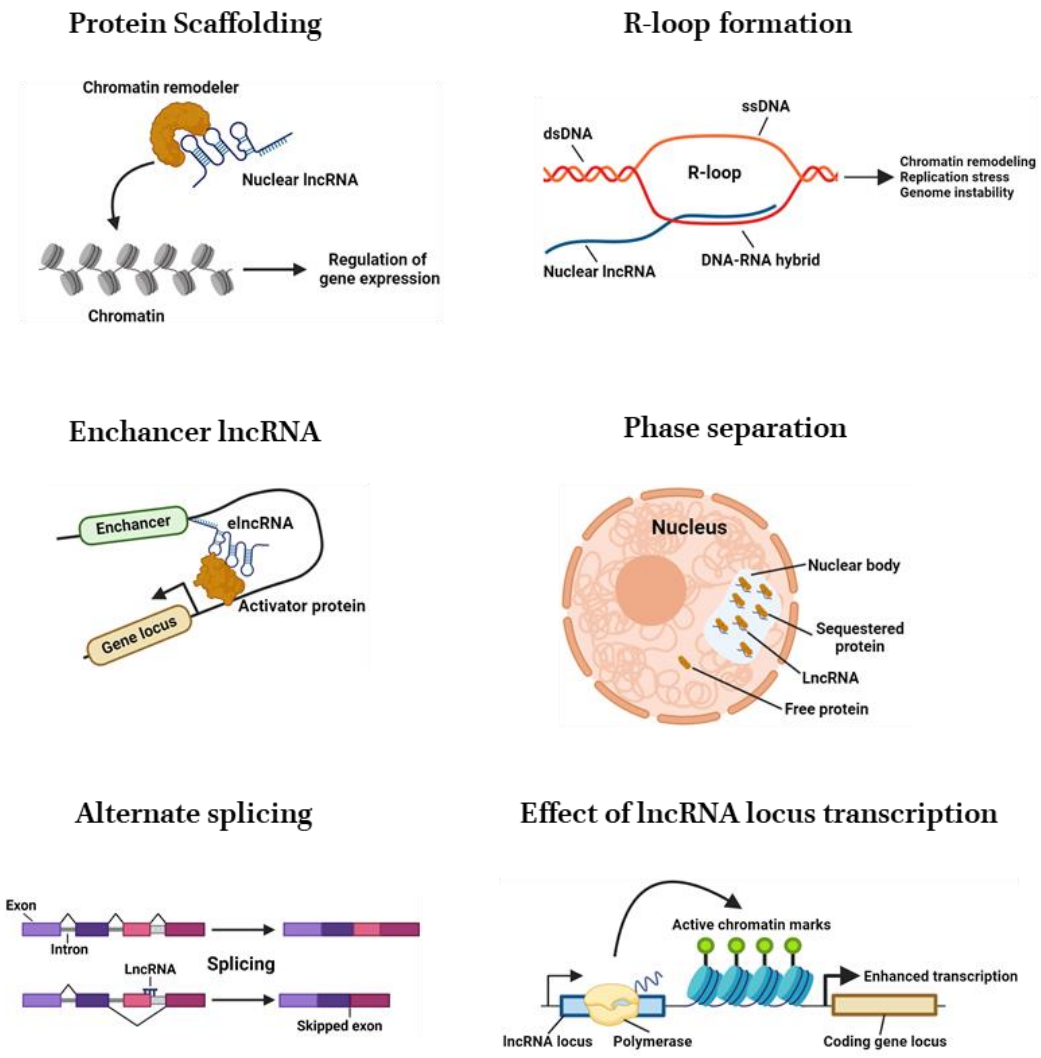


Figure 5: Nuclear lncRNA mechanisms.

LncRNAs can bind proteins directly through their complex structures, modulating their function and/or localization. They can also directly bind DNA to form hybrids and structures known as R-loops. These R-loops can change the local genomic landscape by influencing chromatin mark deposition. LncRNAs can also be transcribed from enhancers to form an active chromatin environment for a distant locus by recruiting activator proteins. They are also able to trigger phase separation in the nucleus to sequester pools of proteins and other RNAs to modulate their function. In addition, they can also influence RNA splicing by covering splicing sites through direct binding with the RNA. Lastly, the transcription of lncRNA loci can act intrinsically by influencing the surrounding chromatin environment, potentially promoting or inhibiting the expression of nearby genes.

Another possible mechanism of action for nuclear lncRNA is the regulation of splicing events. Splicing is an essential step in the processing of RNAs, including lncRNAs, and is carried out by the spliceosome machinery to produce mature and functional molecules [51]. A previous study showed that an antisense lncRNA of the transcription factor ZEB2 is necessary to maintain a large intron containing an IRES in the 5'UTR of the ZEB2 mRNA, necessary for ZEB2 expression [52]. In this example, the aslncRNA covers the 5' splice site of the intron through direct binding, preventing its splicing and promoting ZEB2 expression.

Finally, as mentioned on the previous part on neutral evolution, lncRNA are not directly responsible for regulation of gene expression, but it is simply their transcription that modulates the expression of nearby genes, as shown by the lab of Eric Lander [53]. In this paper, they focus on different lncRNA loci, for which the deletion of their promoter strikingly affects the transcription of genes close to their respective loci. As a specific example, they showed that deleting the promoter of the lncRNA BENDR reduced expression of the neighbor gene *BEND4* by 57 %, whereas the elimination of the spliced BENDR had no effect on BEND4 expression. Moreover, they also describe that other types of RNA loci such as mRNAs can carry out this function, but it is particularly interesting that lncRNAs harbor this mechanism as it could also explain the importance of some lncRNA showing a very low number of copies per cell, their locus carrying the function. Instead of inhibiting nearby genes transcription, promoters of lncRNAs can also represent enhancer sequences for distant loci through chromatin looping [54].

Thus, nuclear lncRNAs display a variety of mechanisms of action often linked with regulation of transcription. Whether they interact with proteins or nucleic acids, they can change the chromatin landscape of the nucleus, promoting or repressing gene expression. They are also able to modulate the splicing of certain transcripts by covering splicing sites or to influence nearby genes by the transcription of their locus. Interestingly, they are even able to form condensates within the nucleus, trapping proteins and nucleic acids in those structures. These examples highlight the fact that, even if often present in small numbers in cells, lncRNAs can influence critical pathways and basal cell functions thanks to their structure and interactions with proteins and nucleic acids. Nuclear lncRNA mechanisms are recapitulated on **Figure 5**.

2. In the cytoplasm

In contrast to the above, lncRNAs present in the cytoplasm tend to act through different mechanisms, even though the general principles of lncRNAs function (protein and nucleic acids interaction) are retained. As mentioned above, lncRNA functions in the cytoplasm are less well documented than in the nucleus, even if the number of articles describing cytosolic lncRNAs is constantly rising.

One mechanism is the ability of lncRNAs to impact the stability of cytosolic mRNAs, as is the case for the lncRNAs ½-SBS [55]. It was demonstrated that ½-SBS RNAs contain a sequence enabling them to directly interact by base pairing with the 3'UTR of their mRNA targets. This interaction facilitates the binding of the STAU1 protein, a known inducer of RNA decay, which was shown to bind translationally active mRNAs. Those lncRNAs carry their function by forming double-stranded RNA structures promoting their target degradation, resembling the action of other RNAs species such as miRNAs. In an opposite manner, lncRNA binding to a mRNAs can also promote their stabilization as shown with the mRNA BACE1 and its aslncRNA BACE1-AS [56]. Mechanistically, BACE1-AS binds directly BACE1 mRNA to from a duplex structure to increase the half-life of the mRNA. The precise mechanism has not been elucidated but the authors showed that depletion of BACE1-AS reduces BACE1 mRNA and protein levels in cells. Thus, lncRNAs seem able to modulate the stability of a set of mRNAs in the cytosol, influencing indirectly the proteome of the cell.

As it has been shown for mRNAs, lncRNAs are also able to stabilize cytoplasmic proteins. To illustrate this, consider the example of lncRNA MaIL1 [57]. This lncRNA can stabilize the protein OPTN by direct interaction, which prevents ubiquitination of OPTN and its subsequent degradation. Of course, lncRNAs can also promote the degradation of proteins through their binding as it is the case for the lncRNA NRON which promotes degradation of the HIV protein Tat through direct binding in association with other proteins implicated in the proteasome system [58]. Here, NRON display a scaffolding function, by bringing protein Tat closer with other proteins that promote its degradation.

These examples illustrate that similar to nuclear lncRNAs, cytoplasmic lncRNAs are also able to interact with proteins and nucleic acids to modulate their stability and function through direct interactions. Nevertheless, cytoplasmic lncRNAs can also harbor specific mechanism adapted to their localization. One of those is the potential for some lncRNAs to act as "sponges" to retain

miRNAs such as the lncRNA TYRP1 studied by the Galibert group [59]. miRNAs are a group of small, single-stranded RNAs, processed by the RISC complex and able to degrade cytosolic RNAs trough base-pair recognition [60] . TYRP1 was shown to bind a miRNA designated miR-16, without being degraded. As TYRP1 and miR-16 were shown to be present in similar levels in cells, TYRP1 then seemed to act as an efficient "sponge" to dampen miR-16-induced degradation of RNAs. It is also possible for lncRNAs to harbor a large number of miRNAs binding sites, enabling one lncRNA molecule to retain a population of miRNA, as demonstrated for the lncRNA ciRS-7 [61]. This RNA belongs to a class of lncRNAs known as circular RNAs (circRNAs) and harbors 70 miRNAs binding sites across its length. CircRNAs are generated by the spliceosome machinery through a back-splicing event [62] are known to be stable RNA molecules as they are insensitive to exonucleases degradation [63]. This type of mechanism is particularly efficient, enabling the cell to dampen the effect of a specific set of miRNAs through the expression of a very limited number of lncRNA molecules.

Most importantly, the cytoplasm is the compartment where translation takes place, and lncRNA can also be implicated directly in this process, as they are often found in ribosomal cell extracts with up to 70 % of cytoplasmic lncRNA bound to ribosomes [64,65]. However, debates are still ongoing whether those lncRNAs carry a function or if they are simply degraded at ribosomes, as blocking translation elongation seem to enhance the stability of the majority of ribosome-bound lncRNAs [64]. However, there is also the possibility that some of those lncRNAs modulate the translation process, as shown for instance in a study of the lincRNA-p21 [66]. In this work, authors showed that lncRNA-p21 was able to bind selectively the mRNAs JUNB and CTNNB1, reducing the amount of those mRNAs in polysomal fractions, thus lowering their translation efficiency.

Yet, there is also another option to explain the presence of lncRNAs in ribosomal fractions: are those RNAs all really non-coding? Indeed, lncRNAs are annotated often automatically, based on the presence of open reading frames and known related proteins [8], but what happens if there is an unconventional start codon or if the produced peptide is very short and has not yet been characterized? Following this idea, it has been shown that, as pervasive transcription exists and could be the source of different lncRNAs [32], pervasive translation also occurs in cells and allow translation of short ORFs in lncRNAs into peptides [67]. On the other hand, it seems that peptides

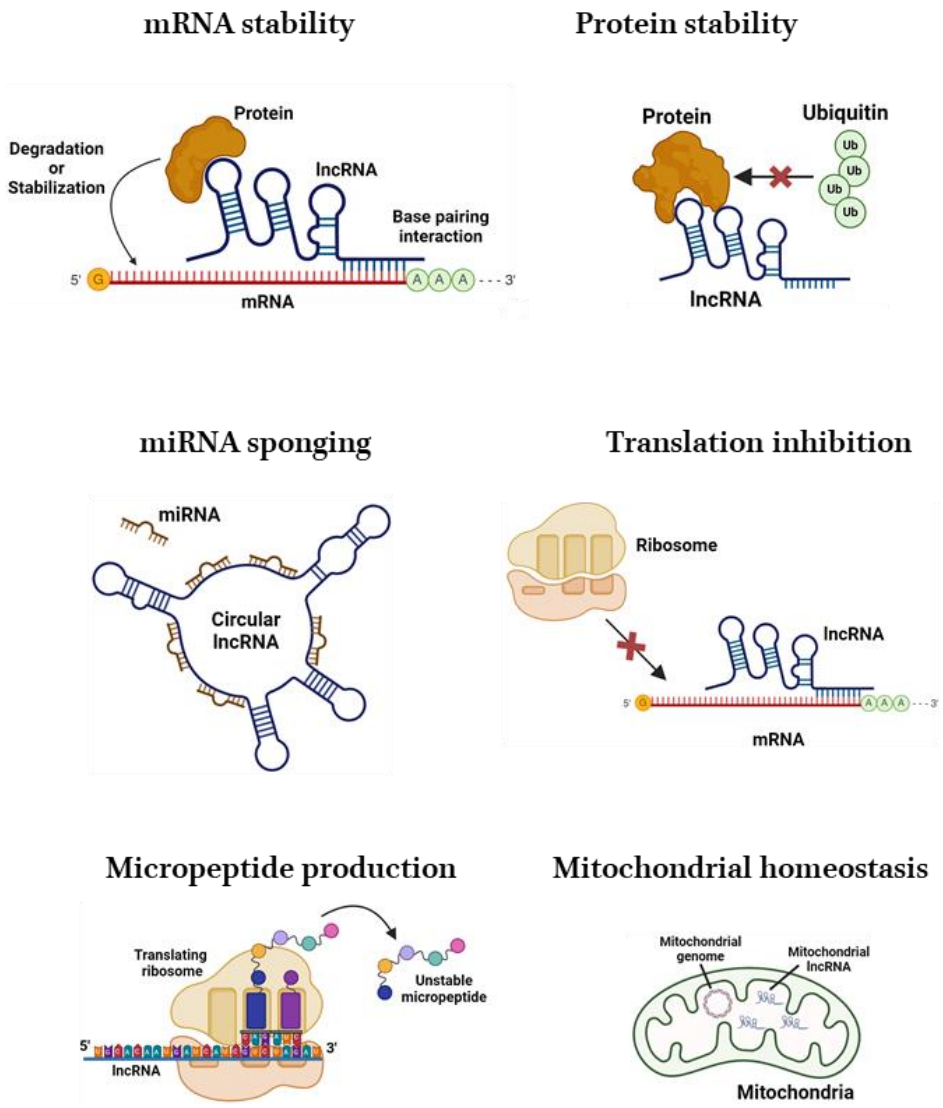


Figure 6: Cytoplasmic lncRNA mechanisms.

LncRNAs are able to directly bind cytosolic mRNAs through base pairing, bringing different modulators of mRNA stability with them. Similarly, lncRNAs can also bind proteins to modulate their stability, by preventing their ubiquitination for instance. On the other hand, lncRNAs can be circular composed of many miRNA binding sites, enabling them to efficiently buffer different mRNA populations. They can also directly modulate mRNA translation through direct interaction, preventing the mRNAs from accessing the translation machinery. Surprisingly, lncRNAs can also encode small peptides which can be translated, even if most of them are rather instable. Last, they can also be transcribed directly from organelle (mitochondria, chloroplasts) genomes, potentially modulating their homeostasis.

produced from lncRNAs are often unstable because of the presence of degradation signals and hydrophobic residues at the C-terminus of those proteins [67]. Thus, non-canonical peptides might not harbor a function in cells as they seem rapidly degraded.

Nevertheless, evidence has been reported that non-canonical peptides translated from lncRNA could act on cellular processes, but with only very few examples well described [68]. An impressive work was carried out by Chong and colleagues, coupling whole exome sequencing, bulk and sc-transcriptomics, ribosome profiling and mass spectrometry [69]. They were able to identify hundreds of non-canonical peptides derived from noncoding RNAs and TEs, expressed in different tissues and cell types, but also to confirm many lncRNAs as "true" noncoding. In addition, they were able to demonstrate the immunogenicity of several of the peptides and their upregulation upon IFNγ treatment, proposing that they are not just products of pervasive translation but could be deliberately translated to modulate signaling pathways. Subsequent work corroborated this hypothesis as IFNγ treatment seem to induce ribosomal frameshifts which produce aberrant peptide presentation (known as neo-antigens), modulating immune cell recognition [70].

To conclude on their function in the cytoplasm, it is also worth mentioning that three lncRNAs are transcribed from the mitochondrial genome, and thus found inside the mitochondria [71]. It was observed that those mitochondrial lncRNAs have a cell and tissue specific expression and are regulated by nuclear-encoded proteins such as the RNase MRPP1, leading to the supposition that they could regulate mitochondrial function to some extent. It is also the case for the chloroplast genome in plants which produces a long antisense RNA [72].

Strikingly, lncRNAs are able to act on a plethora of cellular processes, ranging from transcription and splicing in the nucleus, to mRNA degradation and translation in the cytosol. They can interact with proteins, DNA and RNA to modulate the stability of their target and their localization but can also sequester them inside different structures based on phase separation or sponging proprieties. LncRNAs can also be found at organelles, notably mitochondria, to influence different cellular functions. Surprisingly, they can also be translated into short peptides which can presented at cell surface for instance, revisiting their status as non-coding RNAs. Cytoplasmic lncRNA mechanisms are recapitulated on **Figure 6.**

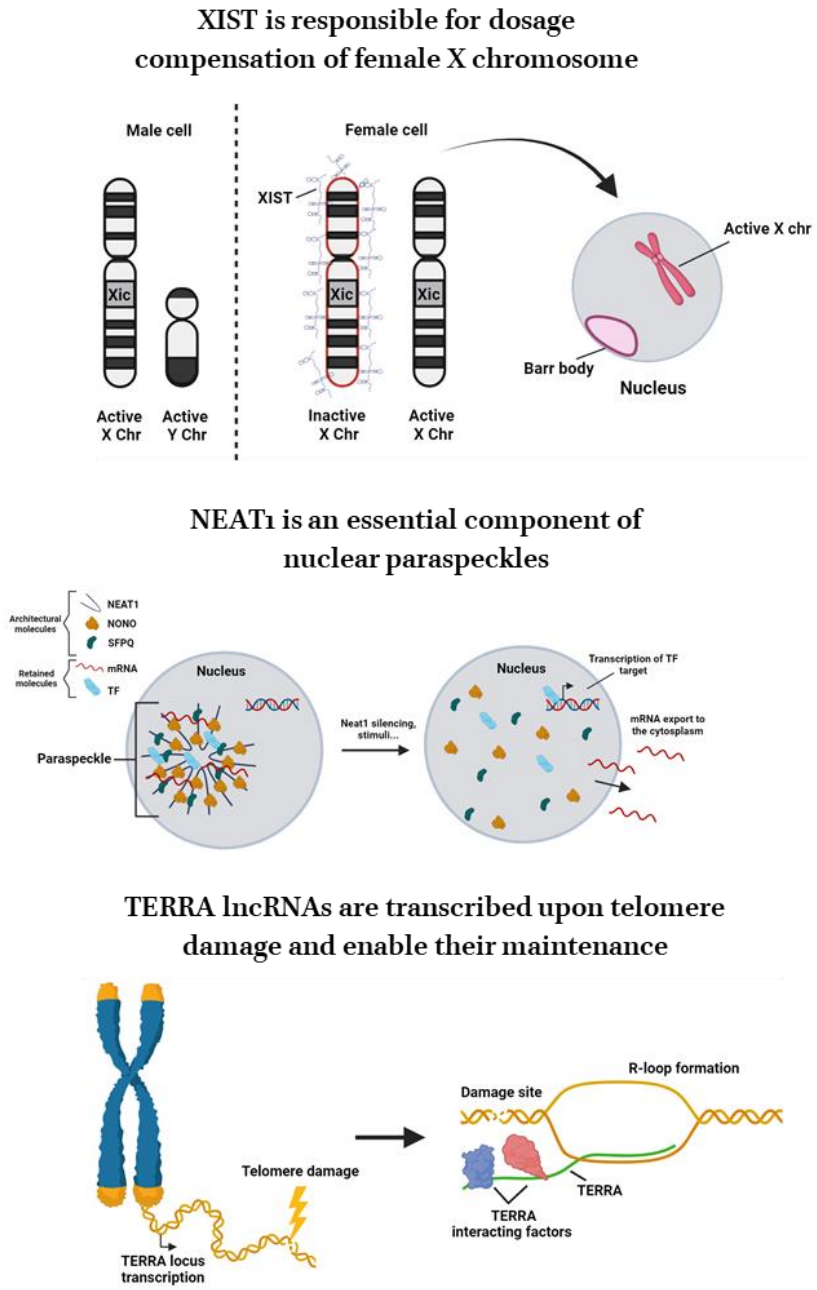


Figure 7: Well conserved and expressed nuclear lncRNAs participate in cell homeostasis.

XIST is expressed from the X chromosome (Xic = X locus) and covers one copy of the X chromosome in female cells, a process known as XCI. XIST modifies almost the entire X chromosome landscape, promoting heterochromatin structure. The silenced chromosome appears as a condensed body in the nucleus, called a Barr body. NEAT1 is an architectural lncRNA, responsible for the formation of membrane-less nuclear bodies known as paraspeckles. By interacting with other factors, NEAT1 create a liquid-liquid phase separation, trapping other RNAs and proteins. Upon NEAT1 silencing, paraspeckles are destabilized, releasing many factors which can then influence gene expression or mRNA export to the cytoplasm. TERRAs are transcripts produced from the sub-telomeric region upon telomere damage. They can directly bind DNA, recruiting factors involved in DNA repair and promoting telomere maintenance.

C. Functions of lncRNAs in health and diseases

Now that we have seen the diversity of molecular mechanism displayed by lncRNAs, it is time to explore their roles and functions on a larger scale, as it becomes clear that those molecules are crucial to the progression of various diseases, with more than 200 000 lncRNA-disease associations reported in the lncRNADisease 2.0 database [73]. As in the last section, several examples will be discussed, focusing more on their importance in disease progression and normal physiology.

1. Normal conditions

LncRNAs are not necessarily linked to a pathology, they can also be essential for the normal functions of the organism. Interestingly, these lncRNAs are often among the most conserved and expressed in cells, highlighting their importance [10]. The best example is the lncRNA XIST, regularly considered as one of the most globally expressed and conserved lncRNA, but also one of the most studied [21] and earliest described [74]. Of note, it has been shown that XIST is exceptionally conserved in eutherians (placental mammals) and evolved through the loss of coding potential of a set of protein coding genes present only in non-eutherian vertebrates [75] XIST is transcribed from the X chromosome and essential for XCI (X-chromosome inactivation) which is established in the female organism of mammals and crucial to prevent autoimmune diseases due to toxic effects arising from double dose expression of certain chromosome X genes [76]. It acts to spread heterochromatin across the X chromosome, repressing almost 1000 genes, by forming foci in different location of the chromosome [77] and recruiting a variety of proteins, often linked with gene silencing such as SHARP [78]. XIST works as the central actor of a larger machinery as shown by comprehensive identification of RNA binding proteins by mass spectrometry (ChIRP-MS), identifying more than 80 protein partners for XIST [79]. This capacity of XIST is enabled by its length of 17 000 nucleotides, representing a large interaction interface for protein scaffolding, a mechanism described in the previous section.

Following the discovery and description of XIST, researchers have been looking for similar, ubiquitous, and highly expressed lncRNAs. In this context, two lncRNAs have been identified designed as NEAT1 and NEAT2 (known now mainly as MALAT1), both well conserved in mammalians [80]. Although they share a similar name, those two transcripts share no homology. Since its identification, NEAT1 has been shown to be essential for the assembly of paraspeckles in the

nucleus by associating with paraspeckle proteins, thus making it known as an "architectural lncRNA" [22,81]. Paraspeckles are membrane-less structures capable of retaining proteins and nucleic acids to regulate gene expression and are present in a variety of cells [82]. Loss of NEAT1 completely destabilize paraspeckles [81], potentially leading to cell death by apoptosis through the release of pro-apoptotic factors [82]. The NEAT1 mechanism resembles that of the previously mentioned NORAD, also an abundant and conserved lncRNA [49]. Indeed, it is able to directly interact with proteins such as NONO and SFPQ, triggering condensation and liquid-liquid phase. Regarding MALAT1, it was described that this lncRNA could be involved in the regulation of splicing [83] but also gene expression through chromatin binding along with NEAT1 [84], although it does not display similar architectural proprieties. Surprisingly, loss of MALAT1 does not produce a readable phenotype in mice, although local transcription seems altered [85], suggesting that the role of this lncRNA could be restricted to specific conditions in a non-disease context, potentially more in adult cells.

One last interesting example are the TERRA lncRNAs, produced from chromosomes ends and ranging from several hundred nucleotides to several thousand [86]. TERRAs are expressed in a plethora of human and mouse cell lines, showing their conservation beyond primates. They directly bind short or damaged telomeres, forming R-loop structures, in association with factors linked with DNA repair to promote telomere maintenance, essential in cell aging [87]. It has also been suggested that TERRAs could promote chromatin modifier action at telomeres, modulating epigenetic mark deposition during cell cycle phases but also TERRA recruitment through a feedback loop [88].

Through those few examples (**Figure 7**), we can appreciate that several lncRNAs carry out basal functions in the organism and are often nuclear, well expressed and conserved, and modulate directly or indirectly gene expression. Thus, those molecules once considered as "transcriptional noise", can in fact display functional features similar to other regulatory RNAs classes or proteins and should be regarded as crucial modulators of cell homeostasis.

2. Genetic diseases and non-cancerous disorders

As stated previously, lncRNAs often display specific expression in a certain type of tissue and/or disease, making them interesting biomarkers and/or therapeutic targets. Several disease-associated

lncRNAs have been characterized, even though the number of studies on diseases other than cancers remains limited.

Involvement of LncRNA in genetic diseases was suggested in the early 2000s with the study of a patient displaying an inherited form of anemia called Thalassemia, displaying a deletion between two gene clusters, enabling abnormal antisense transcription of the *HBA2* gene [89], normally producing a hemoglobin essential for oxygen transport in the blood. This was explained by the removal of the adjacent gene transcription termination site by the deletion, enabling the RNA Pol II to transcribe a region of the strand opposite to the *HBA2* locus. The presence of those abnormal transcripts was then linked to silencing of the *HBA2* locus by methylation. Thus, loss of this hemoglobin by abnormal expression of an aslncRNA highlighted the potential for lncRNAs to be responsible of genetic diseases. In this case, a novel lncRNA was formed by a genetic event, but several known lncRNAs have also been implicated in genetic diseases onset.

Conversely, events involving lncRNA translocation cause various disorders documented in different case studies. For instance, two families with brachydactyly (short fingers and toes) displayed translocations on chromosome 12, leading to downregulation of a parathyroid growth hormone designed PTHLH [90]. The consequence of this event is the displacement of the elncRNA DA125942 locus, normally necessary for PTHLH transcription by forming a chromatin loop between its enhancer and the hormone locus. In addition, this event seems to promote DA125942 expression, leading to dysregulation of additional genes through an unexplored process. In another case, a rare translocation event was shown to disrupt the lncRNA RMST locus, downregulating its expression [91]. As RMST is essential for neural development via its association with the transcription factor SOX2, the patient displaying this translocation presented Kallmann Syndrome (impairment of smell and delayed or absent puberty). As a final example, a study focused on lncRNAs implicated in autism spectrum disorder and intellectual disability by producing copy number variation morbidity maps of a cohort of children [92]. This study identified a lncRNA called lnc-NR2F1 specifically disrupted in those patients due to translocation events. Depletion experiments of lnc-NR2F1 in ES cells showed that this lncRNA controls neural gene expression and is essential for neuronal cell maturation.

In addition to genetic diseases, lncRNA functions in other non-cancerous diseases have also been explored in different studies. For instance, it has been shown that the lncRNA DANCR could

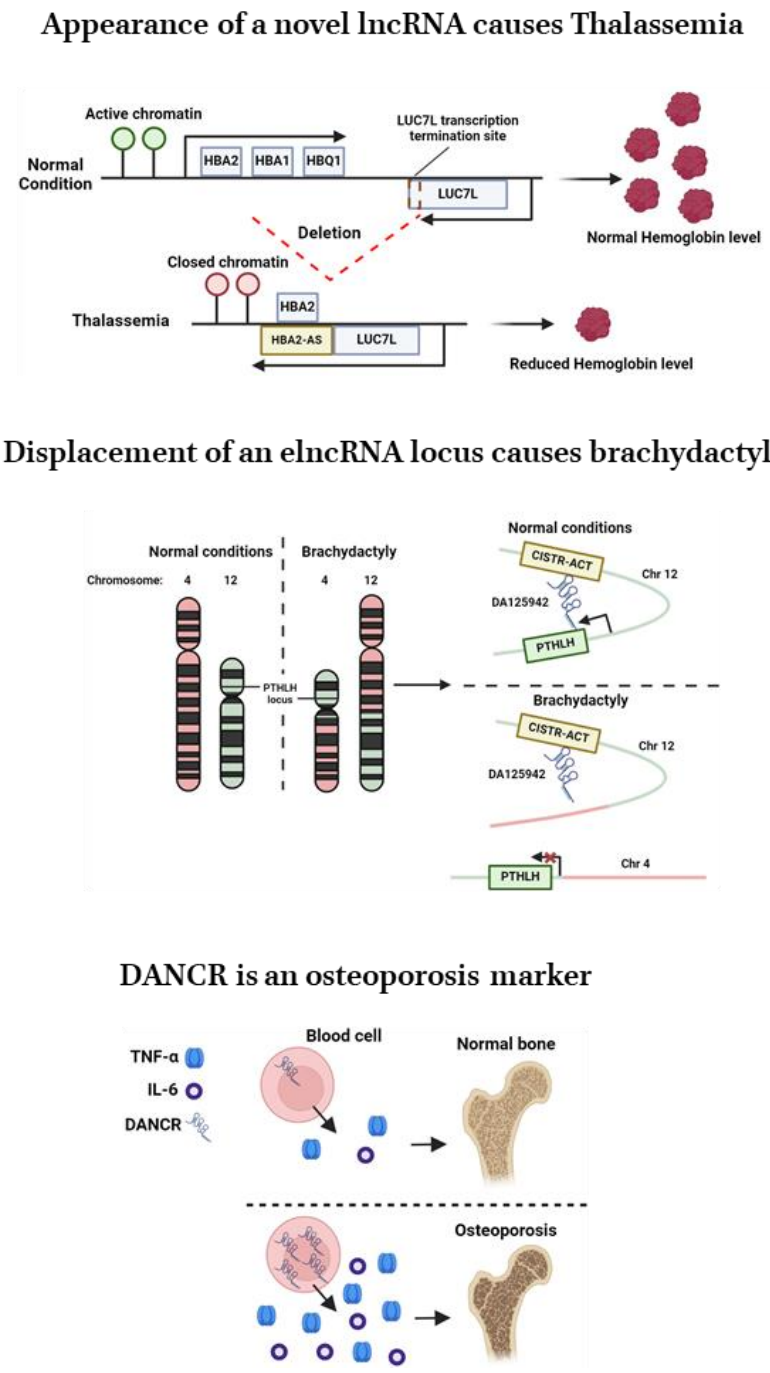


Figure 8: LncRNAs are implicated in genetic and other non-cancerous diseases.

A novel aslncRNA is formed by the removal of a transcription termination site by a genomic deletion. The presence of this RNA modifies the chromatin environment, repressing the expression of hemoglobin, leading to thalassemia. A translocation event between chromosomes 4 and 12 causes the displacement of the DA125942 elncRNA locus (CISTR-ACT), preventing its association with the PTHLH hormone locus. PTHLH is thus downregulated, leading to brachydactyly. The lncRNA DANCR has been identified as an osteoporosis marker as its high expression in blood cell is correlated with increases in cytokine production, enabling bone resorbing activity of these cells.

be used as a biomarker of osteoporosis [93], a condition where bones become brittle, partially due to monocyte bone-resorbing proprieties. Indeed, DANCR was found significantly and specifically over-expressed in immune cells of patients showing signs of osteoporosis. The authors showed that DANCR promotes the production of IL6 and TNF-a, two cytokines enabling bone-resorbing activity by immune cells. Another study focused on lncRNAs specifically dysregulated in type 2 diabetes (T2D) by analyzing transcriptomic data from pancreatic cells of patients [94]. A little more than 1000 lncRNAs were found to be expressed specifically in pancreatic cells amongst which a few were found dysregulated in T2D, namely KCNQ1OT1 and HI-LNC45. The functions of these lncRNAs in T2D have not been characterized yet, but for now they could be used as biomarkers of T2D onset. A last example is work on the role of aslncRNAs implicated in hypertension [95]. This study described a set of seven genes implicated directly in the regulation of blood pressure and focused on one of them called NPPA as it modulates blood flow through smooth muscle contractions and presents a natural antisense transcript designated NPPA-AS. This asRNA is poorly conserved outside of primates and is co-expressed with the sense NPPA, modulating its splicing through direct interaction and thus could regulate its expression. The precise function of NPPA-AS remains unexplored but this lncRNA could represent a main regulator of blood pressure and a potential biomarker for hypertension.

To summarize, lncRNAs are directly implicated in various genetic or non-genetic diseases (**Figure 8**). These transcripts can be novel, generated due to genomic events enabling aberrant transcription of loci and be responsible for abnormal regulation of important developmental factors. Their specific expression in certain disease contexts can also make them interesting biomarkers, whose expression in biological samples can be detected by transcriptomic analysis or RT-qPCR. Moreover, specific tools have been recently created to discover new lncRNA biomarkers such as Firalink, a pipeline already tested in cardiovascular and brain pathologies [96], which in the future could enable researchers to identify an increasing number of disease-related lncRNAs.

3. Cancers

It is now established that lncRNAs can display disease-related expression and functions. Interestingly, they have been correlated with various cancers in the last decade [97], being regarded as new biomarkers and potential therapeutic targets [98]. The topic of melanoma-specific lncRNAs will not be discussed here but in the second section on melanoma.

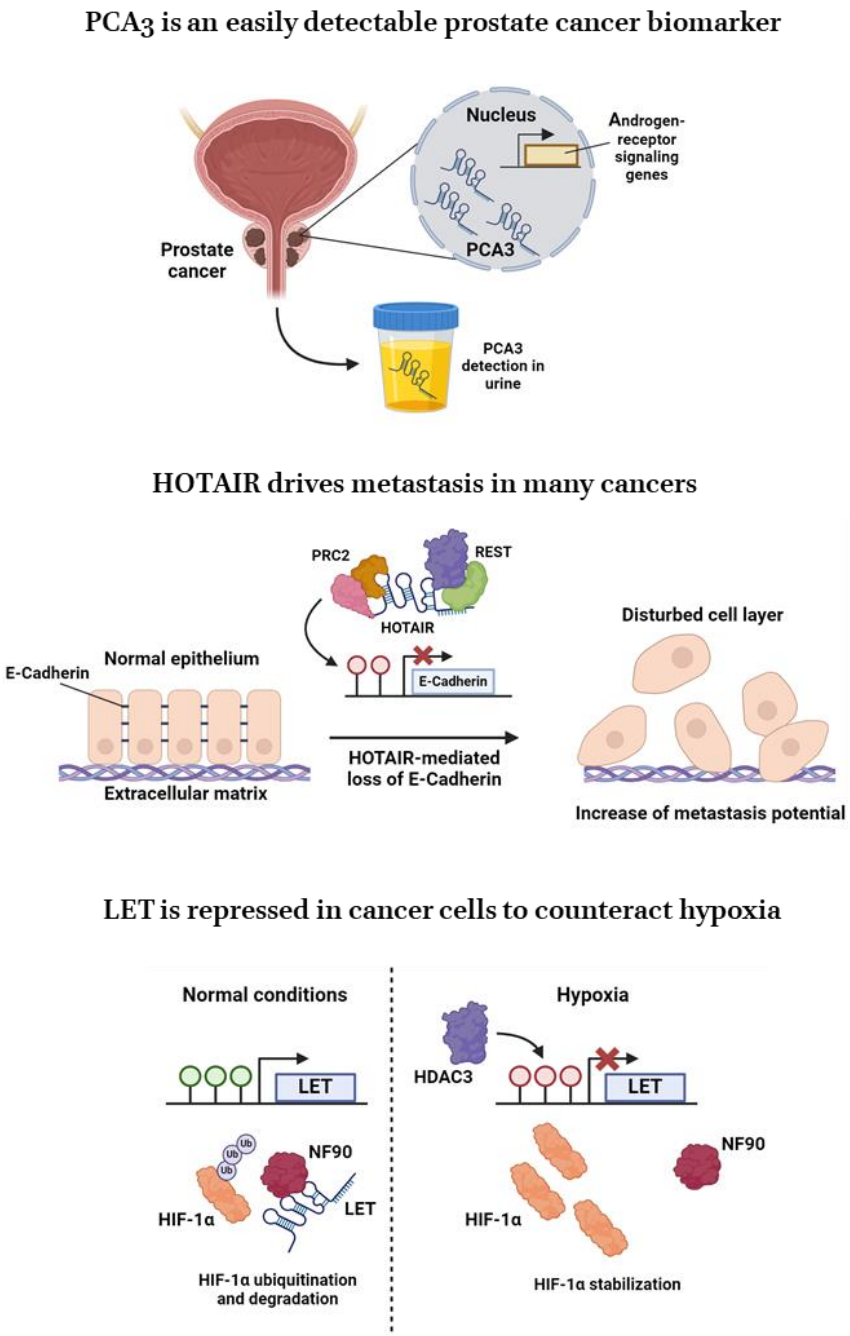


Figure 9. Many LncRNAs are implicated in cancer onset and progression.

PCA3 is a lncRNA specifically expressed in prostate cancer cells and promotes androgen receptor signaling to boost proliferation. PCA3 can be easily quantified in urine to detect tumour cells. HOTAIR is a lncRNA upregulated in many cancers. It is able to repress E-Cadherin transcription through methylation of its promoter by binding chromatin remodelers. E-Cadherin is essential to make cell-cell contacts and maintain cell layer structures. Upon its loss, cells gain mobility, promoting metastasis. In contrast, LET is expressed in normal condition to repress HIF-1α, a factor promoting cell respiration and proliferation. In conditions of hypoxia as encountered in tumours, LET is repressed by HDAC3, stabilizing HIF-1α which enables cancer cells to resist this oxygen stress.

One of the first described cancer-associated lncRNA and potential biomarker is PCA3 discovered in prostate cancer, almost 20 years ago [99]. Authors have shown that PCA3 is expressed up to 66-fold higher in prostate cancer cells compared to normal conditions. Moreover, PCA3 could be detected in urine of patients by RT-PCR, providing an easy method to detect prostate cancer onset. Since then prediction models have been refined and risk-assessment has been improved using PCA3 detection [100]. Nowadays, tests have been developed to achieve an easy and fast detection of PCA3 in urine samples, such as paper-based detection devices [101], providing an easy and efficient way of detecting cancer onset using PCA3 detection making it the first lncRNA used in the clinic as a cancer biomarker. Moreover, the function of PCA3 in cancer cells has also been assessed by depletion studies [102]. PCA3 silencing reduces cancer cell growth and triggers apoptosis showing that this lncRNA is essential for prostate cancer proliferation and survival. In term of mechanisms, PCA3 seems present in the nucleus and its depletion strongly decreases the expression of androgen-receptor signaling related genes, a pathway known as a promoter of prostate carcinogenesis and cell survival [103]. However, its precise mechanism of action remains unknown.

Another well described lncRNA involved in cancer progression is HOTAIR, known to be upregulated in a set of human cancers such as in breast, colon or gastric cancers [104]. The enhanced expression of HOTAIR was detected in gastric cancer cells compared to adjacent tissues and linked with a short overall survival in patients [105]. By depleting HOTAIR, it was observed that cancer cell proliferation was not affected, but cells showed lower colony formation capacity *in vitro* and lower metastatic capacity *in vivo*. Mechanistically, depletion of HOTAIR triggered the expression of E-cadherin, which acts as an adhesion protein that holds cell together [106]. The loss of E-cadherin is common to many forms of cancers as it promotes cell mobility and invasion by disrupting epithelial cell contacts [107]. To explain this effect, it was shown that HOTAIR was able to recruit the PRC2 and REST chromatin modulator complexes at the E-cadherin locus and promote its methylation, lowering the expression of E-cadherin and thus promoting cancer cell metastatic capacity [106]. Similarly, to PCA3, HOTAIR could be considered as a novel biomarker for cancers due to its strong expression compared to normal tissues, even though it is not specific to a particular cancer type and would represent a more general marker.

In contrast to PCA3 or HOTAIR which are upregulated in cancers, some other lncRNA can be downregulated, promoting cancer onset and progression, such as the lncRNA LET (Low Expression in Tumor) [108]. Indeed, LET has been detected at significantly lower levels in hepatocellular, colon or lung carcinomas compared with adjacent tissues. In mice, xenograft experiments with colon cancer cells ectopically expressing LET resulted in lower metastatic capacity. LET pulldown followed by mass spectrometry showed an association with the protein NF90, implicated in the stabilization and transport of many mRNAs [109]. In association with NF90, LET is able to lower the level of HIF1a, a well-known pro-hypoxic factor [110]. In conditions of hypoxia, often found in tumors, the LET locus is deacetylated by HDAC3, repressing its expression and enhancing HIF1a levels [108]. High HIF1a levels are then directly correlated with higher metastatic capacity of cancer cells, enabling cancer progression via inhibition of LET. Strikingly, lncRNAs do not only act as oncogenic factors, they can also display tumor suppressor activity and their expression can be inversely correlated with cancer progression.

On another hand, MALAT1, already mentioned above, is involved in normal cell homeostasis but can also be involved in the biology of cancer cells. MALAT1 formally known as NEAT2 (Nuclear Enriched Abundant Transcript) [80] was renamed MALAT1 (Metastasis-Associated Lung Adenocarcinoma Transcript 1) due to its recognized role in lung cancer progression [111,112]. In these studies, MALAT1 was identified as a prognosis marker for stage 1 non-small cell lung cancer and was upregulated in metastatic samples compared to adjacent tissues. The mode of action of MALAT1 is still unclear, but is was shown that MALAT1 depletion lead to deregulation of 23 metastasis-associated genes without affecting their splicing [111]. *In vitro*, MALAT1 depletion decreased cell motility, in line with its association with metastasis progression.

LncRNAs do not necessarily modulate cancer progression via their expression directly in cancer cells directly, as they can also be expressed by immune cells such as lymphocytes, known to infiltrate tumours and promote cancer cell clearance [113]. For example, the cytoplasmic lnc-EGFR is expressed by regulatory T cells and promote the stability of the growth hormone receptor EGFR by preventing its degradation through direct binding [114]. EGFR is known to be more expressed by tumour infiltrating T cells compared to those present in the adjacent tissue and its upregulation seems

correlated with tumour growth [115]. Thus, in this example, a lncRNA present in the tumour microenvironment may also able to indirectly enhance tumour growth.

The above examples indicate that many lncRNAs often display cell-type specific expression and can be crucial in various disease settings, but are also sometimes necessary in normal tissues and cells. They represent a new class of disease biomarkers, both in cancer (**Figure 9**) and other diseases.

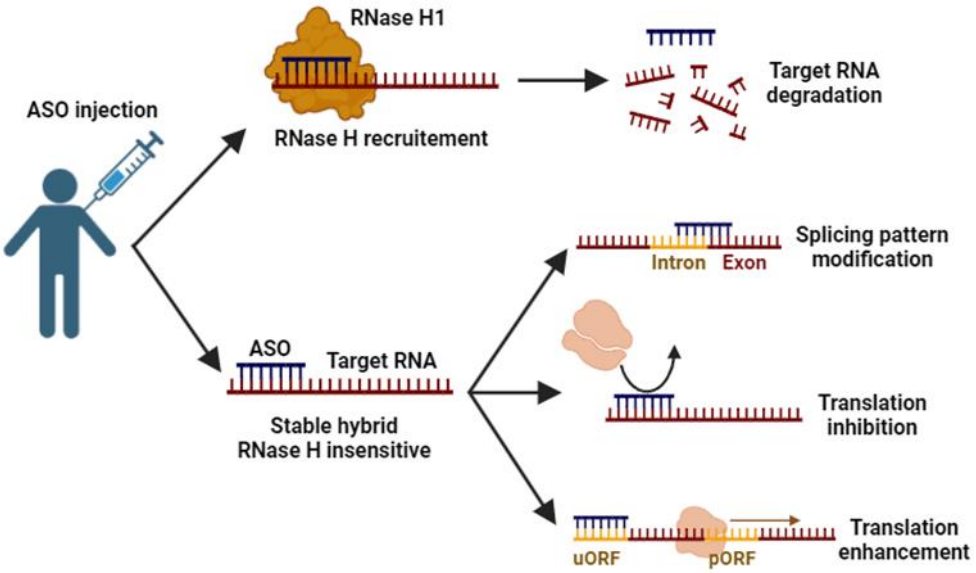


Figure 10: ASO mechanisms of action.

Once injected, the ASO will bind its complementary target sequence leading either to processing of the hybrid by RNAse H1 target degradation, or formation of a hybrid too stable for RNAse H1 degradation. In this case, the ASO can be used to cover splice sites to rescue splicing defects for instance. It can also prevent ribosomes from accessing mRNAs, inhibiting their translation. Conversely, ASOs can also cover uORF to promote translation from the pORF of the target mRNA.

D. LncRNA targeting for therapy

As described above, the critical role of lncRNAs in disease could make them potential therapeutic targets [116]. In this section, we will discuss how they may be targeted particularly in patients. Antisense oligonucleotides (ASOs) are widely considered as versatile strategies for targeting lncRNAs both *in vitro* and *in vivo* [117,118]. ASOs can be defined as synthetic single stranded DNA or RNA molecules, in general between 12 and 25 nucleotides-long, complementary to their target sequence. However, other types of RNA therapeutics exist such as double stranded siRNA [117] which are able to deplete cytosolic RNAs, but in contrast to ASOs, are not imported in the nucleus [119].

1. Mechanism of ASO action

The most straightforward approach to inhibit the action of an oncogenic lncRNA is to trigger its degradation. RNA clearance occurs naturally through the action of RNase in the cell [120]. Selective degradation of RNAs can be mediated by a well described process known as RNA interference (RNAi) [121]. Double stranded RNA present in the cytoplasm is rapidly cleaved into small pieces by the protein Dicer and loaded in a complex called RISC which will then use one of the strands of the cleaved RNA (small interfering RNA or siRNA) to mediate the degradation of complementary RNA sequences. This mechanism is used by the cells to eliminate viral dsRNA [122], but also to process miRNA precursors [60]. Of note, degradation only occurs when the complementarity between the siRNA and the target RNA is perfect, otherwise a mechanism of translation inhibition can occur if the target is a mRNA [121]. Moreover, DNA-RNA hybrids can also be specifically processed by RNAse H1, triggering degradation of the RNA molecule in the hybrid, but keeping the DNA intact [123].

Selective RNA degradation by complementary nucleic acids was quickly exploited by biologists to target cellular RNAs, as demonstrated several decades ago in a work showing thymidine kinase mRNA inhibition by expression of an antisense RNA [124]. In this example, the antisense RNA is expressed with plasmids injected in the nucleus, but it is also possible to directly insert nucleic acids in cells, for instance by transfection or electroporation methods [125]. Once in the cell, the injected oligonucleotide will bind to its complementary target through base pair recognition and trigger its degradation or inhibition through the mechanisms described above [121,123]. These strategies can be

used to target lncRNAs or mRNAs, but also other ncRNA species such as miRNA, for which ASOs known as antagomirs have been designed to trigger their degradation [126].

A pioneering study showed that ASOs could be used to correct splicing events responsible for thalassemia, a blood disorder caused by a lack of hemoglobin [127] which was discussed above. This study highlighted that a mutation in the gene of β-globin caused a splicing defect generating a non-functional protein, responsible for hemoglobin deficiency. The authors managed to correct this defect by designing a DNA ASO targeting the incorrect splicing site, making a stable DNA-RNA hybrid insensitive to RNAse H degradation. In this way, the splicing site was covered, favoring the correct splicing pattern and restoring the normal levels of hemoglobin without degradation of the mRNA.

ASOs can also be used to prevent interaction of a target RNA with a protein exemplified in the context of Myotonic dystrophy type 1, a genetic disease characterized by a shrinking and overall weakness of muscles [128]. The origin of this disease is the presence of GUG repeats on the mRNA DMPK, that promotes retention of the mRNA in the nucleus where it interacts with and sequesters the MBNL1 protein, a splicing regulator crucial for muscle development. The authors designed an ASO antisense to the toxic mRNA, insensitive to RNAse H degradation, that promoted the release of the MBNL1 protein, reversing the defective splicing.

In a similar way, ASOs can also directly impact mRNA translation through direct interaction with their start codon or UTR regions as shown for instance for the BCL2 mRNA [129]. The binding of the ASO seems sufficiently stable to prevent ribosome binding to the mRNA through steric blockade. Protein expression can also be modulated by ASO through poly A site coverage, modifying target mRNA stability and also translation efficiency [130].

Additionally, it is also worth mentioning that ASOs do not always trigger negative regulation of their target RNA, they can for instance enhance translation efficacy. This was demonstrated by using an ASO targeting the mRNA encoding LRPPRC [131]. The idea was to use an ASO covering an uORF in the 5'UTR of the target mRNA. UORFs are present upstream of the canonical or primary ORF (pORF) and are often found in less favorable context for translation initiation [132]. Indeed, they are scanned by the preinitiation complex and can represent modulators of translation efficiency as they are able

to trigger ribosome detachment from the mRNA. Using an ASO to cover the uORF is therefore a means of upregulating the translation efficiency of the mRNA as it seems to be the case for LRPPRC and also other tested genes [131].

However, the major mechanism of action of ASOs is to use ASOs complementary to the target RNA sequence to form DNA-RNA hybrids that are cleaved by RNase H leading to degradation of the target RNA and release of the ASO that can then target another RNA molecule. These mechanisms are represented on **Figure 10**. Nevertheless, irrespective of the chosen strategy, to be successful in the clinic the ASOs ultimately need to be delivered *in vivo*, implying that their structure and characteristics need to be carefully studied beforehand.

2. Nucleotide modifications commonly used for ASO production

As stated above, delivering nucleic acids to an organism is not a straightforward process. Several mechanisms impair ASO delivery, first because cell membrane surfaces are negatively charged, similarly to nucleic acid which would prevent ASO from entering the cell [133]. In addition, many exonucleases exist that can degrade ASOs rapidly, especially if ASO are delivered in body fluids [134]. In addition, immune response must also be taken in consideration as cells are able to recognize foreign nucleic acid and trigger a type 1 interferon response leading to inflammation and potential apoptosis [135]. To overcome this, researchers have been looking to modify classical nucleic acids to make them resistant to nucleases and able to enter cells without the use of transfection agents but also to escape the immune system. This gave rise to three generations of ASOs, products of continuous development by biochemists (**Figure 11**) [118].

The first generation consisted mainly of modifications of the phosphate backbone of the nucleotides by addition of sulfur or nitrogen ions, but also methyl groups. These modifications enhanced ASO stability and resistance to nucleases but still enabled RNAse H cleavage of the targeted RNA [136]. Unfortunately, severe side effects due to unpredicted interactions with RNA binding proteins by this generation of ASOs [137] pushed scientists to develop a novel generation based on different chemistry.

The second generation was developed around modifications of the ribose on the 2' position, sometimes in addition to a phosphate backbone modification. Compared to the previous

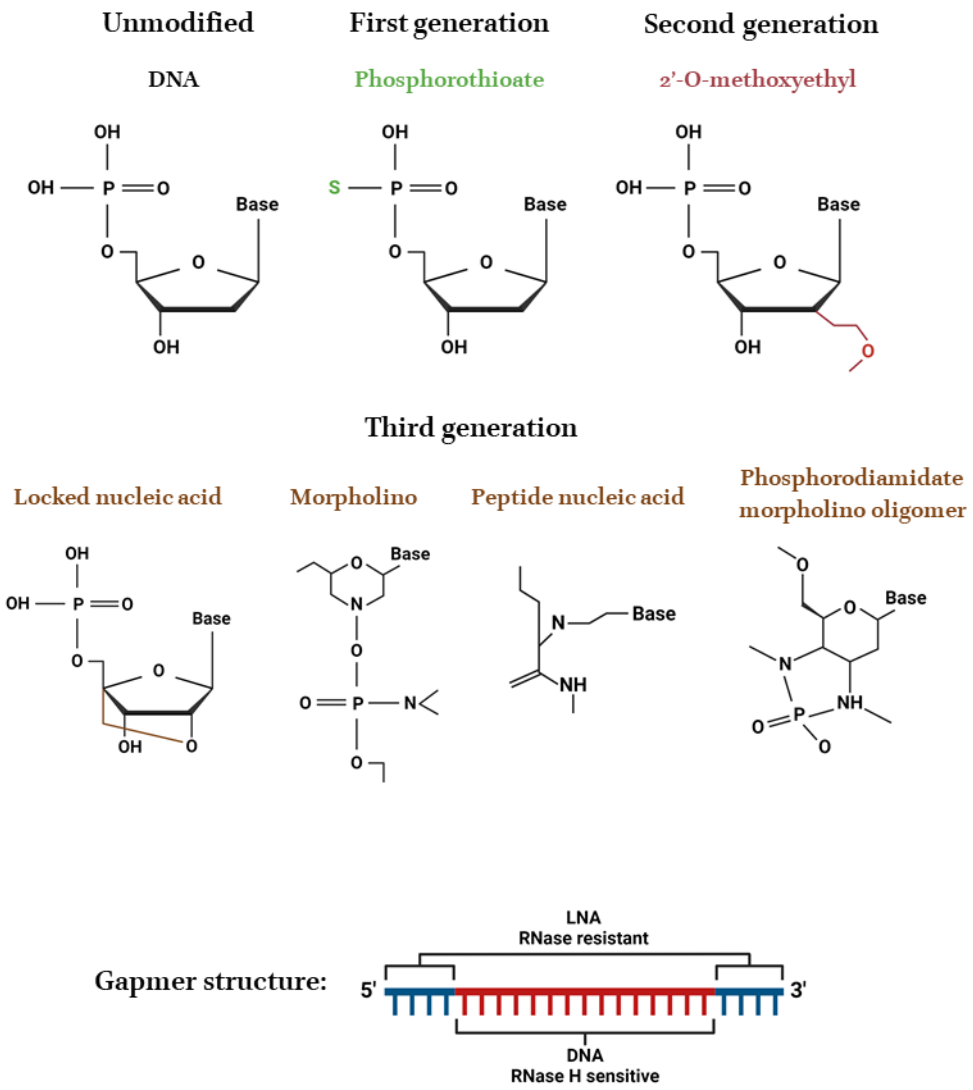


Figure 11: Nucleotide modifications found in ASOs.

Three ASO generations exist, based on the type of chemical modifications. The first is characterized by changes in the phosphate backbone, the second present modifications on the ribose 2' position and the third by heavily modified nucleotides. ASOs can be designated as Gapmers when they are composed of a sequence sensitive to RNAse H cleavage and flanked by modified nucleotides resistance to RNAses.

generation, those ASOs were proven to show enhanced specificity to their target and less toxic, but unfortunately were unable to mediate the degradation of their target trough RNAse H recruitment [138]. Thus, they are of interest in situations involving targeting of an RNA without triggering its degradation.

A third generation of ASO comprise more extensively modified nucleotides. Similarly to the second generation, these modifications greatly improved ASO stability and binding to their target, but are refractory to RNAse H activity [139,140]. However, a class of ASOs known as "GapmeRs" combines chemical modifications such as LNAs to prevent their degradation by exonucleases in their single stranded form on the nucleotides flanking a central sequence sensitive to RNAse H cleavage (**Figure 11**).

Other crucial modified nucleotides found in ASOs are 5-methylcytosines and pseudo-uridines which limit immunogenicity by regulating type I interferon response [141]. Those modifications were notably used to design SARS-CoV-2 RNA vaccines as a way to reduce reactogenicity [142].

In summary, ASO chemistry has been under constant development for several decades to produce stable, deliverable, non-toxic and RNAse H sensitive oligonucleotides for disease treatment. Moreover, new technologies enable researchers to improve ASO composition in the pursuit of more efficient molecules by developing novel modifications [117].

3. ASO-based therapies

Given the improvement of ASO composition, clinical trials began during the 90's for ASO therapies, which gave rise to the first ASO approved by the federal drug agency called Fomivirsen in 1998 (**Figure 12**) [143]. Fomivirsen was a first generation ASO targeting the mRNA of a protein from the human cytomegalovirus, responsible for various symptoms including severe retina infection, potentially harmful especially for babies and immuno-depressed patients [144]. Fomivirsen was injected directly in the eyes, showing an efficient reduction of CMV particles but failed to diffuse to other parts of the body. Ultimately, the drug was withdrawn from the market in 2002, not in relation with potential side effects but because the demand for Fomivirsen fell as new anti-HIV drugs efficiently cured CMV infections [143]. Even if this drug did not last long in the clinic, it proved that ASOs could represent novel therapeutic opportunities. Yet, it also showed that ASOs should be

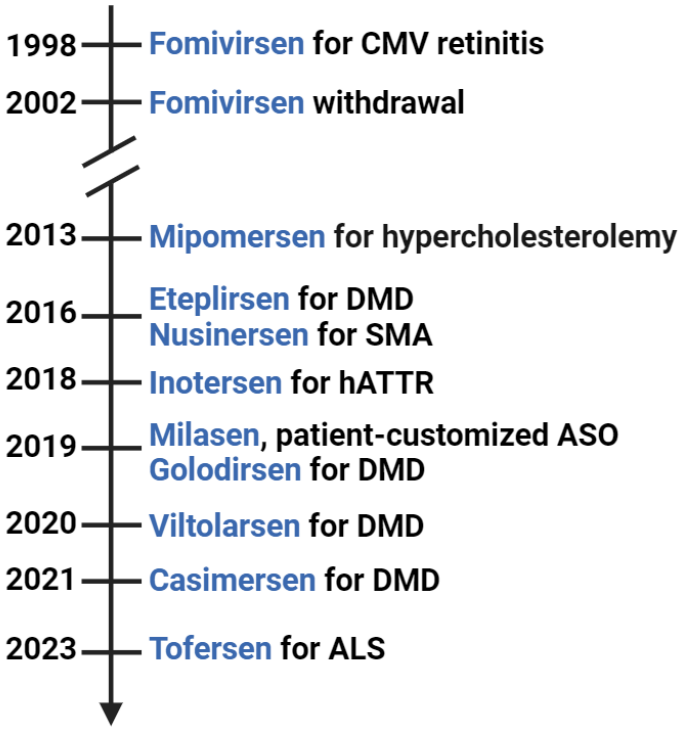


Figure 12: Timeline of FDA approved ASO therapies

All currently approved therapies are used to treat genetic diseases. Of note Milasen is derived from Nusinersen. CMV: Cytomegalovirus; DMD: Duchenne muscular dystrophy; SMA: Spinal muscular atrophy; hATTR: polyneuropathy of hereditary transthyretin-mediated amyloidosis; ALS: Amyotrophic lateral sclerosis.

injected near their target organ or tissue as it seems that their diffusion in the human body is relatively limited.

A decade passed before the approval of a new ASO, Mipomersen in 2013 used to treat patients with familial hypercholesterolemia (FH), correlated with cardiovascular disease [145]. It is a second generation GapmeR delivered by sub-cutaneous injection and then transported to the liver to repress apolipoprotein B mRNA. The inhibition of this protein reduces low-density lipoprotein cholesterol, the main cause of FH. Even though the drug presented satisfactory benefits on the disease, it triggered severe side effects on patients such as hepatotoxicity, inciting the European Medicine Agency to refuse its use in the clinic [146]. The drug is still used in the USA, but patients need to be heavily monitored to limit the adverse effects of the drug. Mipomersen highlighted a downside of ASO use; their tendency to induce hepatotoxicity through off-target binding [147].

The following years were marked by an increase of ASO-based therapies with seven new drugs approved between 2016 and 2023 [117,118]. Some of them are GapmeRs (Mipomersen, Inotersen and Tofersen) while others are characterized by various other nucleotide modifications. Without describing in detail all of these examples, it is striking that all ASOs currently aim to cure genetic diseases, in particular for Duchenne muscular dystrophy [148], but none are used for cancer therapy.

Among those drugs, it is worth mentioning the development of Milasen in 2019, an ASO customized to treat a young girl called Mila affected by Batten disease (a fatal neurodegenerative condition) [149]. A known pathogenic mutation (G to C substitution) was identified in the gene *MFSD8*, a lysosomal regulator, as well as a transposon insertion in the same gene. This genetic event prevented the correct splicing of the 6th exon of the *MFSD8* gene by forming a new splicing site, producing an incorrect peptide. Milasen was designed to cover the abnormal splicing site, restoring the normal splicing pattern of the gene. The structure of this ASO is derived from Nusinersen, an FDA-approved drug usually prescribed for patients with spinal muscular atrophy. It resulted in an improvement of the patient's quality of life and halted disease progression even if the patient eventually passed away a few years after the beginning of the treatment as the disease was already at an advanced stage. Thus, Milasen was the first patient-customized ASO, demonstrating the possibilities for personalized medicine trough ASO therapies [149].

Interestingly, new possibilities are emerging for ASO delivery as it seems that even with modified nucleotides, their stability and transport is still not optimal. The new approaches mainly concern conjugation of ASO with peptides [150] or antibodies [151] to improve their capacity to pass cell membranes and their specificity, but also the development of delivery vehicles such as liposomes [152] widely developed for RNA-based vaccines. In the first case, coupling an ASO with an antibody specific to a surface marker of a specific cell type, such as cancer cells, could represent an opportunity to limit ASO toxicity and enhance their effect as they would be concentrated to cells of interest. Delivery vehicles are often composed of lipids, allowing ASO delivery through cell membrane fusion and protecting them from nucleases. Those new delivery systems are not used so far in the clinic but could become new pertinent ways to improve ASO efficiency *in vivo*.

E. Conclusion of Section 1

In this section, we explored the biology of lncRNAs. LncRNAs are often poorly conserved and less expressed than other RNA classes but display a wide variety of functions acting through a plethora of mechanisms in different cell compartment. Even though they have been described and linked with diseases for several decades, they have gained exponential interest in recent years thanks to the development of novel therapeutic drugs called ASOs with several FDA approved ASOs currently used to treat genetic diseases. Indeed, although the roles of several lncRNAs in cancer have been described, no ASO-based therapy for cancer treatment have been approved by the FDA. Nevertheless, current efforts by several drug companies are directed toward the development of ASO-based therapies for cancer, that hold promise for the future.

In this context, we focused our efforts of the role of lncRNAs in skin cutaneous melanoma. To provide a background to the work in this thesis, we will take a break from lncRNAs to focus on the biology of melanoma, its onset, progression, heterogeneity and current treatments as a first step to explore the roles of lncRNAs and their potential as therapeutic targets in this disease.

Section 2: Skin Cutaneous Melanoma

A. Melanoma cell of origin

1. Melanoblasts to Melanocytes and MSCs

The first step to understand melanoma biology is to describe its origin. Human skin is composed of several layers (**Figure 13**): the hypodermis that contains adipose tissues, the dermis which contains collagen fibers, blood vessels and nerves, and finally the epidermis which is in contact with the environment and composed mainly of keratinocytes [153]. Hair bulbs are also part of the epidermis as they are surrounded by keratinocytes, forming an invagination in the dermis. Melanocytes are cells present at the basal epidermis that arise from neural crest during development [154] through Wnt and BMP signaling pathways [155]. They begin their life as precursor cells called melanoblasts and undergo a dorsolateral migration to colonize the skin, the eyes, the inner ear and also the heart [156]. Different origins are possible for melanoblasts through the differentiation of neural crest progenitors or neural crest-derived Schwann cell precursors (**Figure 14**) [157,158]. The spreading of melanoblasts during development is a tightly regulated process, orchestrated by β-catenin to ensure their even distribution in the skin and to limit their proliferation [159].

In mice, earliest melanoblasts were observed around E 8.5, with the first determinants of melanocyte identity observed between E 9.5 and E 10.5 [160]. Those earliest melanocytes markers are MITF, a transcription factor and DCT, an enzyme essential for melanin production. Among those two, MITF is regarded as the master regulator of melanocyte biology [161–163]. It contains a basic helix-loop-helix leucine zipper domain which binds double stranded DNA by recognizing an M-box motif and belongs to the MiT subfamily which includes other TFs such as TFE3, TFEB and TFEC, able to form heterodimers with MITF [164]. MITF is central to melanoblast and melanocyte function, coordinating many basal cell functions such as proliferation, metabolism, differentiation or senescence by regulating the expression of almost 500 genes [165]. Several MITF isoforms exist based on alternative promoter use [166], the A isoform presenting an ubiquitous expression, H being mainly expressed in the heart and M being exclusive to melanocytes and melanoblasts [167,168]. Given its specificity, MITF-M is considered as the main marker of melanocytic identity. Moreover, it has been

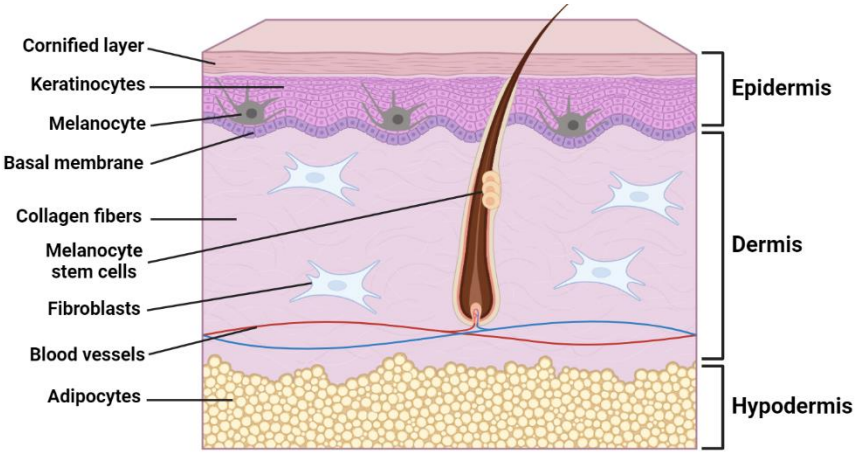


Figure 13: Organization of the skin.

The skin can be divided into three layers: the hypodermis at the bottom, the dermis in the middle and the epidermis on top. The hypodermis contains mainly fat storing and energy producing adipocytes acting as a thermic barrier. The dermis mainly comprises fibroblasts and is rich in collagen. The epidermis is mainly composed of keratinocytes which create a basal, suprabasal and cornified layer that form a barrier protecting the body against the external environment. Melanocytes are also found in the follicular and interfollicular epidermis with the MSCs located in the bulge region of the hair follicle.

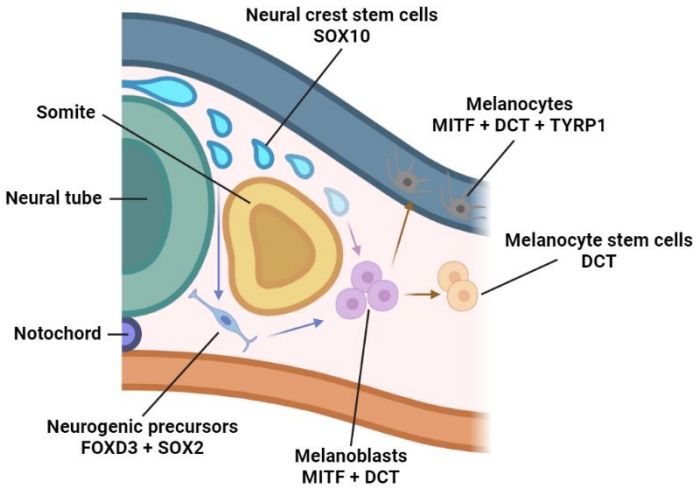


Figure 14: Melanocyte development.

Melanoblasts are the precursors of melanocytes and MSCs. They arise from either neural crest stem cells expressing SOX10 or neurogenic precursors expression FOXD3 and SOX2. The first determinant of melanoblast identity are MITF and DCT. To become functional melanocytes, melanoblasts migrate from the dermis to the epidermis and gain key enzymes for melanin production such as TYRP1. On the other hand, melanoblasts can also become melanocyte stem cells through the loss of MITF, representing a pool of future available melanocytes.

shown that several phenotypes are associated with MITF loss of function in mice and humans, including microphthalmia, loss of pigmentation across the body, osteoporosis and deafness [169]. Mutations of the *MITF* locus also lead to Waardenburg syndrome, characterized by hearing loss and abnormal pigmentation of the skin, hairs and eyes [170].

Following E 11.5, melanoblasts leave the dermis for the epidermis where they will start to proliferate and express melanin synthesis enzymes which will turn them into functioning melanocytes [159]. Mature melanocytes then tend to accumulate in the epidermis in humans or hair follicles in mice and other mammals [171]. They stay in the dermis in mice for instance but are depleted in human dermis [172]. Additionally, it has been shown that keratocytes present in the external layer of the epidermis produce attraction cues for melanocytes whereas melanocytes secrete repulsion cues for each other [173]. This system promotes melanoblast migration to the epidermis and ensures that melanocytes are well dispersed in human skin.

In addition to giving rise to functional differentiated melanocytes, melanoblasts also establish melanocyte stem cells (MSCs), post-embryonic stem cells present in the bulge of hair follicles and responsible for renewing the melanocyte population during the anagen phase of the hair cycle or in the inter follicular epidermis [174]. They exist as a quiescent cell population that do not produce melanin as they lack key factors such as MITF [175], but upon the appropriate stimuli, they undergo asymmetric cell division generating the transitory migrating progenitor cell population and a new population of stem cells. Depletion of the MSC population leads to progressive hair/fur greying in humans and mice respectively [176]. In addition, it has been shown that MSCs are also important in zebrafish where they participate in pigment stripes pattern and fin regeneration [177].

At present, it is still unclear whether melanoma arise from melanocytes or MSCs, as both seem able to transform into melanoma [178,179]. Thus, it is possible that the two populations could give rise to melanoma, even if more work is needed to fully elucidate this question.

2. Roles of melanocytes and MSCs

Now that we have described their origins, we will describe the functions of those cells as they are implicated in several processes. The major function of melanocytes in the body is to enable the pigmentation of the skin, hairs and retina [180]. To accomplish this, they synthetize a pigment called

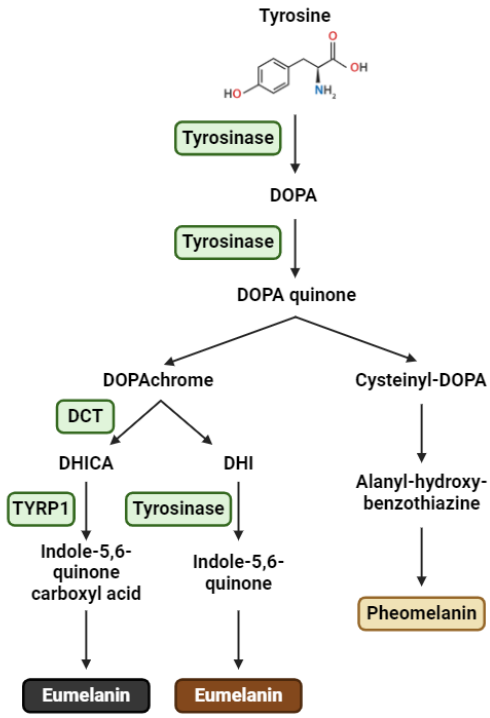


Figure 15: Melanin synthesis pathway.

Melanin is synthetized from tyrosine through a series of reactions catalyzed by enzymes (in green boxes) which are expressed under the control of MITF. Two distinct pathways give raise to pheomelanin (yellow pigment) and eumelanin (dark and brown pigments).

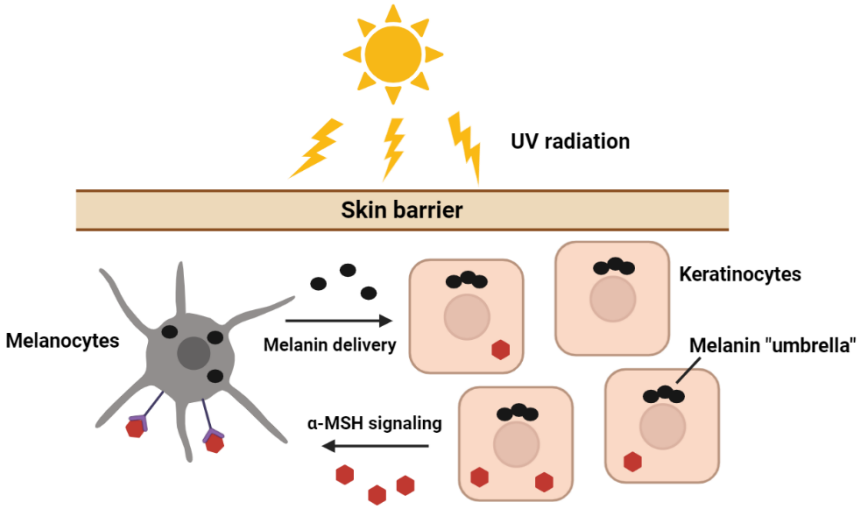


Figure 16: Function of melanocytes in the skin.

Upon exposure to UVR, keratinocytes secret a hormone (α-MSH) which triggers melanin production. Melanin is transported from melanocytes to keratinocytes where it forms a barrier between the nucleus and the UVR source absorbing and deflecting the radiation. In absence of melanin, UVR can cause DNA damage and oxidative stress resulting in cell death and increased mutation rate.

melanin which is transported to adjacent cells, notably keratinocytes, through single layer vesicles known as melanosomes [181]. Indeed, melanocytes possess a specific machinery enabling the production of melanin by expression of a set of genes controlled mainly by MITF [165]. Among those, the tyrosinase enzymes encoded by *TYRP1* and *TYRP2* are essential as they catalyze the first steps of melanin synthesis by processing the amino acid tyrosine [182]. Separate pathways will then produce two types of melanin: the eumelanin which is a dark pigment and the pheomelanin which is yellow/reddish (**Figure 15**) [183]. The ratio between those two pigments determines notably the skin color of an individual. In addition, melanocytes also need a machinery to produce melanosomes in order to transport melanin, which is composed of structural proteins such as PMEL [184] and MART1 [185] but also proteins involved in sorting and trafficking such as AP-3, BLOC1 and OCA2 [186]. Quite interestingly, melanin and melanosomes represent an ancient system of pigment delivery, found in all kingdoms of life [187]. In addition, melanin can also be delivered in the form of melanin granules called melanocores which lack a lipid membrane [188].

Nevertheless, skin pigmentation induced by melanin production is not just a visual characteristic, it is an adaptive trait which is determined by several environmental factors, the main one being UVR exposure [189]. Indeed, UV light of different wavelengths (UV A, B and C) are emitted by the sun and impact living organisms as they can induce DNA damages and lead to inflammation, sunburns and immunosuppression [190]. Strikingly, populations more exposed to UVR tend to have a darker skin, suggesting that melanin production is correlated with sun exposure and DNA damage [191]. Melanin molecules transported into keratinocytes form a barrier next to the nucleus, toward the sun-exposed side of the cells (**Figure 16**) [192]. This barrier will shield the nucleus from UVR in order to prevent DNA damage as melanin is able to absorb and deflect UVR [193]. A single melanocyte is surrounded by approximately 36 keratinocytes, explaining their even repartition in the skin and their limited proliferation as only one cell can ensure melanin delivery to many keratinocytes [194]. Mechanistically, the tanning response triggered by UVR induces p53 activation in keratinocytes, followed by transcription of the *POMC* gene, encoding the a-MSH hormone [195]. This hormone is secreted and binds the MC1R receptor on the surface of melanocytes [196], also able to adjust the ratio of eumelanin/pheomelanin in the skin as it is one of the most polymorphic genes of the human genome [197]. MC1R activation promotes ATP conversion to cAMP which activates PKA, then able to activate CREB and MITF. As seen previously, MITF will then promote the transcription of genes

responsible for melanin production. MSC are also involved in the tanning response as they get activated upon exposure to UVR. After their activation, they migrate to the epidermis through MC1R signaling in order to enhance the melanocyte population [198,199]. However, this process is not the only factor explaining the diversity of pigmentation in humans.

Indeed, the Vitamin D-Folate hypothesis was also proposed to explain the evolution of skin pigmentation in human populations [200]. Vitamin D is an essential nutrient synthetized upon UVR exposure, able to influence approximately 5 % of the gene to regulate calcium levels in the body [201]. Folate is another essential nutrient (known as vitamin B9) essential for DNA synthesis and repair [202], degraded by UV light on the opposite of vitamin D [203]. Thus, a balance between vitamin D and folate is necessary to maintain essential functions of the body. Our ancestors living in Africa needed a protection against folate degradation, linked with elevated melanin synthesis and a darker skin tone. Migrations to northern parts of the world, where UVRs are weaker, lead to a loss of skin pigmentation to facilitate vitamin D production as melanin deflects UV lights.

Thus, melanocytes and MSCs are essential component of the skin as they modulate pigmentation, essential for DNA damage protection and vitamin D / folate balance. Moreover, melanocytes are heterogenous depending on their location [204,205] and are essential for other processes as impaired pigmentation is often linked with additional symptoms in diseases [170]. Conversely, a range of genetic and external factors can lead those cells to abnormal proliferation and cell death resistance, leading to cancer.

B. Melanoma onset

1. Mutations implicated in melanoma development

The first task to understand why cancer arises from melanocytes is to characterize the mutational landscape of melanoma cells which represent the cancer type displaying the highest mutational burden, mostly due to UVR and characterized by a specific UV signature (signature 7) marked by C > T substitutions [206]. Strikingly, approximately 85 % of SKCM carry mutations in the MAPK pathway, either in the kinase BRAF and or in the GTPase NRAS but also in NF1, a negative regulator of the pathway (**Figure 17**) [207]. The other 15 % are considered as triple-WT, often driven by uncharacterized mutations. The MAPK pathway is central to melanocyte biology as it controls proliferation and senescence but also MITF activity and is composed of a cascade of kinases, activated through RTKs growth factor receptors [208]. It has been shown that the mutation V600E in BRAF [209], present in half of SKCM cases, is a gain of function mutation leading to constitutive BRAF activation and hyper activation of the MAPK pathway stimulating melanocyte proliferation [210]. Even if the vast majority of SKCM seem indeed driven by UVR [211], *NRAS* and *BRAF* mutations do not display classical UVR signature mutations [212], suggesting the implication of other additional stresses. Indeed, it has been shown that oxidative DNA damage and UV-induced photoproducts are common in melanocytes which lack efficient repair pathways due to melanin interference, potentially creating those mutations not displaying an UVR signature [213,214].

Of note, SKCM is not the only melanoma subtype (**Figure 18**). Several others exist based on their site of origin such as mucosal melanoma where half also display mutations in *BRAF* and *NRAS* [215]. In contrast, these mutations are much less prevalent in acral melanoma [216] (from non-glabrous skin such as the palm and soles, driven by *KIT* mutations) and uveal melanoma [217] (from the retina, driven by *GNAQ* and *GNA11* mutations [218]). This observation is related to the fact that *BRAF* and *NRAS* mutations are indirectly UV-driven and become rare in regions of the body less exposed to the sun which is the case for non-skin cutaneous melanomas [219].

Nevertheless, it has been observed that BRAF V600E is not sufficient to transform melanocytes as the mutated cells enter oncogene-induced senescence, arresting their proliferating [220,221], which suggests that other mutations are necessary for senescence escape and melanoma progression. The

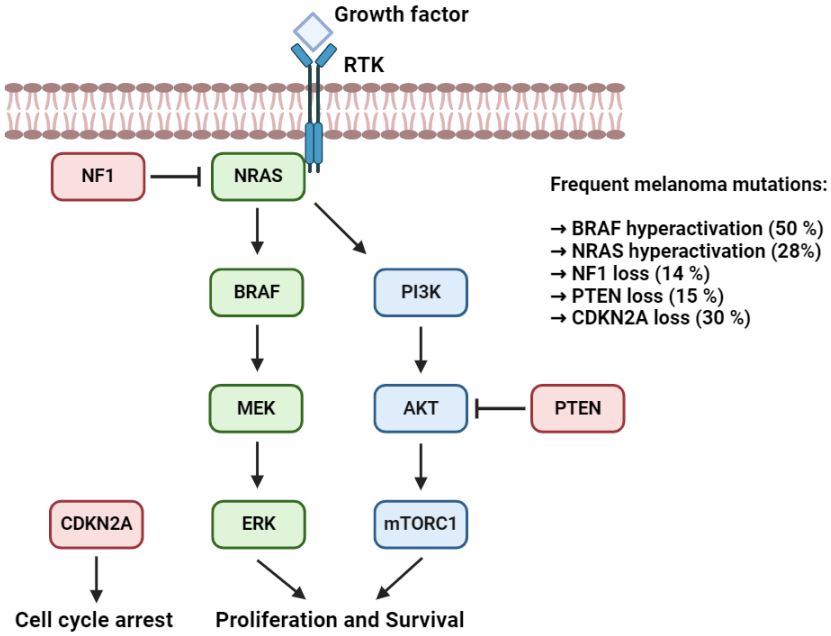


Figure 17: MAPK pathway and mutations leading to SKCM.

The MAPK pathway is composed of a cascade of kinases, activated by growth factors through receptor tyrosine kinases that stimulate cell proliferation and survival. The upstream GTPase NRAS also activates the PI3K/AKT/mTOR pathway to stimulate cell growth. The BRAF and NRAS are the most frequently mutated proteins in melanoma cells, leading to hyperactivation of the pathway. Several additional proteins are also often mutated in melanoma (NF1, PTEN and CDKN2A) overcoming senescence and leading to uncontrolled cell proliferation and survival.

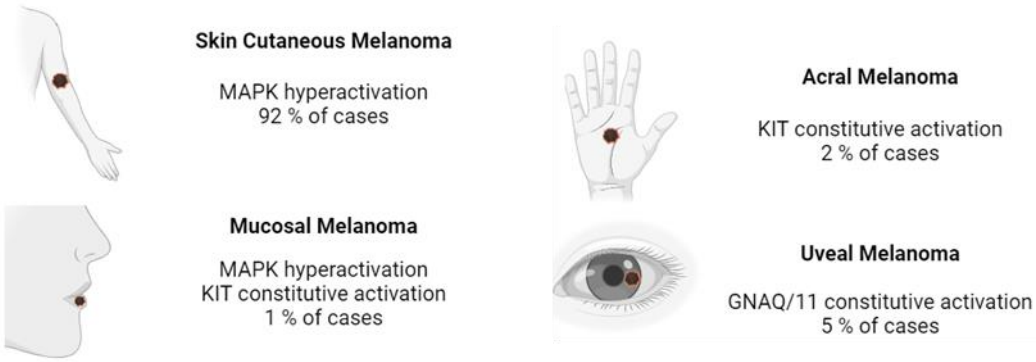


Figure 18: Melanoma subtypes and mutational characteristics.

The vast majority of melanoma cases are SKCM, characterized by MAPK pathway mutations as seen above. Mucosal melanomas also often display MAPK mutations but show higher rate of mutation in KIT than SKCM which is also the case for acral melanomas. Uveal melanomas have a unique mutational profile as mutations in GNAQ/11 are found in almost 90 % of cases. Numbers were provided by the melanoma research alliance.

first identified genes came by looking at genetic mutations which predispose to melanoma such as in *MITF* [222] or in cyclin *CDKN2A* which is lost in half of melanoma cases [223]. CDKN2A is responsible for cell cycle arrest in G1 and G2 phases by interacting with other cyclins such as CDK4 and CDK6[224]. The loss of this cyclin in addition to a mutation in the MAPK pathway could then enable melanocytes to escape senescence and abnormally proliferate. Additional tumour suppressor genes such as *PTEN*, *TP53* and *CDH1* are also often lost in melanoma cells, enhancing senescence escape and cell motility [225]. Of note, *TP53* is the most frequently altered gene in human cancer, as 50 % of human tumours carry loss of function mutations of *TP53* [226], highlighting its role as a master tumour suppressor gene implicated in many metabolic processes [227].

In summary, a set of mutations, rather than a single event, drives melanocyte toward melanoma through the hyper-activation of the MAPK pathway and the loss of tumor suppressor genes due to genetic alterations, UVR-induced damage and other stress factors.

2. Disease progression

Having defined the major mutations responsible for melanoma initiation we can see how they contribute to the different stages of the disease. Indeed, melanoma mutations seem to appear sequentially, driving disease progression step by step (**Figure 19**) [228]. This process of sequential progression was proposed already in 1969, where melanoma disease was assigned a score (Clark) based on thickness of the lesion [229].

As mentioned above, the first event is generally a mutation in the MAPK pathway leading to melanocyte proliferation. A cluster of melanocytes will then undergo senescence and induce skin pigmentation in a limited zone, which is referred as a nevus [220] , benign skin lesions occurring in most individuals during their life. Subsequently, either due to familial predisposition or DNA-damage induced mutation, different tumour suppressor genes can be lost such as CDKN2A or PTEN described above [225]. A cell in the benign nevus will then escape senescence and proliferate to become dysplastic, pushing against the basement membrane separating the epidermis from the dermis, leading to the radial growth phase where cells will gain in motility and start to spread within the epidermis[158]. Amplification of the *CCND1* gene and activating mutations leading to TERT expression are often found during this step, promoting melanoma cell proliferation by impairing the cell cycle

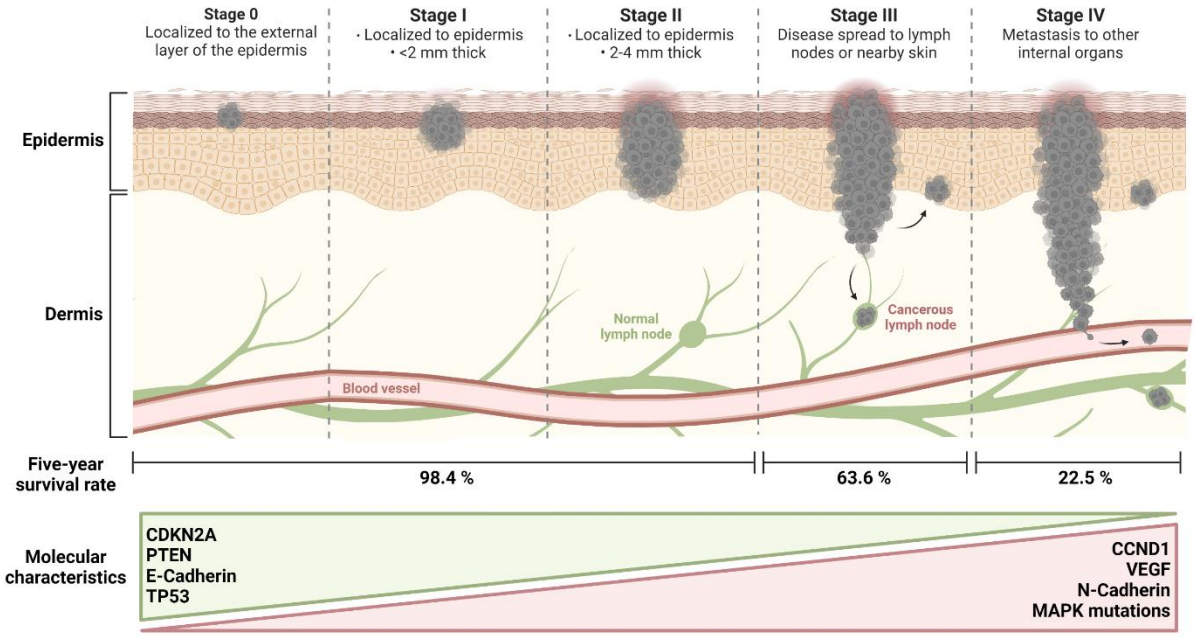


Figure 19: Melanoma progression.

In the first stages, melanoma is localized to the epidermis is characterized by MAPK pathway (mainly BRAF and NRAS) mutations. Additional mutations will then lead to the loss of tumor suppressors (CDKN2A, PTEN, E-Cadherin and TP53) and gain of proteins enhancing tumour invasion (CCDN1, VEGF, N-Cadherin). While early stage melanoma can be easily removed through surgical resection, the five-year survival rate of melanoma patients greatly decreases if they are diagnosed in stage 3 or 4 when the melanoma cells have started to spread to adjacent tissues and distant organs.

and maintaining telomere length and immortalization [230]. The tumour can then progress to a vertical growth phase, characterized by invasion of the dermis [231]. This event is promoted by the loss of E-cadherin, involved in epithelial cell contacts [107] and the expression of N-cadherin promotes cell motility and metastasis formation [232]. By colonizing the dermis melanoma cells will eventually reach blood vessels and enter the circulation. Circulating cells will then reach different organs where they will be able to proliferate and form metastasis [233]. Those secondary tumors are the major cause of death as they are found in lungs, liver or also the brain. In addition, metastases are transcriptionally different from primary tumours, based on their anatomical site [234,235].

Of note, only a minority of melanomas derive from a pre-existing nevus, the majority have already accumulated multiple mutations that allow them to bypass senescence and arise *de novo* on the skin and seem more aggressive than their nevus-derived counterparts [236]. A checklist (7PCL) has been implemented to distinguish benign nevi from melanoma, mostly based on the color and shape of the lesion as melanoma tend to be irregular and darker than benign nevi [237].

3. Classification

As discussed above, melanoma initiation involves a sequential series of events to lift several molecular barriers and progress. An official staging nomenclature based on histopathology has been proposed for melanoma to characterize the progression of the disease [229]. The Clarck classification is mainly based on tumour size, thickness, the invasion of adjacent lymph node and the presence of metastasis. Stage 0 is also known as melanoma *in situ* and concerns tumours limited to the epidermis of the skin. Stage 1 describes tumours less than 2 mm thick and stage 2 tumours less than 4 mm thick. Signs of ulceration (breaking of the skin) can appear from stage one but are not necessarily present.

From stage 3, adjacent skin portions and lymph nodes are invaded by melanoma cells, irrespective of tumour size. Lymph nodes are organs responsible for immune cell production and are implicated in the elimination of cancer cells by collecting them through the lymphatic tube network [238]. Lastly, when cancer cells reach distant organs, the disease progress to stage 4, irrespective of primary tumour size or adjacent lymph node invasion. According to the Melanoma Research Alliance, the five-year survival rate of melanoma patients is 98,4 % for localized tumours (stage 0, 1 and 2), 63.6 % for regional tumours (stage 3) and 22.5 % for distant tumours (stage 4).

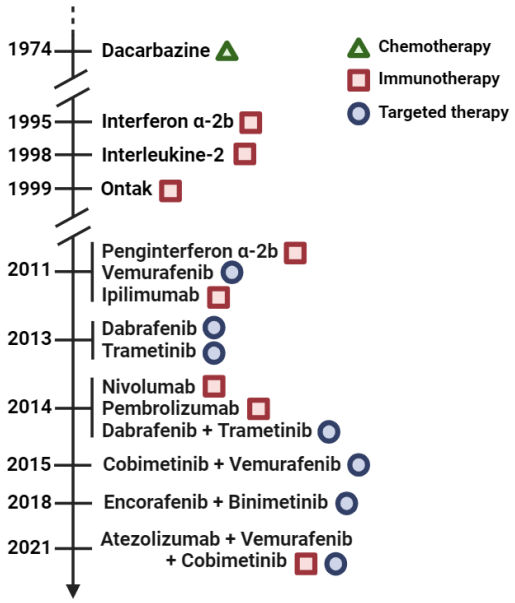


Figure 20: Timeline of FDA-approved drugs to treat melanoma.

Drug development against melanoma went through several periods marked by increased response rates and PFS. Initial chemotherapy was accompanied by severe side effects and poor responses. Melanoma therapeutic management was revolutionized around 2010 with the arrival of targeted therapies against the MAPK pathway and subsequently by modern immunotherapies that show much better response rates but still lead to important side effects

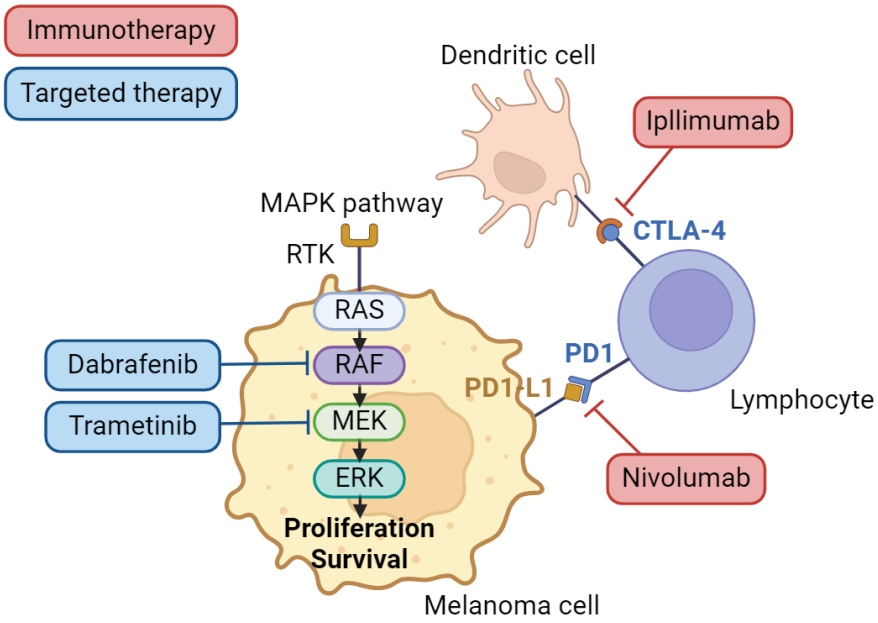


Figure 21: Mechanisms of immune-checkpoint and targeted therapies.

Targeted therapies specifically impact mutated kinases of the MAPK pathway in melanoma cells. Several kinases of the pathway can be targeted especially MEK1/2 in the most recent therapies as it is downstream of NRAS and BRAF. Modern immunotherapies work by injection of antibodies targeting receptors at the surface of immune cells such as lymphocytes promoting their anti-tumoral activity.

C. Treatments for melanoma patients

Even although melanoma accounts for only 1 % of all skin tumours, it represents 80 % of skin cancer-related deaths due to its aggressivity [239]. In total, 325 000 new melanoma cases and 57 000 deaths were estimated for the year 2020 worldwide and those numbers should rise in the following decade [240]. Even if localized melanoma is easily curable by surgical resection, advanced stages greatly impair patient prognosis inciting scientists to develop therapies against advanced stage melanoma. In the past, the first line treatment option for melanoma patients was chemotherapy with the chemical Dacarbazine approved by the FDA in 1974 (**Figure 20**) [241]. This alkylating agent had a very limited effect on cancer progression with less than 5 % of patient complete response rate. It has a cytotoxic effect on cells and often results in severe side effects for patients. Thus, researchers were prompted to discover new therapeutics means against melanoma, more efficient and less toxic, which gave raise to immunotherapies and targeted therapies [242].

1. Immunotherapy

It was suggested very early in cancer research that cancer cells were not the only determinant of disease progression, as the immune system also has a central role in response to cancer [243]. Notably, several immune cell types such as lymphocytes are able to invade tumours and trigger cancer cell death, providing a powerful system against cancer onset [244] . One way to kill cancer cells and to boost the immune environment is the secretion of cytokines such as interferons which have antiproliferation and anticancer proprieties [245]. Thus, a therapy consisting of IFN α-2β injection was approved in 1995 against melanoma. The treatment improved patient prognosis by reducing cancer cell proliferation [246]. Nevertheless, only a minority of patients responded to IFN treatment [247]. This therapeutic approach was later improved with the approval of Peginterferon α-2β in 2011, a mix between interferons and PEG, enabling a more efficient response [248]. Another cytokine displaying antitumoral activity is interleukin-2 (IL-2) which is able to expand the lymphocyte population. Again, only a limited response was observed upon IL-2 injection, which was approved by the FDA in 1998 [249]. Lastly, another strategy consisted in the elimination of an immune population called Tregs which are correlated with a poor clinical outcome as they limit the action of other immune cells [250]. A drug known as Ontak was approved in 1999, constituted by Il-2 coupled with a toxin, specifically depleting Tregs [251], even though the true effects of this drug have now been questioned [252].

Those treatments showed a rather poor response rate and a high percent of patients were refractory to them. The field of melanoma immunotherapy greatly progressed in the 2010's with the development of new therapies consisting in the injection of antibodies to boost immune cell action on cancer cells. The first one, ipilimumab, was approved in 2011 and is an anti-CTLA-4 antibody [253]. The receptor CTLA-4 is an inhibitory checkpoint present on T cell surface that blocs T cell activation and prevents autoimmune diseases [254]. These antibodies work as antagonists in order to promote T cell expansion and tumour infiltration. Strikingly, combination of ipilimumab and Peg-IFN produced an overall response of 40 % in melanoma patients [255].

Subsequently, another checkpoint called PD-1/PD-L1 was targeted by immunotherapy in melanoma. PD-1 is presented on the surface of T cells whereas PD-L1 is located on tumour cell surface. The interaction between the ligand and the receptor prevents T cell activation, and thus cancer cell elimination [256]. Nivolumab is an anti-PD-1 antibody which was approved in 2014 for melanoma therapy [257]. It greatly enhanced patients' prognosis and showed a PFS of 6.9 months whereas ipilimumab displayed a PFS of 2.9 months and chemotherapy a PFS of 2.2 months [258]. Another anti-PD-1 approved therapy is Pembrolizumab in 2015, showing a lower toxicity than ipilimumab for instance [259]. Of note, it is possible to combine immunotherapies, which improves response rate to 58 % with a combination of nivolumab and ipilimumab compared to 52 % with nivolumab alone and 34 % with ipilimumab alone [260].

Thus, immunotherapy (immune checkpoint therapy) represents a powerful and versatile approach and is now the first line treatment for metastatic melanoma patients. Nevertheless, a significant proportion of patients are unresponsive to immunotherapy and severe side effects can occur as those drugs tend to provoke autoimmune diseases [260,261]. Notably, immune cells infiltration is a direct predictor of response to immunotherapy as cold tumours (no immune infiltration) are globally insensitive to treatment on the opposite of hot tumours (immune infiltration) [262].

2. Targeted therapies

Immunotherapies are not the only treatment possibility for advanced melanoma. As mentioned above, the majority of melanomas harbor mutations in the MAPK pathway, leading to its

hyperactivation [207]. This observation led researchers to design inhibitors against the kinases implicated in this pathway as a way to reduce melanoma cell proliferation and senescence escape.

The first drug of this kind approved by the FDA was vemurafenib in 2011, which was designed to target BRAF V600E, found in many melanomas [263]. Vemurafenib treatment greatly improved metastatic patient prognosis with an overall survival of 84 % at 6 months compared with 64 % with a group treated by chemotherapy. Moreover, a response rate of 48 % was observed with vemurafenib treatment compared to 5 % for chemotherapy. Other drugs targeting mutated BRAF have been developed such as Dabrafenib [264] or Encorafenib [265]. Those different drugs led to an increased response rate and better survival chances for melanoma patients [266] but present also two major downsides. The first is that only patients with melanoma presenting a BRAF mutation can benefit from those treatments as half of melanoma harbor other mutations [225]. The second is related to resistance mechanisms leading to re-activation of the MAPK pathway due to upregulation or mutations of other kinases (CRAF, MEK1/2) of the MAPK pathway or of other pathways, limiting clinical benefits [267,268]. Notably, BRAF gene amplification was seen in 20 % of tumours resistant to BRAF inhibitors [269] along with overexpression of CRAF and ARAF [270], restoring ERK signaling.

A possibility to overcome those limitations is to target the MAPK pathway downstream of BRAF, for instance by using MEK inhibitors such as Trametinib approved in 2013 [271]. The following year a combination of Trametinib with Dabrafenib has been approved by the FDA as it showed better effects than targeted monotherapy with an objective response rate of 76 % compared to 54 %, respectively [272]. A similar combination was approved in 2015 with Cobimetinib (another MEK inhibitor) and Vemurafenib, which also improved the PFS of melanoma patients [273]. Nevertheless, those treatments are often accompanied by adverse effects, especially with drugs combination [266]. Of note, recent efforts have been made to combine immunotherapy with targeted therapies in melanoma, even if no significant improvements have been observed so far compared to classic targeted therapy combination [274].

In summary, a great deal of effort in the last two decades have been made to treat patients with metastatic melanoma. Targeted therapies and immunotherapies are available for melanoma patients, alone or in combination, which greatly improved patient PFS and response rate compared to chemotherapy or older immunotherapies (**Figure 21**). Conversely, a consequent portion of

patients are still refractory to immunotherapy and resistance mechanisms impair targeted therapies efficiency [267,275]. In addition, it was shown that resistance to targeted therapy in melanoma also confers cross-resistance to immunotherapy [276]. Those mechanisms of resistance are still not fully understood, but many studies point out to the role of cancer microenvironment heterogeneity in this process [235], and will be discussed below.

D. Heterogeneity of melanoma tumours

1. Melanoma cell types

As stated above, the tumour microenvironment is complex, composed indeed of tumour cells but also of immune cells as well as stromal cells of adjacent tissues. In addition to this diversity, each cell population is also composed of different cell types, including cancer cells which can be found in different states [277,278]. The use of new technologies such as scRNA-seq which consist of the dissociation of the tumour into single cells in solution followed by RNA-seq has revolutionized the analyses and understanding of tumour heterogeneity [279]. Each cell is labeled individually, enabling the analysis of the single cell transcriptome revealing the complexity of the cell population instead of bulk RNA-seq which assess the overall transcriptome of a heterogenous collection of cells.

Before scRNA-seq, two major melanoma cell subtypes were already identified [278,280]. The first displays high expression of MITF, seen previously as the main melanocytic marker [171]. Many genes associated with MITF are also expressed in these cells such as *TYR*, *PMEL* and *MLANA*, all implicated in melanin production. Conversely, a second melanoma cell population was identified with low or no MITF expression. This MITF-low population instead express AXL, a receptor tyrosine kinase [281] and NGFR, a stem cell marker [282]. The high MITF subtype was characterized as proliferative whereas the low MITF subtype represent invasive cells (**Figure 22**).

One of the first scRNA-seq on melanoma tumours was performed in 2016 from 19 patient tumours, enabling the sequencing of almost 5000 cells [283]. The well-known MITF-low and MITF-high subtypes were identified in this analysis. This study also highlighted that invasive melanoma cells represent the minority of malignant cells in these tumours, a characteristic potentially undetectable using bulk RNA-seq. Of note, five non-malignant cell populations were found in this analysis: T cells, B cells and macrophages of the immune cell compartment as well as endothelial cells and CAFs from the stroma. Interestingly, authors found that the AXL-high invasive melanoma cell population was enriched upon treatment with BRAF inhibitors, suggesting that invasive cells are more resistant to targeted therapy than proliferative cells. CAFs abundance in tumours was also associated with a greater AXL-high melanoma cell proportion, suggesting that CAFs could be involved in the

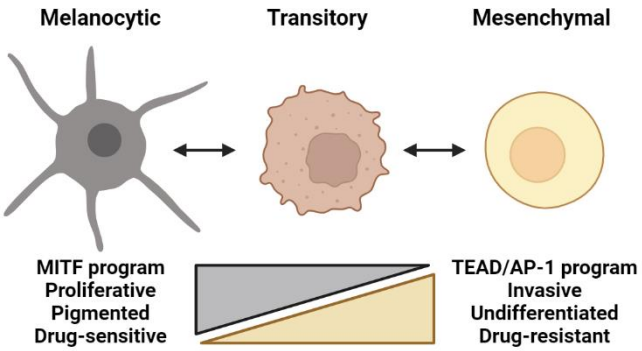


Figure 22: Melanoma cell heterogeneity.

Melanoma cells can be mainly classified between the more differentiated melanocytic state and the most undifferentiated mesenchymal cells. Melanocytic express MITF, SOX10 and all the enzymes necessary for melanin synthesis. They are highly proliferative and globally sensitive to therapies. Mesenchymal cells rely on the TEAD/AP-1 program and express AXL, TCF4, JUN and ZEB1. They are highly invasive but show lower proliferation. They are associated with therapy resistance and are enriched in tumours after treatment. Cells that express intermediate levels of these markers can be considered intermediate or transitory state cells.

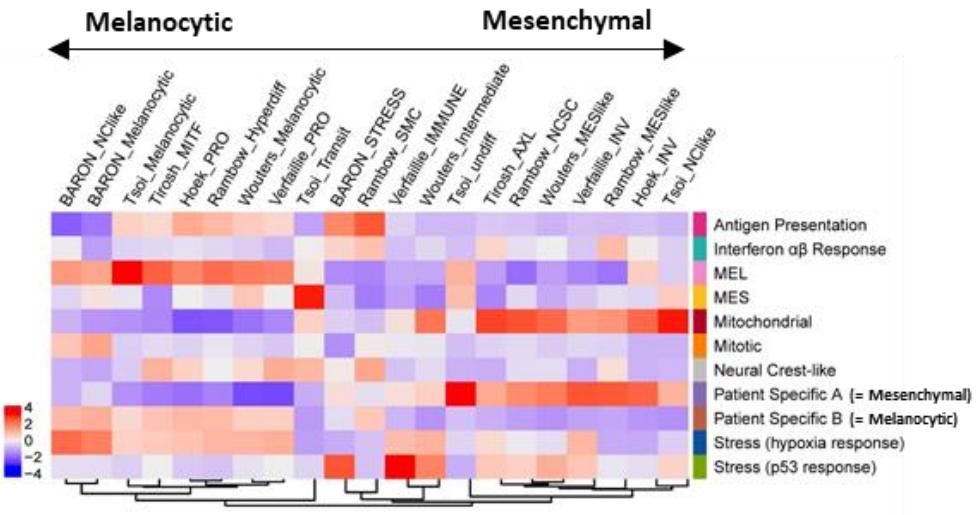


Figure 23: scRNA-seq studies identified several melanoma cell types.

From Pozniak *et al.,* 2024**.** Through the years, different studies highlighted melanoma cell populations based on their transcriptomic programs and classified them on a scale from melanocytic to mesenchymal. In their recent study, the Marine group identified novel melanoma subtypes that fall on this axis. They show that the MEL, Patient-specific B and hypoxia clusters represent melanocytic cells. Antigen presenting, MES, interferon response and neural crest-like cells represent an intermediate phenotype. Mitochondrial, Patient specific A and p53 response cells are more mesenchymal. Finally, the mitotic cluster was not enriched in markers of the other cell types.

appearance of this malignant cell population. Heterogeneity of T cell populations was also observed in the different biopsies.

Two subsequent studies were conducted in 2018, enabling the discovery of new melanoma cell types in tumours. In the first study [284], the authors designated these subtypes from melanoma cell lines and patient tumours: melanocytic, transitory, neural crest like and undifferentiated. The melanocytic population is similar to the MITF-High population from Tirosh [283], presenting a strong expression of *MITF* and other related genes such as *SOX10* and *CTNNB1*. On the opposite, the AXL-High population was also found and annotated as undifferentiated cells, with a high expression of *AXL*, *SOX9* and *EGFR*, corresponding to a set of genes known to be upregulated upon SOX10 loss [285]. They seem characterized by a loss of melanocytic identity with the absence of expression of many key melanocyte genes. In addition to those previously described subtypes, the authors presented a population designated as neural crest-like, characterized by the expression of SOX10 and NGFR, but with low MITF expression that resembled the transcriptome of neural crest cell during development before their differentiation into melanoblast trough MITF expression [154]. Transitory cells were marked by an intermediate expression of many of those markers including MITF, SOX10, AXL and NGFR, suggesting that they could mark intermediate states between melanocytic, neural crest and undifferentiated cells. Interestingly, melanocytic and transitory types were linked with a proliferative phenotype whereas undifferentiated and neural crest-like cells were correlated with an invasive phenotype [280]. Again, it was observed that dedifferentiated melanoma cells were resistant to targeted therapy, but were sensitive to ferroptosis-inducing drugs, proposing a way to target this cell population.

In a similar manner, the work carried by Rambow and colleagues [286] described analogous melanoma subtypes in PDX treated or not with MAPK inhibitors. The MITF high population was designated as pigmented cells as melanin production was elevated through the MITF pathway. An invasive population was characterized displaying a low MITF and expression of previously identified invasion markers such as SLIT2, BGN and TNC [280]. The same neural crest -like cluster was identified, again with expression of SOX10 and NGFR, but loss of MITF. In contrast to the study of Tsoi however, transitory subtypes were not found in Rambow analysis. Instead, authors identified a cluster of 'starved'-like melanoma cells (SMCs) which express a set of genes correlated with nutrient-

deprivation [287], strongly enriched upon treatment. These SMCs could represent a pool of drug resistant cells, able to differentiate into proliferative melanoma cells at later stages.

Moreover, scRNA-seq analysis was also performed on a collection of cultured melanoma cells, enabling the refinement of subtype markers [288]. In this study, 10 different cell transcriptomes were assessed, showing six pigmented or melanocytic lines, three undifferentiated or mesenchymal-like lines, and one neural-crest cell like. Among the six melanocytic cell lines, three were further designated as intermediate cells, displaying diminished expression of MITF and SOX10 and resembling the transitory subset from Tsoi [284].

Furthermore, the study of Karras and colleagues showed that melanoma growth is hierarchically organized [289]. Perivascular niches contain melanoma stem-like cells exhibiting an embryonic transcriptional program, originating from dedifferentiation of melanoma cells. Endothelial cells present in these niches seem responsible for tumour growth signaling. This observation echoes with the concept of cancer stem cells (CSCs), proposing that a limited number of malignant cells fuel tumour growth as already observed in other models such as in breast, colorectal or brain cancers [290]. CSCs are highly plastic and could represent a reservoir of tumour cells able to differentiate upon various stimuli. Authors also characterize the TF PRRX1 as an essential driver of metastasis onset as it is specifically expressed in undifferentiated melanoma cells [289].

More recently, a broader scRNA-seq analysis of almost 60 000 tumour cells coming from 20 patients was reported [291]. The previously identified melanoma subtypes were retrieved, corresponding to melanocytic, mesenchymal and neural crest-like populations (**Figure 23**). Moreover, novel subtypes were characterized. An antigen presenting cluster showing elevated expression of multiple HLA class 1 and II genes, responsible for cell surface antigen presentation [292]. Another cluster was named interferon response and showed elevated levels of interferon response genes which are activated upon exposure to interferon secreted by adjacent immune cells [245]. Two additional populations known as mitochondrial and mitotic were described and seem routinely found in many human biopsy samples [293]. Of note, the mitochondrial cluster was referred to as the "low quality" cluster showing no particular enrichment for molecular pathways [294]. Lastly, two stress response clusters were found, one more linked to hypoxia response often found at the tumour center which is oxygen-deprived [295] and the other linked with p53 response linked with cell apoptosis and

DNA damage response [296]. In addition, two patient specific clusters were retrieved, one corresponding to a highly proliferative signature and the second correlating with a highly invasive signature.

Together these studies reveal that melanoma cells can adopt different cell states with distinct gene expression signatures and biological properties. Melanoma tumours can be segregated into two major cell types mainly based on MITF expression [297], those expressing MITF can be considered melanocytic, proliferative or pigmented, whereas cells not expressing MITF are designated as mesenchymal or undifferentiated. Cells displaying intermediate levels of MITF can be considered as transitory or intermediate, whereas cells expressing SOX10 but not MITF are designated as neural crest-like. This classification is analogous to that seen in other carcinoma characterized by Epithelial to Mesenchymal Transition (EMT), with a loss of identity, gain of invasive capacity and cell plasticity [277,293]. In the case of melanoma cells, it is more accurate to speak of a pseudo-EMT as melanocytes are not epithelial cells. Further subtypes can be identified based on the expression of different markers, showing specialization of melanoma cells in the tumour in response to external stimuli such as treatment, immune or stromal cell influence. Again, cells with the characteristics of the immune-like HLA-expressing cells and those with the IFN-response signature as well as those with a stress response signatures were all found in other solid tumours suggesting that despite the differing nature of the cells of origin, analogous tumour cell states can be found in tumours perhaps due to the influence of the TME. Future work may further refine the complex composition of melanoma tumours, potentially identifying additional novel and sparse malignant cell subtypes.

2. The tumoral micro-environment is a driver of therapy resistance

As mentioned above, current melanoma therapies are limited due to resistance mechanisms of melanoma cells. Several cancer cell subtypes exist in the tumour [278], which lead researchers to speculate that not all melanoma cells respond to treatment in a similar manner (**Figure 24**).

As stated above, the first scRNA seq study of Tirosh [283] supported the idea that mesenchymal melanoma cells may be resistant to therapy as AXL was known to be implicated in intrinsic resistance to RAF and MEK inhibition [281]. To investigate this, they analyzed RNA-seq data from six melanoma biopsies before and after treatment with BRAF inhibitors and observed a shift toward the

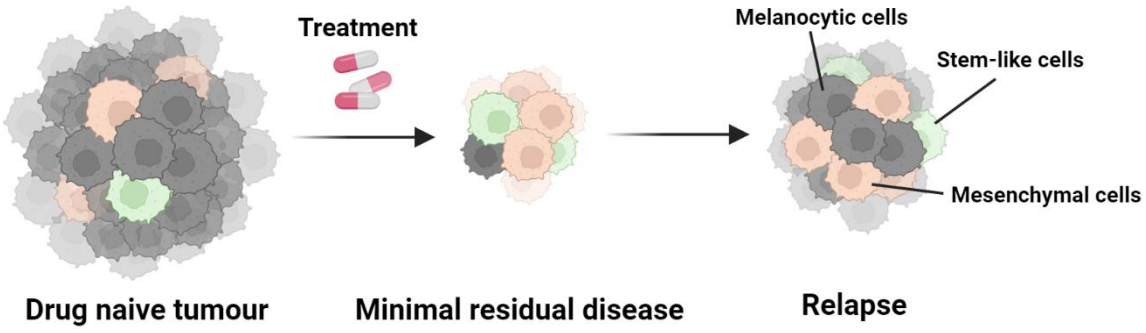


Figure 24: Melanoma cell types involved in therapy resistance.

Melanoma tumours are mainly composed of drug-sensitive melanocytic cells. Upon treatment, resistant neural crest stem cell-like and mesenchymal cells persist and contribute to minimal residual disease. Through phenotype switching some of those cells can recover proliferative capacity driving tumour growth and leading to relapse.

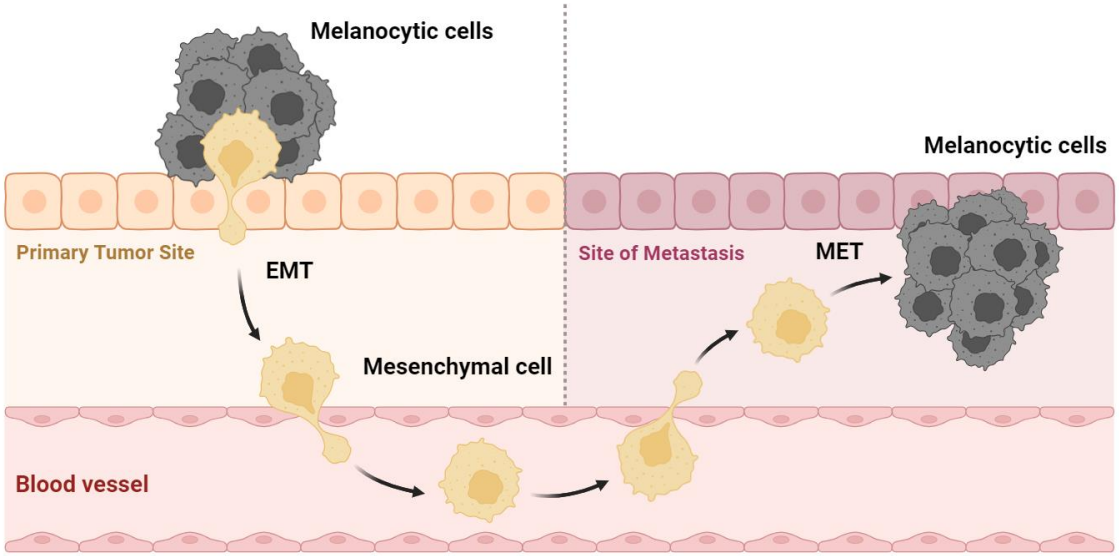


Figure 25: Phenotype switching enables cancer cell dissemination.

As stated above, primary tumours are mainly composed of melanocytic cells. Upon external stimuli such as nutrient deprivation or immune infiltration, melanocytic cells can undergo pseudo-EMT leading to invasion through the basal lamina and dissemination to the draining lymph node and adjacent tissues. If close to blood vessels, mesenchymal cells can pass the endothelial barrier and enter the circulation. They can then exit the blood vessels to invade distant organs. Some cells can undergo the reverse MET switching back to a melanocytic-proliferative phenotype and forming metastasis.

AXL program in the treated biopsies. Similar results were obtained with melanoma cells in culture. Interestingly, melanoma cells expressing AXL were also linked with resistance to anti-PD1 treatment , suggesting that this subtype represent a global drug-resistant state [298]. CAFs are also implicated in this resistance as they seem to be related with the presence of mesenchymal cells [283] and are responsible for an immune suppressive environment [299]. Several previously mentioned studies observed the same pattern where the mesenchymal melanoma cell population tended to increase upon different treatments, and this may also be the case for neural-crest like cells [284,286,288,291].

Mechanistically, it has been proposed that TCF4, a mesenchymal cell TF, could be central to immunotherapy resistance as its activity suppress antigen presentation and IFN signaling resulting in a dampened immune response [291]. These authors also found that treatment of mesenchymal cells with BET inhibitor, which recapitulates TCF4 silencing, sensitizes cells to BRAF and MEK inhibitors. Similarly, another major resistance driver is the aryl hydrocarbon receptor which is activated upon treatment [300]. Use of an AhR inhibitor in combination with BRAF inhibitor reduced the number of resistant cells and prevented tumour growth. ATF4, the major driver of the integrated stress response (ISR), also participates in therapy resistance as it is able to activate AXL and to reprogram the translation of melanoma cells towards an invasive state [301]. Thus, mesenchymal melanoma cells seem to be the major drivers of therapy resistance as they express specific factors that promote immune escape and invasion whereas melanocytic cells are more sensitive to drugs in general. Nevertheless, a key question arises following those observations: how are resistant mesenchymal melanoma cells able to drive tumour growth if they are not able to highly proliferate?

A potential explanation would be the ability of cancer cells to dynamically and reversibly switch between different phenotypes in response to external stimuli. Evidence for such phenotype switching in melanoma has been provided by multiple studies [288,302]. For instance, a study showed that the depletion of MITF in melanocytic melanoma cells rendered them drug resistant, but impaired their proliferation [302]. Mechanistically, it has been shown that MITF directly downregulates the expression of AP-1-driven genes encompassing many invasion markers by antagonizing C-Jun, a TF promoting inflammation-induced dedifferentiation of melanoma cells [303,304]. This model explains the appearance of mesenchymal cells upon treatment as they would not arise *de novo* but would represent a population of drug-sensitive cells that switched their transcriptional program from the

SOX10/MITF network towards a resistant phenotype through the AP-1 program [305]. This process, discussed before, is known as EMT [306] and can also occur in the opposite sense, where mesenchymal cells can switch back to a melanocytic phenotype, known as MET (Mesenchymal to Epithelial Transition), which occur in many different cancers [307]. MITF and the kinase CDK7 also repress a program dependent on the TF GATA6, expressed in mesenchymal cells [308]. When active, GATA6 promotes the expression of AMIGO2, essential for melanoma cell survival [309], and ABCG2, an efflux pomp implicated in drug resistance [310].

Additionally, phenotype switching may also help to account for melanoma disease progression in the body (**Figure 25**). Primary tumours are known to be mainly composed of melanocytic cells [291], highly proliferative but with poor invasive capacity. Some of those cells, close to blood vessels, could undergo transcriptional changes, giving them a mesenchymal phenotype leading to tissue invasion [289]. Upon their arrival in a new environment, those invasive cells could reverse their phenotype from mesenchymal to melanocytic, giving rise to metastasis [311]. This plasticity is in part due to the activation of an embryonic migration program in melanoma cells, normally repressed in melanocytes [312]. This view of melanoma progression is different from the model were a set of sequential mutations, seen previously (*CDKN2A*, *BRAF*, *TP53*…) [225], is responsible for the transition from radial to vertical growth phases. Mutations represent irreversible genetic events whereas phenotype switching is mainly regulated through reversible, epigenetic modifications [288,313]. Thus, a combination of both genetic and epigenetic events is responsible for melanoma progression through loss of tumour suppressor genes and phenotype switching toward invasive and drug-tolerant cell states [314].

In summary, melanoma tumours are composed of a variety of cancer cell types which participate in disease progression. Many of these cells are highly plastic and can switch from melanocytic to mesenchymal and *vice versa in vivo* [302]. Melanocytic cells express the MITF program, are highly proliferative and drug sensitive whereas mesenchymal cells express the TCF4/PRRX1/AXL program, are highly invasive and resistant to treatments. This potential for a dynamic switch between melanoma cell types impairs treatment efficiency and promotes relapse. An important physiological difference between melanocytic and mesenchymal cells is their metabolism. Understanding the basis for these differences could be a key to uncover new therapeutic targets.

E. Melanoma metabolism at the center of phenotype switching

1. Metabolic pathways in melanoma cells

It was observed already in the 1920s that cancer cells tend to use a peculiar metabolic pathway compared to normal cells which consist of aerobic glycolysis, described by Otto Warburg and hence known as the Warburg effect [315]. Usually, in presence of oxygen, cells tend to privilege oxidative phosphorylation to produce ATP through glucose conversion to pyruvate and its utilization by the mitochondria, known as aerobic respiration (**Figure 26**) [316]. This process is way more efficient (36 ATP molecules per molecule of glucose) than glycolysis which uses available glucose to produce lactate and 2 molecules of ATP. However, cancer cells tend to perform glycolysis even in presence of oxygen [317], as demonstrated in melanoma cells [318]. However, the precise benefits for cancer cells of using this pathway remain debated [319], but are likely linked to the fact that intermediates in the glycolytic chain are used for nucleotide and amino acid synthesis that are essential in rapidly proliferating cells [320].

In melanoma, the major driver of glycolysis is the MAPK pathway which is known to be hyper-activated in this context [210]. Promotion of the MAPK pathway upregulates MYC and HIF1α signaling which promote the glycolysis program through the activation of several factors such as GLUT1, LDH or HK2 [321]. HIF1a is also known to be activated by the PI3K/AKT/mTOR pathway [322] which is often activated in melanoma cells [323]. Thus, melanoma cells highly depend on glycolysis compared to melanocytes [324]. This addiction is also present as the MAPK pathway downregulates OXPHOS, promoting glycolysis use even more. At the molecular level, it was shown that oncogenic BRAF downregulates MITF, known to promote the expression of PGC1α, a transcription factor implicated in mitochondria biogenesis [325].

Nevertheless, as described above, some subtypes of melanoma cells maintain high MITF activity and thus express PGC1α promoting OXPHOS in those cells [326]. The same study showed that a high OXPHOS signature is also linked with an overall poor prognosis for melanoma patients. The precise mechanism driving OXPHOS activity in those cells remain unclear as mTOR has been shown to promote glycolysis or OXPHOS, depending on which mTOR complex is activated [327]. Thus, it seems that melanoma cells can be also discriminated based on their metabolism, melanocytic cells

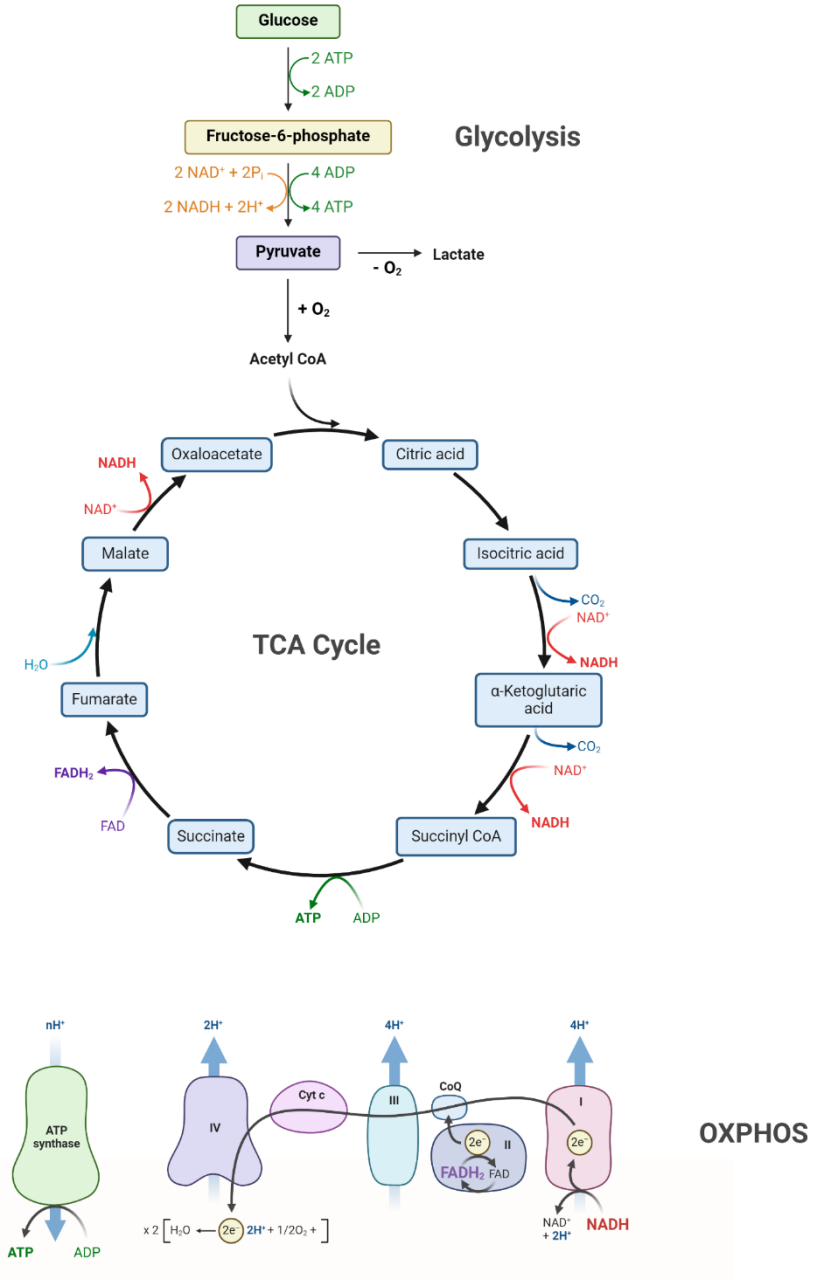


Figure 26: General cell metabolism.

Cells need ATP to fuel essential molecular processes, especially translation. To generate ATP cells uptake glucose from their environment that is catabolized into pyruvate through glycolysis. In presence of oxygen, pyruvate is converted into Acetyl-CoA which is used by the TCA cycle to produce NADH and FADH2. Those two molecules are then used by the electron transport chain to form a proton gradient used by the ATP synthase to convert ADP into ATP. Under hypoxic conditions pyruvate is rather converted into lactate, a process known as fermentation. In cancer cells, glycolysis to produce lactate is favored compared to acetyl-CoA, even in presence of oxygen, a phenomenon known as the Warburg effect.

expressing MITF depending more on OXPHOS than mesenchymal cells which rely more on glycolysis [328].

As melanoma cells can switch phenotype during cancer progression [302], could the associated metabolic switch be used by cancer cells to drive tumour growth and invasion? Indeed, it has been shown that melanoma cell metabolism is influenced by the microenvironment. In case of glutamine or glucose limitation, melanoma cells activate ATF4 which suppresses MITF and downregulates global translation, but maintains a subset of translationally active mRNAs to promote ISR [301,329]. However, the ATF4 program is also able to activate AXL to promote an invasive phenotype while also activating OXPHOS as this pathway is more efficient than glycolysis, especially when glucose is limited [330]. Additionally, lactate produced by glycolysis can also be secreted in the microenvironment which acidify the extracellular Ph, promoting invasion and suppressing the immune system [331]. This property of melanoma cells could enable them to change their environment and their motility when resources are lacking. Conversely, hypoxia conditions which promote HIF1a activation, drive glycolysis use as oxygen is not available enough to fuel aerobic respiration [322]. The E3 ubiquitin ligase MDM2, a known p53 repressor, also translocates to the mitochondrial matrix upon hypoxia, repressing *MT-ND6* to reduce complex I activity and cell respiration [332]. It also promotes mitophagy by binding PARKIN, depleting mitochondria and reducing respiration [333]. Thus, a model proposes that tumours contain hypoxic niches far from blood vessels [334] and composed of cells depending mainly on glycolysis while cells closer to blood vessels would rely more on OXPHOS (**Figure 27**) [335]. Moreover, it seems that both pathways could fuel each other and thus, that melanoma cells change their metabolism dynamically in response to external cues, a process known as “metabolic symbiosis”, also described in other cancers [336].

Thus, cancer cells are able to modulate their metabolism based on their environment and nutrient availability. Stress linked with hypoxia and nutrient deficiency promote invasion and glycolysis as a way for cells to attain a more suitable environment for their proliferation. Once in presence of oxygen and nutrients, cells can promote the OXPHOS pathway to enhance their growth, limiting their motility in return. In summary, metabolism switch is a consequence of environmental changes and drives cancer progression by influencing cancer cells invasion and proliferation.

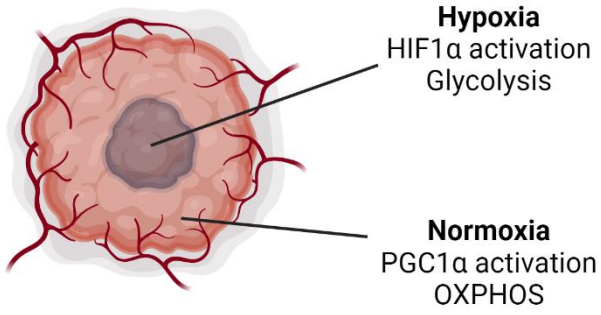


Figure 27: Tumour metabolism.

The tumour center far from blood vessels often represents a hypoxic environment. To survive, cells induce HIF1α which promotes ATP production via glycolysis and represses MITF expression. The periphery of the tumour or regions close to blood vessels increased oxygen represses HIF1α and promotes mitochondrial respiration and cell proliferation through MITF and PGC1α activation.

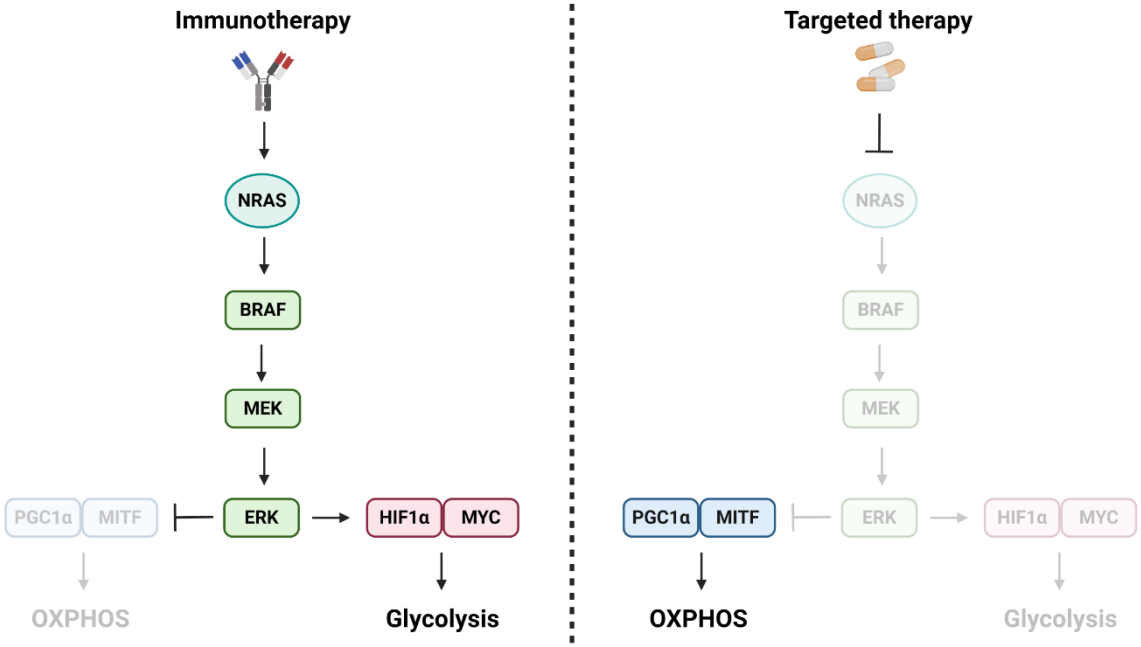


Figure 28: Effects of treatment on melanoma cells metabolism.

MAPK activation promotes glycolysis over OXPHOS as it activates HIF1α and reduces MITF activity leading to reduced PGC1α levels and mitochondrial respiration. Immunotherapies were shown to activate pathways promoting glycolysis in cancer cells, but also in immune cells. On the other hand, MAPKi inhibits glycolysis inducing a metabolic shift towards OXPHOS via altered mitochondrial homeostasis and upregulated PGC1α expression.

2. Metabolic changes upon melanoma therapy

So far, we have seen that melanoma tumours are composed of different cancer cell types, relying on different metabolic programs and responding differently to therapeutic drugs. Consequently, it is interesting to assess what happens to cell metabolism upon treatment and if this could represent an opportunity to elaborate new therapies (**Figure 28**).

It has been shown that CTLA-4 and PD-1 activation can inhibit the PI3K-AKT pathway [337]. Thus, immune checkpoint inhibitors which block CTLA-4 and PD-1, tend to promote the PI3K-AKT pathway and glycolysis whereas they decrease OXPHOS activity [338]. Glycolysis is also used by effector T cells which could explain why immune checkpoint inhibitors create an immunosuppressive environment. It is then possible to speculate that combining immunotherapy with glycolysis inhibitors could be an efficient way to deplete melanoma cells but would be challenging as several immune cell types also rely on glycolysis [339].

Conversely, MAPK inhibitors will downregulate glycolysis, promoting OXPHOS [340]. More precisely, OXPHOS induction has been observed in almost half of BRAF-mutated melanomas resistant to targeted therapy [341]. PGC1α expression has also been linked with therapy resistance and poor patient prognosis, as discussed before [326]. Different studies also underlined that some melanoma cells are highly dependent on OXPHOS by TFAM activation, a mitochondrial TF, and not by the classical MITF/PGC1a axis [342]. Other pathways and metabolic regulators can be considered such as the E4F1 TF, known to control the expression of genes involved in mitochondria homeostasis and conversion of pyruvate to acetyl-CoA to fuel the TCA cycle [343,344]. Furthermore, several studies attempted to couple MAPK inhibitors with OXPHOS inhibitors [341,342,345,346], showing that mitochondrial inhibitors were able to overcome resistance of melanoma cells to targeted therapies, showing a promising therapeutic opportunity.

One of those drugs currently explored in combination with melanoma therapies is metformin, an inhibitor of complex 1 of the electron transport chain in OXPHOS which is usually given to diabetic patients [347,348]. Its impact on melanoma progression is actively disputed as it was shown that vemurafenib + metformin treatment displays a strong anti-proliferative effect on melanoma cells [349] whereas another study showed that metformin stimulates tumour growth [350]. In any case, OXPHOS inhibitors are rather unspecific drugs that also impact the tumour microenvironment and are

notably known to drive the emergence of autoimmune diseases as they enable chronic lymphocyte activation [351].

To overcome this, it would then be important to identify specific therapeutic targets in melanoma cells that contribute to the OXPHOS pathway. LncRNAs could be interesting candidates given their specific expression and their essential roles in cancer development [19]. Interestingly, two lncRNAs have been recently identified as melanoma-specific and are implicated in mitochondrial function, named SAMMSON and LENOX.

F. Melanoma-specific lncRNAs impact cell metabolism

1. SAMMSON

The first identified melanoma-specific lncRNA is Linc01212, renamed as SAMMSON (Survival Associated Mitochondrial Melanoma Specific Oncogenic Non-coding RNA) [352]. Its locus is located adjacent to that of MITF and comprises its own promoter. Interestingly, SAMMSON expression is regulated by SOX10, but is only expressed in melanoma cells and not in melanocytes. SAMMSON is also not expressed in other cancer types. ASO-mediated SAMMSON depletion reduces clonal capacity of melanoma cells and triggers their apoptosis, showing that this lncRNA is essential for their proliferation and survival. Mechanistically, the authors characterized the protein interactome of SAMMSON by biotinylated-oligonucleotide pulldown followed by mass-spectrometry, a technique known as RAP-MS and often used to understand lncRNA functions [79]. This experiment identified two protein partners for SAMMSON: the exoribonuclease XRN2 and p32 (= C1QBP) which is linked with mitochondrial processes. XRN2 is an RNA binding protein interacting with several RNAs and is involved in their degradation [353] , whereas p32 is more specific to SAMMSON and is interesting given its role in cancer and OXPHOS regulation [354] .

SAMMSON depletion had no effect on p32 levels but depleted it from the mitochondrial fractions, promoting instead its nuclear localization and reducing activity of OXPHOS complexes potentially caused by a reduced level of several mitochondrial proteins implicated in respiration. Indeed, mitochondrial deficiency can cause cellular stresses leading to cell death [355]. One of these is known as mPOS (mitochondria precursor over-accumulation stress) and occur when mitochondrial membrane potential is lost, leading to an impaired protein import and thus, an accumulation of mitochondrial protein in the cytosol [356]. Indeed, the majority of mitochondrial proteins are nuclear-encoded and translated in the cytosol before being imported inside mitochondria [357]. Signs of mPOS occur upon SAMMSON depletion as nuclear-encoded mitochondrial proteins such as ATBP and SDHA accumulate in the cytosol and even in the nucleus, leading to cell death [358]. This concept of mPOS should be kept in mind for the results obtained on LENT where similar observations were made in our study.

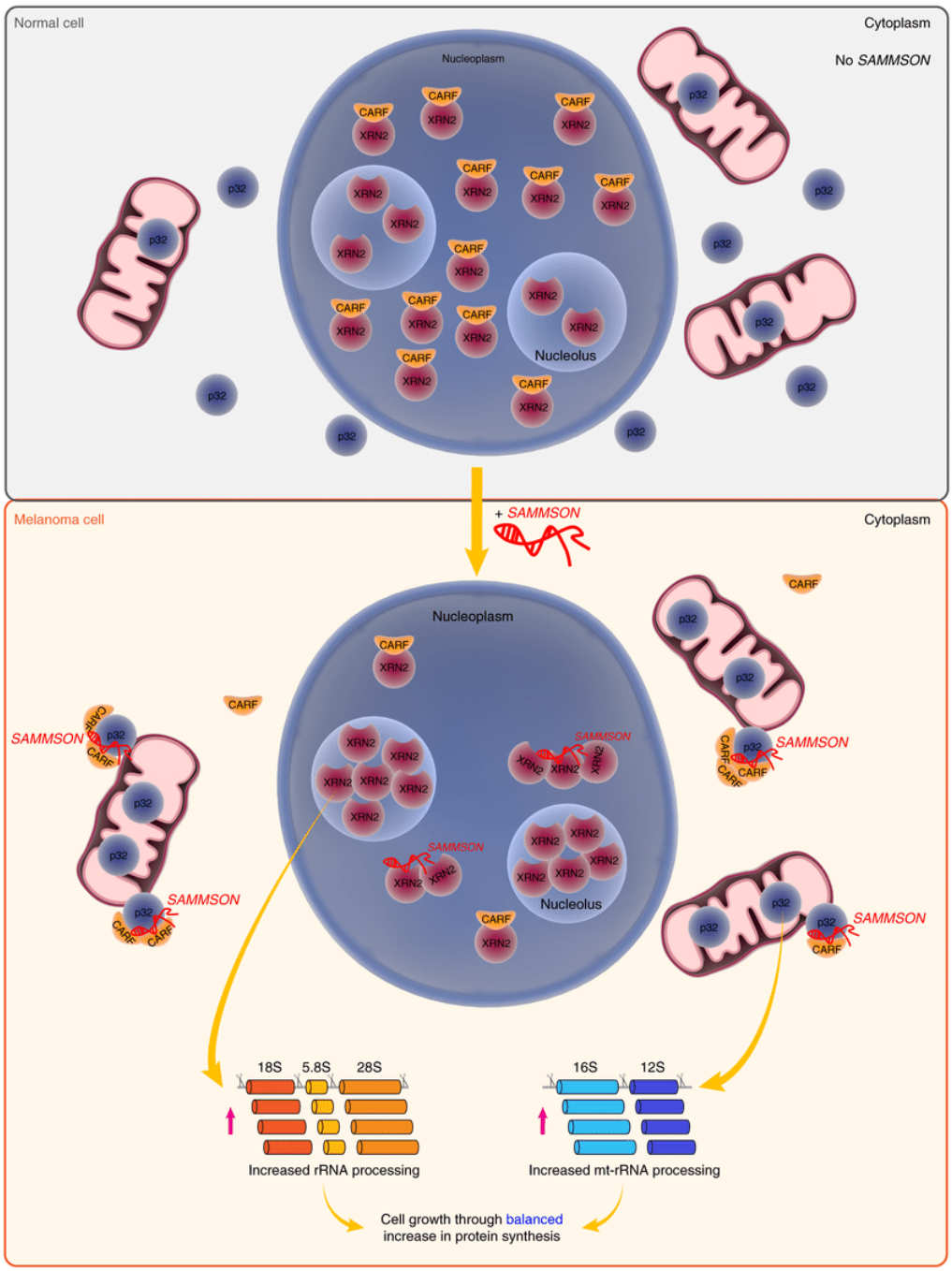


Figure 29. SAMMSON mechanism of action.

From Vendramin et al., 2018. In normal (SAMMSON-negative) cells, CARF controls the nuclear localization of XRN2 by sequestering a pool of XRN2 in the nucleoplasm. In the context of melanoma, SAMMSON expression promotes the interaction of CARF with p32 in the cytoplasm at the expense of the CARF-XRN2 interaction, thus favoring p32 mitochondrial localization and XRN2 nucleolar localization. By modulating these interactions SAMMSON determines a balanced increase in rRNA maturation and protein synthesis in the cytosol and mitochondria. As a result, SAMMSON promotes cell growth.

In agreement with the idea that MAPK inhibitors block glycolysis and promote a shift to OXPHOS [345] SAMMSON depletion coupled with dabrafenib treatment impaired PDX growth more efficiently that the treatment with dabrafenib alone. Thus, targeting SAMMSON with ASOs could represent an interesting therapeutic opportunity in complement to BRAF inhibitor as SAMMSON expression is specific to melanoma cells compared to OXPHOS inhibitors which are unspecific. SAMMSON was also shown to be expressed in mesenchymal melanoma cells, more dependent on glycolysis but still sensitive to SAMMSON depletion.

A subsequent study confirmed previous observations on SAMMSON and also gave new insights on its function [359]. In addition to p32 and XRN2, the authors found that SAMMSON also interacts with CARF, a p53 interacting protein regulating DNA damage response in the nucleus [360]. In cells, in absence of SAMMSON, CARF binds to XRN2 and limits its presence in the nucleolus where it is involved in rRNA maturation. SAMMSON disrupts this complex allowing XRN2 into the nucleolus and scaffolds CARF interaction with p32 that is involved in mitochondrial rRNA processing. In this way SAMMSON coordinates mitochondrial and cytoplasmic translation optimizing mitochondrial homeostasis and preventing mitochondrial protein accumulation in the cytoplasm and subsequent mPOS-induced cell death (**Figure 29**).

The benefit for melanoma cells of SAMMSON expression could be to give them a growth advantage through an increased and coordinated translation potential, a process often coupled with cell proliferation and metabolism as this process necessitates large amounts of ATP [361].

2. LENOX

More recently, a second melanoma-specific lncRNA called Linc00518 or LENOX (LincRNA-ENhancer of Oxidative phosphorylation) was described. LENOX expression is highly elevated in SKCM and UVM compared with normal skin and is associated with an overall poor prognosis in patients. It is expressed in all types of melanoma cells and is regulated by MITF and SOX10 but also by TFAP2A whose locus is adjacent to that of LENOX. TFAP2A is implicated in melanocyte differentiation and could act as a potential tumour suppressor even if it's role in melanoma is still disputed [362]. LENOX has even been proposed as a biomarker for melanoma as it can be detected by using non-invasive patches and in biopsies [363]. Just like for SAMMSON, ASO-mediated depletion of

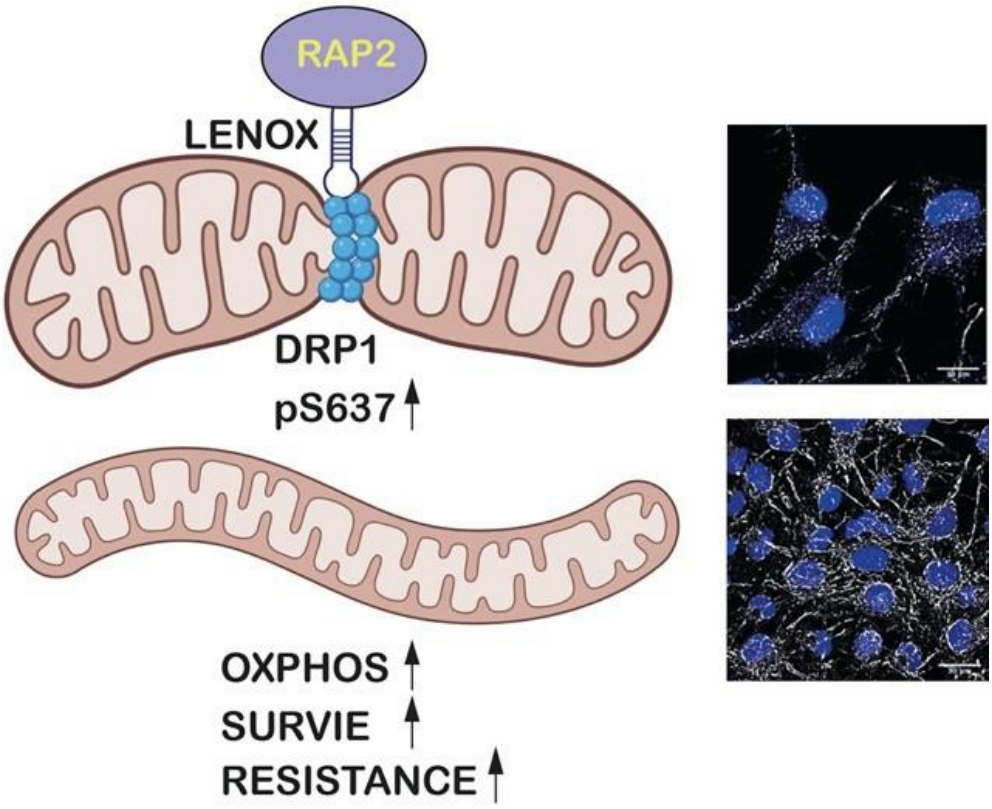


Figure 30. LENOX mechanism of action.

From the INSB.CNRS website. The lncRNA LENOX interacts with the small GTPase RAP2C scaffolding its interaction with DRP1. This interaction promotes DRP1 phosphorylation on the serine 637 which inhibit DRP1 function, preventing mitochondrial fission. The resulting increased mitochondrial fusion leads to enhanced OXPHOS activity which promotes melanoma cell survival and resistance to MAPKi.

LENOX reduces melanoma cell proliferation *in vivo* and *in vitro* and triggers apoptosis. LENOX proteome was also determined by RAP-MS identifying a specific interaction with a group of small GTP-binding proteins designed RAP2 A, B and C whose roles are poorly described in the literature [364,365]. Interestingly, RAP2 localizes to mitochondria in melanoma cells suggesting it may be involved in mitochondrial processes.

Strikingly, LENOX or RAP2C depletion reduces the reserve respiratory of melanoma cells and modulates mitochondrial homeostasis with accumulation of shorter mitochondria which is associated in the literature with impaired respiratory functions [366]. Mechanistically, LENOX seems to scaffold an interaction between RAP2C and the large GTPase DRP1, the major effector of mitochondrial fission [367]. LENOX depletion reduced DRP1 S637 phosphorylation, an event that negatively regulates fission activity of DRP1 and leads to accumulation of longer fused mitochondria and hence enhanced respiratory capacity (**Figure 30**) [368]. Moreover, like SAMMSON, LENOX depletion combined with MAPK inhibitor treatment leads to a cooperative reduction of melanoma cell survival.

Conversely, even if both SAMMSON and LENOX seem to promote mitochondrial homeostasis, they display opposing regulation upon vemurafenib treatment. SAMMSON levels go down in the short term before being restored after three weeks, whereas LENOX is upregulated upon treatment before going down in the longer term. Given its capacity to promote OXPHOS, LENOX is upregulated to facilitate the glycolysis to OXPHOS switch induced by BARF inhibition. In contrast, as BRAF inhibition blocks proliferation, SAMMSON is repressed due to the lower demand for protein translation, a process known to be globally repressed in stressed cells [301,369]. Interestingly, it is possible to deplete both SAMMSON and LENOX together by using sub-optimal ASO concentrations, which produced a strong induction of apoptosis and proliferation arrest compared to single depletions. Thus, LENOX and SAMMSON represent two potential therapeutic targets in melanoma, targetable together in addition to MAPKi inhibitors to achieve an efficient depletion of all types of melanoma cells. Of note, other melanoma-associated lncRNAs have been described in the literature, showing that a complex network of lncRNAs modulate cell functions and promote melanoma progression [59,370,371].

G. Conclusion of Section 2

In this section, I discussed the main features of melanoma; the events involved in initiation and progression of the disease, highlighting both the genetic and epigenetic mechanism that drive these processes; the characterization of melanoma cell states and their relationship to the activity of specific transcription factors, the proliferative, invasive and drug resistance properties of the different cell states; the treatment options and how heterogeneity influences resistance and relapse; the metabolic switch that accompanies phenotype switch and finally the role of lncRNAs SAMMSON and LENOX in the regulation of translation and mitochondrial homeostasis in melanoma and their potential as therapeutic targets.

The characterization of LENOX and SAMMSON and the possibility to combine their targeting led us to search for other melanoma-specific lncRNA, potentially impacting other pathways of melanoma metabolism.

In this context, the major focus of this thesis is characterization of Linc00520 or LENT (LncRNA ENhancer of Translation), a novel lincRNA highly overexpressed in SKCM and essential for melanoma cell proliferation and survival. However, I will first provide some background to introduce a helicase called DHX36, known to resolve DNA and RNA structures known as G-quadruplexes (G4s). I describe the roles of G4s and DHX36 below.

Section 3: Context for the Results

A. G-quadruplexes

G4s are DNA or RNA structures composed of stacks of guanine tetrads, stabilized by ions and held together by hydrogen bonds (**Figure 31**) [372]. Notably, they are stabilized by K+ but destabilized by Li + ions. Several types of G4 exist based on their conformation as they can be either parallel or antiparallel. They are widespread both in the genomic DNA [373] and transcribed RNA [374] of eukaryotes cells. Strikingly, they seem globally absent from bacteria, suggesting that prokaryotes either lost their need during evolution or that they developed other structures or mechanisms [375]. Conversely, a study described that mitochondria DNA contains three times more G4s than nuclear DNA., challenging this view [376]. It's also worth considering that G4s can be intramolecular or intermolecular as small RNAs can interact together to form this type of structure [377]. Moreover, several ligands and proteins are known to bind G4s, providing tools to assess their localization and their functions [378].

For example, G4access has been recently developed to study the role of G4 in the genome [379]. G4access-profiling showed that G4 formation was associated with open chromatin regions, nucleosome positioning and Pol II transcription, a property also determined by G4 ChIP-seq [380]. Even if the precise mechanism remains elusive, it seems that G4 formation could promote open chromatin at promoter and thus, gene transcription as small ligands stabilizing G4 enhance Pol II recruitment. Similar effects on transcription were also found in mitochondria [381]. Additionally, G4s were also often found at telomere extremities, rich in guanines and partially single-stranded [382]. A proposed function of telomeric G4s is that they represent a cap, stabilizing and protecting the telomere extremities, but experimental proof is still needed to validate this model. DNA G4s were even found in the cytoplasm in response to oxidative stress [383]. They seem implicated in stress granule formation, a process well described in response to stresses as it enables cells to modulate the translation of stress-related transcripts [384]. A proposed mechanism is the synthesis of single-stranded DNA upon DNA damage which could fold into G4s and be exported to the cytoplasm to form these granules.

Interestingly, RNA G4s are also stabilized in the cytoplasm upon different stresses, as a way to stabilize RNA molecules [385]. Moreover, they are also able to trigger stress granule formation [386], like

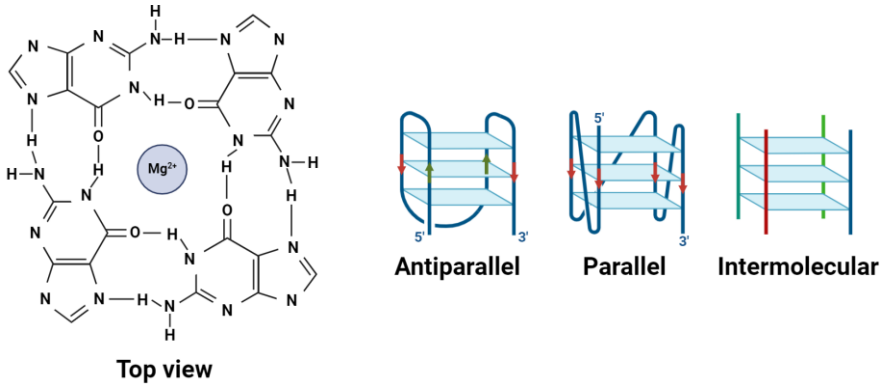


Figure 31. G-quadruplex (G4) structures.

G4s are composed of stacks of guanines stabilized by ions. One layer is made of a guanine tetrad, each guanine forming hydrogen bonds with two other guanines. Intramolecular G4s can be either antiparallel or parallel based on their organization. In parallel G4s, all strands composing the tetrad go in the same direction, whereas if two strands point in opposite directions it is an antiparallel G4. Intermolecular G4s can also be formed where the guanines are contributed from different molecules.

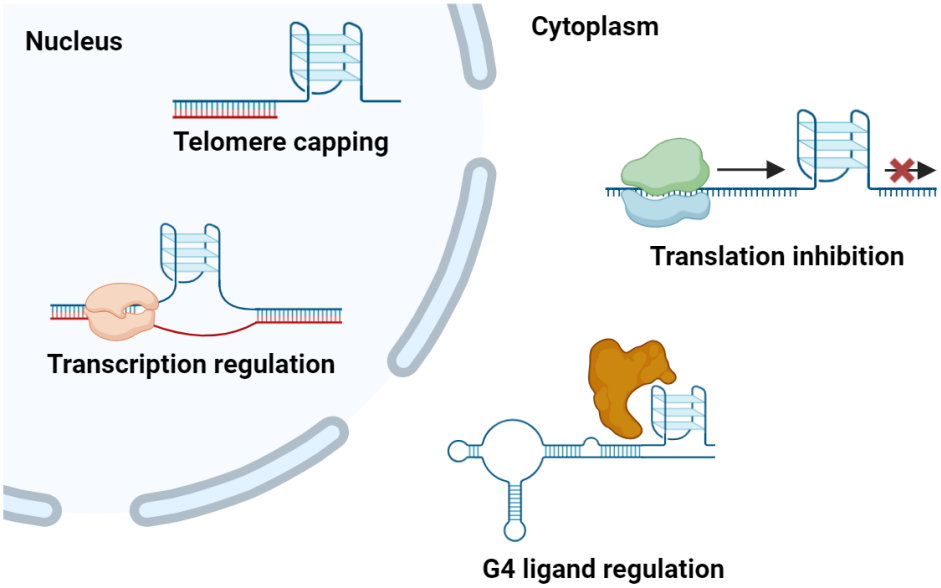


Figure 32. G4 functions. G4s are found across DNA and RNA in eukaryotic cells.

In the nucleus they were shown to be present at the extremities of telomers potentially preventing their degradation. They can also regulate transcription as they were shown to be frequent at proximal promoters and associated with open chromatin. In the cytoplasm, they were mainly found to be involved in translation inhibition where their presence in the 5'-UTR prevents ribosome scanning on the mRNA. They can also be present on ncRNA, representing interaction interfaces for G4 ligands, regulating key biological functions.

their DNA counterparts. RNA G4s are also enriched in UTRs of mRNA relative to the CDS [387], suggesting that they may influence their translation. Indeed G4s were found to inhibit translation as they may be too stable for the ribosome to process them [388]. Thus, G4s represent a widespread structural feature of both DNA and RNA with suggested implications in all kinds of molecular processes, even if their mechanisms of action are still not well understood (**Figure 32**).

Moreover, G4s have been linked with human pathologies as their presence was reported in several transcripts essential for disease onset such as the oncogene *NRAS* or the tumor suppressor *TP53* [389]. In melanoma, studies found that stabilizing G4s with ligands such as RHPS4 or IZTZ-1 inhibited cell proliferation [390,391]. Notably, the authors found that those two ligands target the oncogene *MYC*, lowering its expression. These observations led researchers to propose that targeting G4s could be a therapeutic strategy to target cancer cells. This prompted the development of a G4 ligand named Dicentrine which seem to bind preferentially oncogenic and telomeric G4s, suggesting that differences exist between normal and oncogenic G4 structures [392]. Dicentrine treatment leads to cell cycle arrest followed by apoptosis of cancer cells.

In conclusion, G4s are gaining attention as they could represent novel anticancer targets as they are present and essential for oncogenic transcript expression and function. Nevertheless, efforts are necessary to characterize the diversity of G4 structures and to develop ligands able to efficiently discriminate the different form of G4s. Moreover, tools are still needed to conduct *in vivo* studies of G4 folding as they were found to be globally unfolded in eukaryotic cells [375].

B. DHX36

Supporting the idea that G4s regulate specific processes is the existence of multiple G4 resolvases. Among these, the helicase DHX36 is considered as one of the main G4 resolvases, in particular for RNA molecules [393]. DHX36 is a member of the DEAH/RHA family of helicases and is also known as RHAU or G4R1 in the literature. This helicase was first described as the major source of G4 resolving activity in HeLa cell lysate in presence of ATP [394]. Subsequent studies showed that DHX36 is able to directly bind with high affinity to DNA or RNA through its N-terminal region [395]. The structure of bovine DHX36 associated with the G4 formed by the *MYC* DNA sequence was resolved (**Figure 33**) [396]. The helicase binds DNA through several interfaces. A helix called the DSM

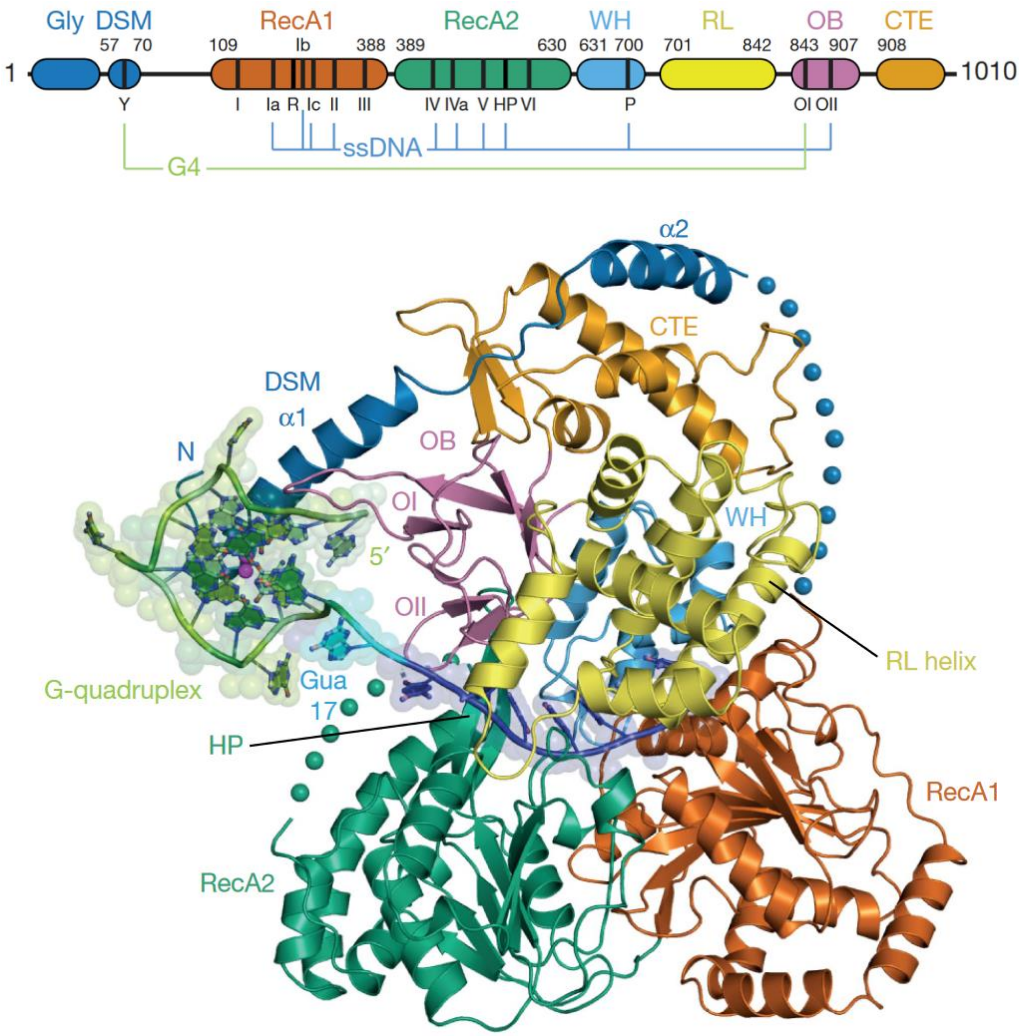


Figure 33. DHX36 organization and structure.

From Chen et al., 2018. Co-crystal structure of the bovine DHX36 with Myc DNA forming a G4. The DHX36 helicase contains several domains forming interfaces that interact with nucleic acids. The DSM and OB domains interact with the G4 structure whereas RecA1, RecA2, WH and another region of OB form a tunnel binding single stranded nucleic acids.

projects away from the core of the protein and stacks on top of the G4, destabilizing the structure with the help of an OB fold. The net result is to pull a single guanine out of the G4 thereby destabilizing its structure. This process is reversible and can occur multiple times but stops if the unwound guanine forms a base pair with the cytosine on the complementary strand, thus leading to unwinding of the G4.

Interestingly, while DHX36 can destabilize G4 in DNA in an ATP-independent manner, DHX36 acts on G4-RNA substrates by a distinct ATP-dependent mechanism. DHX36 first induces a stably unfolded state as seen for DNA substrates independently of ATP hydrolysis followed by successive cycles of ATP-dependent and stepwise refolding of G4-RNA. The ATP-independent unfolding and ATP-dependent refolding of G4-RNA can occur multiple times before DHX36 protein dissociates from the G4-RNA substrate. DHX36 thus maintains the G4 in a dynamic state until the RNA binds to molecules such as the ribosome that gain access to the RNA from the 5' end.

DHX36 localizes mainly in the cytosol in HEK293 and HeLa cells and has a similar binding profile to EIF4A implicated in translation initiation, suggesting a similar role for DHX36 [397]. Moreover, 70 % of DHX36 binding sites were mapped to exons of almost 4500 mRNAs. Interestingly, loss of DHX36 only had a mild effect on cell proliferation and increased the levels of its RNA targets by 20 %, in line with the fact idea that G4 presence could stabilize nucleic acid molecules perhaps by limiting their translation. Indeed, DHX36 loss was also associated with a marginal, but significant reduction in translation efficiency of its target mRNAs. Phenotypically, DHX36 loss correlated with an increase of stress granule formation and an elevation of the stress response marker PKR/EIF2AK2 [398]. These observations are in line with the previous proposition that the presence of a G4 on a mRNA impairs its translation. DHX36-mediated translational control was further demonstrated in muscle stem-cells, where an enrichment of DHX36 in the 5'UTR of mRNAs was observed [399]. Nevertheless, effects of DHX36 loss are limited, suggesting redundancy with other helicases. Multiple helicases have been characterized as G4 resolvases such as the RECQL-family including BLM [400] and WRN [401], as well as EIF4A, DDX5, DDX21, DHX9 or DDX3X [378]. In particular DDX3X, DDX5 and DHX9 are reported RNA G4 resolvases, with DHX9 shown to control mRNA translation efficiency [402]. DHX36 was also found present in the nucleus and seems able to resolve G4s present in the promoter of several genes, enabling their transcription [403,404]. DHX36 is also implicated in telomere maintenance as it binds the telomerase RNA to unwind its G4 [405].

Results

Interaction of lncRNA LENT with DHX36 regulates translation and suppresses autophagy in melanoma.

Alexandre Haller[1, 2, 3, 4], Giovanni Gambi[1, 2, 3, 4, 5], Mattia D'Agostino[6], Guillaume Davidson[1, 2, 3, 4], Antonin Lallement[1, 2, 3, 4], Gabrielle Mengus[1, 2, 3, 4], Guillaume Bec[6], Angelita Simonetti[6], Eric Ennifar[6], and Irwin Davidson[1, 2, 3, 4, 7, 8]

1. Institut de Génétique et de Biologie Moléculaire et Cellulaire, BP 163, 67404 Illkirch, France

2. Centre National de la Recherche Scientifique, UMR7104, 67404 Illkirch, France

3. Institut National de la Santé et de la Recherche Médicale, U1258, 67404 Illkirch, France

4. Université de Strasbourg, 67404 Illkirch, France

5. NYU Grossman School of Medicine, Langone Medical Center, 550 First Avenue, Smilow 1307

New York, NY 10016

6. Institut de Biologie Moléculaire et Cellulaire du CNRS, 15 Rue René Descartes F-67084 Strasbourg

7. To whom correspondence should be addressed: irwin@igbmc.fr

8. Equipe Labélisée Ligue contre le Cancer.

Running Title: LENT suppresses mitophagy

Key words: LncRNA, melanoma, mitophagy, antisense oligonucleotide,

Conflict of interest.

GG, AH and ID are co-inventors on patent: Combination therapy for Melanoma. PCT/EP2023/067998.

Abstract.

The melanocyte lineage determining Microphthalmia-associated transcription factor (MITF) drives proliferation and survival of melanocytic melanoma cells through regulation of both coding genes and long non-coding RNAs (LncRNAs). Here we characterize LINC00520 (hereafter called LncRNA ENhancer of Translation, LENT) regulated by MITF and strongly expressed in melanocytic melanoma cells. LENT is essential for proliferation and survival of cultured melanocytic melanoma cells and xenograft tumours. LENT interacts with the G4 quadruplex resolvase DHX36 and both associate with the ribosome in the 80S and light polysome fractions. LENT modulates DHX36 association with a collection of mRNAs regulating their engagement with polysomes and fine-tuning their subsequent translation. These mRNAs encode proteins involved in endoplasmic reticulum (ER) and mitochondrial homeostasis as well as autophagy. Consequently, LENT silencing leads to extensive autophagy and mitophagy, compromised oxidative metabolic capacity and increased translation and mis-localization of mitochondrial proteins leading to proteotoxic stress and apoptosis. The LENT-DHX36 axis therefore fine-tunes translation of proteins involved in ER and mitochondrial homeostasis suppressing autophagy and promoting survival and proliferation of melanoma cells.

Introduction

Melanoma tumours are notoriously heterogeneous, with melanoma cells adopting multiple cell states with differing proliferative, invasive and stem cell capacities (1–7). Intra-tumour heterogeneity is a major determinant of therapeutic resistance with mesenchymal-type cells playing a critical role in targeted and immune checkpoint resistance (4, 7). The transcription programs associated with the different cell states are driven by of a host of transcription factors, with the lineage-defining MITF (Microphthalmia-associated transcription factor) and SOX10 driving the more differentiated melanocytic cell state, while AP1, TEAD, PRRX1 and TCF4 drive the undifferentiated mesenchymal state (6–11). Other intermediate states have been defined such as the neural crest stem cell (NCSC)-like state that plays a key role in minimal residual disease and the emergence of drug resistant populations (3, 12).

In melanocytic melanoma cells, MITF and SOX10 bind together at cis-regulatory elements to promote expression of genes driving proliferation, survival and oxidative metabolism (8, 9). While, these factors regulate multiple coding genes they also regulate expression of long non-coding (lnc)RNAs such as the melanoma-specific lncRNA SAMMSON (LINC01212) essential for melanoma cell proliferation and survival through coordinating mitochondrial and cytoplasmic translation (13, 14). SAMMSON silencing induced mitochondrial precursor overaccumulation stress (mPOS), a form of proteotoxic stress, resulting in melanoma cell death. SOX10 also regulates the melanoma-specific lncRNA LENOX (LINC00518) that interacts with the small GTPase RAP2C promoting its interaction with DRP1 and impairing mitochondrial fission through enhanced DRP1 S637 phosphorylation (15). LENOX potentiates oxidative phosphorylation metabolism to promote melanoma cell survival and resistance to MAP kinase inhibitors.

Here we characterize LINC00520 (hereafter called LncRNA ENhancer of Translation, LENT) strongly expressed in melanocytic melanoma cells under the regulation of MITF and

essential for proliferation and survival *in vitro* and *in vivo*. LENT interacts with the G4 quadruplex resolvase DHX36 and both associate with the 80S and light polysome fractions. LENT modulates DHX36 association with a collection of mRNAs regulating their engagement in polysomes and their subsequent translation. LENT coordinately regulates engagement with the light polysomes of mRNAs encoding proteins enriched in endoplasmic reticulum (ER) homeostasis as well as autophagy and mitophagy. Consequently, LENT silencing leads to extensive autophagy/mitophagy, compromised OxPhos capacity and increased translation and mis-localization of mitochondrial proteins leading to proteotoxic stress and apoptosis. Our results are consistent with a model where LENT fine-tunes translation of proteins involved in ER and mitochondrial homeostasis by modulating the ability of ribosome-associated DHX36 to unwind G4 structures in their mRNAs and their engagement with polysomes. LENT, LENOX and SAMMSON hence constitute a set of melanoma-expressed lncRNAs that act coordinately to fine-tune translation and/or mitochondrial activity and promote melanoma cell proliferation and survival.

Results.

LENT is expressed in melanocytic melanoma cells and associated with poor patient outcome.

Integration of MITF and SOX10 ChIP-seq data with RNA-seq following MITF silencing in 501Mel melanocytic melanoma cells (9) identified LENT (LINC00520) as a lncRNA directly and positively regulated by MITF. LENT expression was reduced upon silencing of MITF or its cofactor BRG1 and the corresponding locus displayed several MITF bound sites associated with BRG1 and marked by H3K27ac in 501Mel cells (**Fig. S1A**). LENT expression was low in normal tissues (0.854 Log2 normalized counts) with highest expression in the oesophagus mucosa and stomach in the GTEX database (**Fig. S1B** and data not shown).

Expression was highest in cutaneous melanoma (SKCM, 6.189 Log2 normalized counts), compared to other cancer types (1.490) including uveal melanoma (UVM, 1.732) (**Fig. S1B**). Expression was also upregulated in primary melanoma compared to benign nevi and normal tissues, (**Fig. S1C**). Thus, LENT expression was negligible in normal tissues and upregulated more than 60-fold in cutaneous melanoma.

Analyses of scRNA-seq data from human melanoma xenografts (3) showed that LENT was widely expressed except in NCSC and mesenchymal cells (**Fig. S1D**), while in scRNA-seq data from human melanoma patients (7), it was also broadly expressed except in mesenchymal cells and was strongest in the hypoxia-stress cell cluster (**Fig. S1E**). Preferential LENT expression in melanocytic type cells was confirmed by RT-qPCR analyses of a collection of melanoma cell lines (**Fig. S1F**). Cytoplasmic LENT localization in MITF expressing melanoma cells could be directly observed using RNA-scope on human melanoma patient sections, whereas its expression was low in normal melanocytes (**Fig. S1G**). RNA-scope also showed a predominantly cytoplasmic localization in cultured melanoma cells, whereas no signal was seen in Hela cells (**Fig. S1H**). Further analyses showed the 432 nt isoform 5 as the most abundant in 501Mel cells and also in melanoma patients (**Fig. S1I**). LENT is therefore a cytoplasmic melanoma-enriched lncRNA most abundant in melanocytic MITF-expressing cells.

To determine whether LENT expression correlated with patient outcome, we divided the TCGA SKCM dataset into primary and metastatic samples, performed unsupervised clustering of the transcriptome data from each and GSEA analyses of differentially expressed genes to define the signatures of each cluster. In primary melanoma, LENT was co-expressed with MITF and SOX10 in cells defined by an oxidative phosphorylation (OxPhos) and cell cycle signature typical of melanocytic MITF-expressing cells, but was strongly reduced in mesenchymal (designated as EMT, epithelial to mesenchymal transition) cells expressing

markers such as PRRX1 (**Fig. S2A)**. In contrast, as previously described (15) LENOX displayed a broader expression pattern being expressed also in EMT cells.

In metastatic melanoma, LENT again was strongest expressed in the MITF-SOX10 expressing OxPhos/cell cycle cells, but reduced in cells with EMT signatures (**Fig. S2B**). In primary melanoma, high LENT expression was associated with better survival whereas in metastatic samples high LENT expression was associated with poorer outcome (**Fig. S2C-D**). These observations are in line with the idea that low LENT-expressing mesenchymal cells promote metastases of primary melanoma (6), whereas LENT-expressing cells with OxPhos and cell cycle signatures associate with poorer survival in metastatic samples, hence accounting for the differential association of LENT expression with survival.

LENT cooperates with LENOX and SAMMSON to promote melanoma cell proliferation and survival.

To address the function of LENT in melanoma cells, we silenced its expression by CRISPR interference (CRISPRi) using dCAS9-KAP1, transfection of locked nucleic acid GapmeR antisense oligonucleotides (ASO) or by Doxycycline (Dox)-inducible expression of LENT-targeting shRNA. CRISPRi silencing in 501Mel cells with LENT promoter-targeting sgRNAs that potently reduced its expression resulted in strongly reduced colony forming capacity (**Fig. 1A-B**). Transfection of melanoma cells with different phenotypes and driver mutations, with 2 independent ASOs that reduced LENT expression by over 80% compared to a non-targeting control (CTR) (**Fig. 1C)** led to reduced growth of melanocytic, but not mesenchymal melanoma cells nor HEK293T cells that did not express LENT (**Fig. S3A**). ASO-mediated silencing resulted in strongly reduced cell proliferation (**Fig. 1D**) and cleaved caspase 3-expressing apoptotic cells (**Fig. 1E**) with early and late apoptotic cells observed in flow cytometry (**Fig. S3B**). We also silenced LENT with a stably integrated Dox-inducible shRNA

that efficiently reduced 501Mel proliferation (**Fig. S3C**). In contrast, ectopic Dox-induced expression of LENT isoform 5 stimulated colony formation in melanoma cells, but also in HEK293T cells where it was not normally expressed **(Fig. S3D-E**)

To test if LENT silencing could also block xenograft tumour growth, melanocytic IGR37 cells were injected subcutaneously in immunodeficient mice and when tumours reached ≈100 mm^3, mice were subsequently injected subcutaneously every 2 days with LENT ASO. Compared to untreated controls, injection of LENT ASO reduced LENT expression in tumours and strongly reduced tumour growth and tumour weight (**Fig. 1F-G, and Fig. S3F**).

All 3 targeting strategies as well as gain of function therefore revealed the essential role of LENT in the proliferation and survival of melanocytic melanoma cells in culture and *in vivo* xenografts.

We previously showed that ASO silencing of LENOX and SAMMSON cooperated to induce melanoma cell death (15). To assess if LENT also collaborated with LENOX and SAMMSON, melanocytic 501Mel and MM011 cells were transfected with sub-optimal concentrations of ASO targeting LENT alone or together with LENOX or SAMMSON. Compared to LENT, LENOX or SAMMSON alone, a cooperative increase in apoptosis of both lines was observed using the combinations of ASO, and an additive increase in slow proliferation (**Fig. 1H-I)**. LENT silencing also cooperated with MAP Kinase inhibition by the dabrafenib and trametinib combination to eradicate melanoma cells (**Fig. S3G**). MITF and SOX10 therefore coordinately regulate a network of 3 lncRNAs that cooperate to promote melanoma cell survival.

LENT interacts with the G4 resolvase DHX36.

Consistent with the observation that LENT is predominantly cytoplasmic and so less likely to influence transcriptional regulation, RNA-seq from ASO-control or LENT ASO-

silenced cells revealed only minor changes in gene expression with only 97 up-regulated and 82 downregulated transcripts (Log2 fold-change +/-1 p<0,05). LENT silencing did not have a major impact on gene expression (**Fig. S4** and **Dataset S1**) and may therefore act via other cellular processes.

To identify LENT interacting proteins, we performed pulldown from cytoplasmic extracts of 501Mel cells using a tiling array of biotinylated oligonucleotides complementary to LENT or as negative control, the prostate cancer lincRNA PCA3, followed by mass-spectrometry. Compared to several control lncRNAs, LENT was selectively enriched using its cognate oligonucleotides, but not those of the PCA3 control (**Fig. 2A**). Triplicate purifications were performed and LENT-interacting proteins identified by mass-spectrometry. DHX36 was the most enriched protein in the LENT pulldown with no peptides found in the 3 control samples, but an average of 17 in the LENT pulldowns (**Fig. 2B and Dataset S2**). To confirm this interaction, we performed LENT pulldown from native or UV-crosslinked extracts followed by immunoblot. Under both conditions, DHX36 was enriched in the LENT pulldown compared to the PCA3 control, whereas neither the SAMMSON-interacting CARF (14) nor LENOX-interacting RAP2 (15) were enriched (**Fig. 2C**). For further confirmation, we performed LENT pulldown from the HEK293T cells ectopically expressing LENT isoform 5. DHX36 was detected after pulldown from LENT-expressing HEK293T cells, but not from control cells with empty GFP vector (**Fig. 2D**). In the converse experiment, we immunoprecipitated DHX36 from 501Mel cells and found enrichment of LENT compared to the control IgG and compared to SAMMSON and MALAT1 (**Fig. 2E-F**).

To ask if LENT interacts directly with DHX36, we generated and purified recombinant DHX36 in *E.Coli* (**Fig. S5A-B**) and performed electrophoretic mobility shift assay (EMSA) with *in vitro* transcribed LENT isoform 5 RNA (**Fig. 2G**). The presence of increasing amounts of purified DHX36 shifted LENT into slower migrating DHX36-RNA complexes. As DHX36

binds RNA with G4 structures (19), we used the QGRS program (20) that predicted a potential G4 structure in LENT, but with a rather low score. Mutation of three guanines in this sequence decreased complex formation in EMSA, but did not fully abolish the interaction (**Fig. 2G**). Together, these *in cellulo* and *in vitro* experiments revealed a selective and direct interaction of LENT with DHX36 that is partially dependent on a potential G4 forming sequence in LENT.

LENT modulates association of mRNAs with DHX36.

DHX36 unwinds G4 structures in RNA and in particular in the 5'-UTR of mRNAs to facilitate their translation (19, 21–24). This observation suggested that LENT may modify DHX36 interactions with mRNAs and their translation in melanoma cells. We therefore investigated the mRNAs associated with DHX36 and determined if their association was modulated by LENT silencing. We performed triplicate DHX36 or control IgG immunoprecipitations (IP) from 501Mel cells expressing control shRNA and the associated mRNAs were sequenced. Almost 2000 mRNAs were preferentially found in the DHX36 IP compared to IgG, whereas 1949 mRNAs were less present in the DHX36 IP compared to control (Log2 fold-change +/-1 p<0,05) (**Fig. 3A** and **Dataset S3**). One of the most enriched was the DHX36 mRNA suggesting DHX36 acts to regulate its own translation in a positive regulatory loop. Ontology analysis showed that DHX36-associated mRNAs were enriched in those encoding proteins involved in mitochondrial function with protein targeting to mitochondrion, mitochondrial calcium ion homeostasis amongst the most enriched terms (**Fig. 3B**). Comparison with RNA-seq data from 501Mel cells showed no correlation between association with DHX36 and expression levels excluding the possibility that we spuriously enriched highly expressed mRNAs in the DXH36 IP (**Fig. S6A**).

It has been reported that mRNAs that associate with DHX36 are enriched in a GG-rich motif with a propensity to form G4 structures [(25) and **Fig S6B**]. This motif was predicted to

be present in around 20% of the mRNAs enriched in the control IP, but close to 40% in the DHX36 IP (**Fig. S6C**) and increased to 50-60 % when considering the mRNAs most enriched in the DHX36 IP (**Fig. S6C** and **Dataset S3**). Examination of the DHX36 mRNA sequence with QGRS mapper indeed identified several potential G4-forming sequences including the 5'-GGnGGnGG-3' motif (**Fig. S6D**) consistent with the observation that it was one of the most enriched mRNAs. We compared mRNAs enriched in the DHX36 IP in 501Mel cells with previously published RNA-seq data from HeLa or HEK293T cells designed to identify G4-containing RNAs (26, 27). Comparing the overlap between the 2 HeLa datasets showed 1832 common mRNAs representing between 46% and 64% of the identified G4-containing mRNAs using different techniques. Comparison with the 501Mel mRNAs enriched by DHX36 IP showed that 900 (46%) were shared with the HeLa and HEK293T datasets, with 341 common to all (**Fig. S6E**). These common mRNAs were enriched in terms associated with transcription and MAP Kinase signaling (**Fig. S6F**).

We also compared the top 500 RNAs in our dataset with the 500 most enriched in a previously published DHX36 PAR-Clip (28) dataset. Among them, only 41 were common between the two datasets, including the mRNA encoding DHX36 (**Fig. S6G**). We observed a clear enrichment for mitochondria-related terms in the 501Mel DHX36 IP (**Fig. S6H**) compared to Sauer et al, characterized by transcription and TGF-B signaling related terms (**Fig. S6I**).

We then investigated if LENT silencing modified mRNA interaction with DHX36. We directly compared RNA-seq of the DHX36 IPs from the control shRNA compared to shLENT cells and identified 484 genes displaying increased association with DHX36 in absence of LENT and 429 with less association ($p<0,05$) (**Fig. 3C**, and **Dataset S3**). As expected, due to its downregulation by shRNA silencing, LENT was identified as less associated with DHX36. Ontology analyses of mRNAs showing increased DHX36 association revealed enrichment in several process including cell cycle, mitophagy and autophagy, whereas those less associated

were enriched in lysosome, metabolic process and allograft rejection (**Fig. 3D)**. Analyses of the mRNAs whose association with DHX36 was affected with the QUADRatlas software showed that a large majority comprised experimentally described and/or predicted G4s (**Fig. S6J**). Together these data define DHX36-associated RNAs in melanoma cells and identify RNAs whose association with DHX36 was positively or negatively modulated by LENT.

LENT and DHX36 are associated with the ribosome and regulate coordinate engagement of mRNAs encoding proteins involved in ER homeostasis with polysomes.

While preparing the DHX36-associated RNAs for sequencing, we noted that the 28S and 18S rRNAs were strongly enriched in the DHX36 IP, but not the control IP (**Fig. 4A**) and therefore used ribo-depletion kits to prepare the libraries for RNA-seq. This observation however strongly suggested that DHX36 was associated with the ribosome. To assess this, we performed polysome profiling of 501Mel cell extracts and analyzed both RNA and protein contents of the fractions. Based on the RNA absorption profile (**Fig. 4B**) and the distribution profiles of EIF4A2 (initiation factor marking the 40S) and RPL36 (component of the large subunit), we designated the 40S, 60S, 80S and polysome fractions. DHX36 showed association with the 60S, 80S and was additionally present in the polysome fractions (**Fig. 4C**). As expected, the control GAPDH mRNA was enriched in the heavier polysome fractions, whereas LENT showed a strong peak in the 80S fraction, but rapidly decreased in the light polysome fractions (**Fig. 4D)**. These observations suggested that LENT may associate with DHX36 on the 80S and light polysome fractions.

To investigate if the LENT-DHX36 axis regulated mRNA association with ribosomes, we prepared biological triplicate polysome fractions from shControl or shLENT-silenced 501Mel cells. We pooled RNA from four fractions representing the 80S, light or heavy polysome components from each replicate and assessed their composition by RNA-seq (**Fig.**

4E and **Dataset S4**). Few RNAs showed differential presence in the heavy polysome (HP) fractions from the control or LENT silenced cells, whereas 189 and 246 mRNAs were depleted or enriched, respectively in the 80S fraction (Log2 fold-change +/-1 p<0,05) (**Fig S7A-B** and **Dataset S4**). However, the most striking effect was seen in the light polysome (LP) fractions, where 383 mRNAs were depleted in the LENT silenced cells, while only 21 were enriched (**Fig. 4F**).

We interrogated the polysome RNA-seq to determine if the mRNAs whose association with DHX36 was positively or negative regulated by LENT were also differentially engaged with the polysome fractions. Depletion of LENT was clearly seen in all fractions (**Fig. 5A**). Of the 383 RNAs depleted in the LP fractions, 74 were also depleted in the DHX36 RIP upon LENT silencing (**Fig. 5B** and **Dataset S4**) indicating a significant, but incomplete overlap between the 2 experimental approaches. In contrast, almost no RNAs showed discordant regulation, with only a single transcript up in DHX36 RIP and down in the LP fractions. For example, the mRNAs encoding UBE4A and CTSD whose interaction with DHX36 was reduced upon LENT silencing also showed reduced association with 80S, LP and HP fractions, with the most striking reduction seen in the LP fraction (**Fig. 5C**). In contrast, mRNAs encoding NOX4 and RBPJ whose association with DHX36 was increased upon LENT silencing were also increased the 80S, LP and HP fractions and although the fold change was below cutoff, their increased association was statistically significant (**Fig. 5D)**.

An ontology analyses of the 383 mRNAs depleted in the LP fractions revealed their strong enrichment in several pathways pertaining to endoplasmic reticulum (ER) homeostasis such as ER stress, ER-associated protein degradation (ERAD), and protein glycosylation (**Fig. 4G** and **Dataset S5**). Key components of the ER stress/ERAD pathways such as the E3 ligase SYNV1, the HSPA5 chaperone, and the PDIA3, -4, and -6 enzymes were all significant depleted in the LP fractions, with PDIA encoding mRNAs also depleted in the 80S fraction

(**Fig. 5E** and **Fig. S7C**). The mRNA encoding WFS1 involved in ER Ca2+ transport was also depleted in the LP fractions, along with those encoding the NOMO1, -2 and -3 proteins and other components of the multi-pass translocon complex **(Fig. 5F and Dataset S5)**. Similarly, mRNAs encoding the DPAGT1 and GALNT2, 7 and 12 enzymes involved in N-linked or O-linked protein glycosylation, respectively, were all depleted in the LP fractions (**Dataset S4)**. KEGG ontology analyses gave comparable results, but further revealed enrichment in lysosome function (**Dataset S5**). These results indicated that engagement in LPs of mRNAs encoding key components of many processes associated with normal ER homeostasis and/or intracellular protein transport was coordinately regulated by LENT.

Related to the above, mRNAs encoding the MHC class 1 HLA-A; -B and -C antigens as well as the TAP1, TAP2 and CALR proteins involved in their transport and antigen presentation were depleted in the LP fractions (**Fig. S7D-E and Dataset S4-S5)**. Moreover, ontology analyses of the 74 mRNAs whose association with DHX36 was in addition reduced by LENT silencing also revealed their strong enrichment in ER homeostasis, including the above-mentioned HLA proteins, and lysosome function (**Fig. S7F and Dataset S4-S5**). Similarly, analyses of the RNAs depleted in the 80S fraction using relaxed criteria of Log2 fold-change ≥0,7, but with a more stringent adjusted p-value of <0,01 also showed a strong enrichment in many of the same terms related to ER homeostasis (**Dataset S5**).

The above data identified a set of mRNAs whose interactions with DHX36 and/or engagement with the LP fractions were modulated by LENT. QUADRatlas analyses of the 383 depleted mRNAs in the LP fraction indicated the presence of predicted (342/368) and experimental (217/368) G4 forming sequences (**Fig. S7G**). The enrichment of potential G4 forming sequences of in these RNAs was consistent with the idea that the LENT-DHX36 axis regulated their unwinding to facilitate their translation. To test this, we investigated if mRNAs whose association with DHX36 and/or engagement with the LP fractions was modified by

LENT silencing were differentially translated. UBE4A and CTSD whose mRNAs were less associated with DHX36 and the LP fractions upon shLENT accumulated to lower levels following ASO-mediated LENT silencing, whereas RNA-seq showed no change in overall abundance of the corresponding mRNAs (**Fig. 5G and I)**. The decreased protein level was therefore most likely due to altered translation and not mRNA down-regulation. In contrast, protein levels of NOX4, FUNDC1 and RBPJ, whose mRNAs showed increased association with DHX36 and the LP fractions, were increased upon ASO-mediated LENT silencing with again no overall changes in the corresponding mRNA levels (**Fig. 5G and I)**. Furthermore, increased NOX4 and reduced CTSD levels were seen in extracts from IGR37 xenograft tumours treated with LENT ASO (**Fig. 5H)**. Similarly, levels of HSPA5 and WFS1 proteins whose RNAs were less associated with the LP fractions were also decreased despite the fact that we did not detect changes in their association with DHX36 (**Fig. 5G)**. These data showed that LENT modulated interactions of mRNAs with DHX36 and/or the ribosome LP fractions to regulate their translation.

LENT and DHX36 are enriched at mitochondria.

Previous studies showed that DHX36 was predominantly cytoplasmic consistent with its ability to regulate mRNA translation (23, 28). Immunofluorescence revealed that DHX36 was enriched at cytoplasmic structures in 501Mel, IGR37 and A375 melanoma cells (A375; NCSC-type cells, not expressing LENT) that co-staining with HSP60 identified as mitochondria (**Fig. 6A-B**). Co-staining with HSP60 was less prominent in HeLa cells and while little nuclear staining was seen in 501Mel cells, stronger staining was seen in IGR37. ASO-mediated LENT silencing did not modify DHX36 mitochondrial localization, nor did it affect expression of DHX36 mRNA or protein (**Fig. S9A-C**).

DHX36 association with mitochondria in 501Mel cells was confirmed by immunoblots of cytoplasmic and mitochondrial fractions (**Fig. 6C**). Similarly, RT-qPCR showed that LENT was also abundant in the mitochondrial fraction (**Fig. 6D**). We then performed immunoblots of the mitochondrial fraction in presence of increasing quantities of digitonin that was previously used to assess association of proteins with mitochondria (29). While, the control Vinculin (VCL) and Beta-actin (ACTB) proteins were rapidly depleted with increasing digitonin concentration, the mitochondrial protein COX IV was resistant to the highest concentrations (**Fig. 6E**). Both DHX36 and the LENOX-interacting mitochondrial partner RAP2 were also resistant to digitonin showing they were strongly associated with mitochondria. Consistent with this, COX IV and DHX36 showed resistance to tryptic digestion in swelling buffer, whereas VCL and ACTB were sensitive (**Fig. 6F**). Hence in melanocytic melanoma cells, DHX36 was enriched and tightly associated with the mitochondria.

Together with the data from Fig. 4C, the immunofluorescence and biochemical data suggested that LENT associated with DHX36 on the mitochondrial proximal ribosomes. It has previously been shown that mitochondrial localized proteins are translated close to the mitochondria facilitating their import (30–32). Consistent with this idea, immunostaining showed that UBE4A was strongly enriched at mitochondria in 501Mel and IGR37 cells (**Fig. S9A**). Similarly, while a fraction of NOX4 was present in the nucleus, it was also enriched in mitochondria (**Fig. S9B)**. The transcriptional regulator, RBPJ was mainly nuclear, but a fraction of RBPJ could also be detected enriched at mitochondria (**Fig. S9C**). NOX4, RBPJ and UBE4A were further detected by immunoblot in biochemically purified mitochondria (**Fig. S9D)**. These observations supported the idea that the LENT-DHX36 axis regulated translation of mitochondrial localized proteins on mitochondrial-proximal ribosomes.

LENT silencing induces autophagy/mitophagy and proteotoxic stress.

The above observations showed that LENT modulated interaction of mRNAs involved in ER-homeostasis, lysosome and autophagy/mitophagy with DHX36 and/or the LP fractions and their subsequent translation. We therefore investigated if LENT silencing impacted these processes. Electron microscopy showed that LENT-silenced cells were characterized by lower numbers of mitochondria, but an accumulation of numerous autophagosomes not seen in the control shRNA cells (Auto in **Fig. 7A**). Many autophagosomes comprised mitochondria identifiable by their cristae indicating extensive mitophagy. In agreement with this observation, immunoblot with anti-LC3 antibody revealed accumulation of the LC3-II form indicative of autophagy in LENT silenced 501Mel and MM117 cells (**Fig. 7B**). A more modest but detectable LC3-II accumulation was also observed in extracts from LENT ASO-treated IG37 tumours (**Fig. 7C**). Staining of control and LENT silenced cells with both lysotracker and mitotracker showed increased numbers of lysosome-mitochondrial contacts in LENT silenced cells that was further indicative of auto/mitophagy (**Fig. 7D**). Accumulation of autophagosomes was not seen in LENOX silenced cells despite that fact that its silencing impacted mitochondrial homeostasis (**Fig. S10**) (15). Thus, induction of autophagy/mitophagy were major phenotypes of LENT silencing.

Given these observations, we asked if mitochondrial function was impacted by profiling the Oxygen Consumption Rate (OCR) using the Agilent SeaHorse. Compared to control ASO, LENT silencing reduced maximal OCR and reserve capacity, but not basal levels in 501Mel cells, whereas no effect was seen in HeLa cells (**Fig. 7E**). DHX36 silencing reduced maximal and reserve capacities in both cell types revealing its more general role in regulating mitochondrial activity (**Fig. 7E**). The mitophagy and impaired mitochondrial function upon LENT silencing led to increased ROS levels and the appearance of ROS-high apoptotic cells (**Fig. S11A**). LENT silencing was further associated with activation of the DNA damage response with increased gH2AX seen both by immunofluorescence (**Fig. S11B**) and

immunoblot (**Fig. S11C**). Immunoblot analyses showed close to maximal LC3-II accumulation already 16 hours after LENT silencing, whereas gH2AX appeared only after 24 hours suggesting it was a secondary effect of mitophagy (**Fig. S11D**).

The observed autophagy/mitophagy and impaired OxPhos prompted us to investigate changes in translation of mitochondrial proteins of the electron transport complexes. Strikingly, increased protein levels of ATP5A, UQCR2, SDHB, COXII and NDUFB8 were seen after ASO-mediated LENT silencing in 501Mel cells and in melanocytic Mel888 and MM117 cells (**Fig. 8A and Fig S12)**. In contrast, this accumulation was not seen in LENOX silenced cells despite the fact that its silencing was also associated with lowered OxPhos capacity (15) (**Fig. 8A)**. Both mitophagy and OxPhos protein accumulation were therefore specific to LENT-silenced cells. Increased ATP5A levels, the most strongly affected in cells, was also seen in extracts from LENT ASO-treated IGR37 xenograft tumours (**Fig. 8B**).

OxPhos protein accumulation may represent a compensatory response to the mitophagy and impaired mitochondrial function and may at least in part result from their increased translation as the presence of the corresponding mRNAs was up-regulated in the polysome fractions from shLENT silenced cells (**Fig. 8C and Fig. S12**). However, as the increase in association with the LP and HP fractions was modest, their accumulation may also result from their impaired ERAD-mediated degradation. Furthermore, their accumulation was surprising given the lowered OxPhos, suggesting the excess OxPhos proteins were not imported into the mitochondria, but accumulated in the cytoplasm. Immunofluorescence showed accumulation of mitochondrial proteins at the mitochondria, but also in the cytoplasm and the nucleus of the LENT-silenced cells (**Fig. 8D**). Immunoblots on the cytosolic and mitochondrial fractions showed increased presence of ATP5A in the cytosolic fraction, with little change in the mitochondrial fraction showing that the accumulated protein was not imported into the mitochondria but rather accumulated outside the mitochondrial (**Fig. 8E**).

Together these data are consistent with the idea that LENT silencing impaired ER-homeostasis resulting in auto/mitophagy with subsequent impaired mitochondrial function. Accumulation of the mitochondrial proteins around the mitochondria and in the nucleus may then induce mPOS (33, 34) or a more general proteotoxic stress leading to apoptosis (35).

Discussion.

LENT, a multi-functional lncRNA.

Here we characterize LENT as a cytoplasmic lncRNA that interacts with the G4 resolvase DHX36 to promote translation of mRNAs involved in ER homeostasis and mitochondrial function and supressing autophagy in melanoma cells. Mining of public data bases showed that LENT expression was much higher in cutaneous melanoma than in other cancers and in normal tissues. Unsupervised clustering of the SKCM TCGA gene expression data as well as scRNA-seq data from melanoma and RT-qPCR in a collection of melanoma cell lines all converged to show that LENT was primarily expressed in melanocytic, but not mesenchymal type melanoma cells. Its expression did not correlate with poor survival in primary melanoma, that rather correlated with the presence of mesenchymal cells that promote invasion and metastases (6). In contrast, its expression correlated with poor survival in metastatic melanoma where it was expressed in proliferative melanocytic type cells marked by an OxPhos signature consistent with the high expression of MITF and SOX10 (15, 36, 37). ASO-mediated LENT silencing induced apoptosis in cultured melanoma cells and impaired xenograft tumour growth. Moreover, simultaneous ASO targeting of LENT together with LENOX or SAMMSON cooperatively impacted melanoma cell viability. Thus ASO-targeting

of these lncRNAs individually or in combination highlights their potential as therapeutic targets.

LINC00520 has been the focus of previous studies designated as LASSIE (38) or LEENE (39). LASSIE was described as a lncRNA induced in endothelial cells by sheer-stress that interacts with PECAM-1 to regulate vascular homeostasis by stabilizing adherens junctions. On the other hand, Miao et al (39) reported that LEENE was induced by pulsatile or oscillatory sheer stress in endothelial cells, but was localized in the nucleus and acted as an enhancer (e)RNA to regulate eNOS expression. Moreover, LEENE was further shown to promote transcription of pro-angiogenic genes, angiogenesis and tissue repair following ischemia (40). These observations contrast with melanoma cells where RNA-scope and cell fractionation showed that LENT was predominantly cytoplasmic being distributed between the cytosol and the mitochondria. We did not see enrichment of PECAM-1 in the RNA pulldown/mass-spectrometry experiments and LENT silencing induced only minor changes in gene expression and did not affect NOS3 (eNOS) expression. This comparison between our data and that previously reported shows that LINC00520 is a multifunctional RNA functioning in a cell-type and context dependent manner as an eRNA in the nucleus to regulate gene expression or in the cytoplasm to regulate translation.

LENT coordinates ribosomal association of mRNAs encoding proteins involved in ER and mitochondrial homeostasis in melanocytic melanoma cells.

We found that LENT selectively and directly interacts with the DHX36 G4 resolvase suggesting that it may contain a G4 structure. The *in vitro* interaction between LENT and DHX36 was reduced, but not abolished by mutation of a G-rich sequence with predicted potential to form a G4 structure. Nevertheless, this G-rich sequence diverges considerably from more canonical G4-forming sequences and LENT does not comprise the 5'-GGnGGnGG-3'

motif or other motifs previously shown to be enriched in DHX36-associated RNAs (23, 28). G4 forming sequences are however variable with the length and sequence of the loop regions between the G blocks contributing to selectivity (41, 42) with DHX36 showing high specificity for parallel G4 structures (19). Moreover, LENT specifically pulled down DHX36, but no other well characterized G4 resolvases such as the RECQL-family including BLM and WRN, nor DDX5, DDX11, DHX9 or DDX3X (41, 42) (43) (25). While several of these helicases are mainly nuclear and more specific for G4 structures in DNA, DDX3X, DDX5 and DHX9 for example are reported RNA G4 resolvases (25, 43, 44). The selectivity of the LENT-DHX36 interaction may therefore reflect specific features of a potential LENT G4 sequence or alternatively, this interaction may be mediated by sequences or structures in LENT independent of G4-formation.

Here we provide evidence that LENT modulates DHX36 interaction with a subset of mRNAs. DHX36-RIP from control 501Mel cells identified DHX36-associated RNAs of which around 40% were previously identified as harboring G4 structures by other methods in other cell types. The most strongly associated RNAs were enriched in potential G4 forming motifs and the 5'-GGnGGnGG-3' motif (28) (23). Surprisingly however, the overlap between the DHX36-associated RNAs in the HEK293T cells used by Sauer et al (28) and the melanoma cells was much lower than seen with the other G4 enrichment protocols, perhaps reflecting the different gene expression profiles in these cell lines. DHX36 IP after LENT silencing identified RNAs whose association with DHX36 was either increased or decreased. One possibility is that LENT comprises a G4 structure that simply competes with the G4s in other RNAs for DHX36 binding hence explaining their increased association upon LENT silencing. However, this competition mechanism cannot explain the reduced binding of RNAs with DHX36 seen upon LENT silencing. How LENT binding to DHX36 modifies its interaction with these RNAs in a positive or negative manner remains to be determined.

An important observation of this study is the association of DHX36 with ribosomes, in the 80S and the polysome fractions. This observation contrasts with that of Sauer et al, where DHX36 was not readily seen in these fractions prepared from HEK293T cells (28), but rather is in accordance with Murat et al, (43) who found DHX36 associated with the 80S and polysome fractions in HeLa cells. The additional presence of LENT in the 80S and light polysome fractions suggested that it may modulate the selectivity of DHX36 to resolve G4 structures in target mRNAs promoting/inhibiting their translation. In accordance with this idea, RNA-seq following polysome profiling in LENT silenced cells showed a pronounced and selective depletion of a set of mRNAs in the LP fractions. Association of many of these mRNAs with the 80S and HP fractions was also reduced, but to a lesser extent. Of these mRNAs, 74 also showed reduced interaction with DHX36 upon LENT silencing. In contrast, several mRNAs whose association with DHX36 was up-regulated upon LENT silencing were enriched in the 80S, LP and HP fractions. Immunoblots showed that mRNAs whose engagement with the LP fractions was promoted by LENT were less well translated upon its silencing and *vice versa*.

Together the above results support the idea that the LENT, via interaction with DHX36, positively or negatively regulates engagement of mRNAs with polysomes and fine-tunes their subsequent translation. However, not all of the mRNAs whose association with the LP fraction was depleted displayed reduced interaction with DHX36 upon LENT silencing suggesting that their interactions with DHX36 are less stable or take place only when they engage with the ribosome. Despite this, these mRNAs are enriched in G4 sequences suggesting that they nevertheless required DHX36-driven unwinding for engagement in the LP fraction. Nevertheless, we cannot formally exclude the existence of alternative DHX36-independent mechanism by which LENT regulates their polysome engagement. We note that it was previously reported that DHX36 silencing often had only minor effects on translation of its associated mRNAs. Quantitative mass-spectrometry revealed only marginal changes in

translation of DHX36 associated mRNAs upon its silencing (28). Here the effects are even more restricted as only a subset of DHX36 associated mRNAs were regulated by LENT, some positively leading to increased translation and *vice versa*.

Despite the above caveats, the idea that the LENT-DHX36 axis regulates translation is in accordance with previous studies reporting that DHX36 unwinds G4 structures in mRNA to regulate their association with ribosomes and their translation in HEK293T and HeLa cells (28) (43). Similarly, DHX36 binds and regulates translation of the mRNA encoding GNAI1 and other mRNAs involved in skeletal muscle stem cell function (23). In these studies, silencing of DHX36 was shown to fine tune translation suggesting that when silenced DHX36 function can be carried out by one or several of the above-mentioned helicases. We rather showed how the selectivity of DHX36 was regulated by LENT. Indeed, the changes in mRNA engagement with polysomes seen upon LENT silencing were more pronounced than those reported for DHX36-associated mRNAs in HeLa and HEK293T cells upon DHX36 silencing. Thus, when present, DHX36 acts as a predominant mRNA G4 resolvase whose activity was modulated by LENT.

LENT is not the first lncRNA shown to affect DHX36 activity. Matsumura et al, (45) identified a cytoplasmic G4-containing lncRNA designed GSEC that binds and inhibits DHX36 promoting motility of colon cancer cells. Similarly, SMaRT is a lncRNA that binds the G4 of the MLX-γ isoform preventing unwinding by DHX36 and repressing its translation in murine muscle differentiation (46). Nevertheless, while GSEC and SMaRT seem to act as molecular decoys to inhibit DHX36 function, LENT is unique in its ability to both positively and negatively impact DHX36 function in an RNA selective manner.

LENT suppresses autophagy to promote melanoma cell survival.

As described above, LENT promotes engagement of a collection of mRNAs with the LP fractions. Strikingly, these mRNAs are strongly enriched in multiple aspects of ER and protein homeostasis, encoding numerous subunits of several protein complexes or pathways, such as SEL1-SYNV1 required for ERAD, TAP1, TAP2 and CALR involved in HLA transport and antigen presentation, enzymes and machinery involved protein glycosylation or the SEC61-NOMO multi-pass complex. Reduced translation of these mRNAs would be expected to lead to accumulation of mis-folded and/or mis-localized proteins and ER stress. For example, down-regulation of WFS1 has previously been shown to lead to reduced OxPhos capacity, increased mitochondrial-lysosome contact and mitophagy similar to what was observed here (47) (48). Consequently, a major phenotype of LENT silencing is autophagy and mitophagy that rapidly appeared in LENT silenced cells associated with impaired OxPhos capacity.

A further consequence of impaired ER function and mitophagy is accumulation of mitochondrial proteins in the cytoplasm and nucleus through their increased translation and/or impaired ER-degradation leading to mPOS, proteotoxic stress and finally apoptosis that may further involve activation of the DNA damage response. Our data therefore support the idea that the major function of LENT is to fine-tune translation of these mRNAs and optimize ER/protein homeostasis, maintain OxPhos capacity, suppress autophagy/mitophagy and promote survival and proliferation of melanocytic melanoma cells.

Previous studies showed that lncRNA SAMMSON acts to coordinate cytoplasmic and mitochondrial translation in melanoma cells to antagonize mPOS-mediated apoptosis (14) (13). Here we show that LENT also regulates translation, but via a different mechanism, to antagonize proteotoxic stress and autophagy-mediated apoptosis. Fine-tuning of translation and protein homeostasis therefore seem to be critical lncRNA-regulated processes in proliferative melanoma cells with a common feature being optimization of mitochondrial function and OxPhos capacity.

Why melanocytic melanoma cells, as opposed to mesenchymal melanoma cells where LENT is not expressed, specifically require fine-tuning of translation and high OxPhos capacity is unclear. However, one likely possibility is that melanocytic melanoma cells, like normal melanocytes, often synthesize melanin, a process that involves production of reactive oxygen species rendering them particularly vulnerable to oxidative stress (49) (50). Optimization of mitochondrial and ER function may therefore be essential to antagonize oxidative stress and may therefore explain why proliferative melanoma cells exploit such diverse mechanisms to optimize mitochondrial homeostasis and OxPhos capacity and ensure cell viability.

Materials and Methods

Analysis of the TCGA-SKCM cohort

For analysis of TCGA-SKCM, raw-counts were retrieved and primary tumors were separated from distant metastasis samples. The raw-counts matrices were normalized by sequencing depth using DESeq2 size-factors and then gene-counts were divided by median transcript length. Consensus clustering was done in R using the ConsensusClusterPlus v3.17 package following standard procedure. In short, matrices were filtered to keep only coding genes based on their biotype annotation and the 5000 most variable genes were selected with the mad() function. The matrices were median centered with sweep(), apply() and median() functions before performing consensus clustering with ConsensusClusterPlus() using base parameters. The number of clusters were selected based on the curve of cumulative distribution function in order to define 4 clusters for primary tumors (CCP1-CCP4) and 5 clusters for distant metastasis samples (CCM1-CCM5).

Cell culture and transfections

Melanoma cell lines SK-MEL-25, SK-MEL-25R, SK-MEL-28, and 501mel were grown in RPMI1640 w/o HEPES medium supplemented with 10 % fetal calf serum (FCS) and

gentamycin (40 µg/mL); IGR-37 and IGR-39 in RPMI1640 w/o HEPES medium supplemented with 15% FCS and gentamycin (40 µg/mL). MM011, MM117, MM047, and MM099 were grown in HAM-F10 medium supplemented with 10 % FCS, 5.2 mM glutamax, 25 mM Hepes, and penicillin/streptomycin (7.5 µg/mL). M229, M229R, M249, and M249R were grown in DMEM medium supplemented with glucose (4.5 g/L), 5 % FCS, and penicillin/streptomycin (7.5 µg/mL). A375 cells were grown in DMEM medium supplemented with glucose (4.5 g/L), 10 % FCS, and gentamycin (40 µg/mL). HEK293T cells were grown in DMEM medium supplemented with glucose (1 g/L), 10 % FCS, and penicillin/streptomycin (7.5 ug/mL). HeLa cells were grown in DMEM medium with glucose (1 g/L), 5 % FCS and gentamycin (40 µg/mL). To assess cell growth and viability, cells were stained with Trypan Blue (Invitrogen). Trametinib (GSK1120212) and dabrafenib (GSK2118436) were purchased from Selleckchem. SK-MEL-25, Sk-MEL-28, A375, and 501mel were obtained from ATCC, all other cell lines were gifts from collaborators. All cell lines were regularly tested using the Venor GeM Mycoplasma Detection Kit, and used at less than 10 passages.

ASO and siRNA were transfected using Lipofectamine RNAiMAX (Invitrogen) with 20 nM of ASO (Qiagen) or siRNA (Thermo Fisher Scientific). ASO and siRNAs sequences are listed in Supplementary Table S1. For ASO combination experiments, cells were transfected with 15 nM of LENT ASO and/or 15 nM of LENOX ASO and/or 5 nM of SAMMSON ASO. For trametinib+dabrafenib-GapmeR cotreatment, cells were cultured for 3 days in presence or absence of Dabrafenib (100 nM) + Trametinib (100 nM), transfected with 15 nM of GapmeR and then cultured for additional 3 days before harvesting. Colony-forming ability was assessed by plating 500 cells/9.6 cm^2, wait for 10 days, fixing cells in formalin and staining with 0.05 % Crystal Violet solution (Sigma Aldrich).

CRISPR interference

501mel cells were co-transfected with a plasmid expressing dead Cas9 protein fused to the Kruppel-associated box (KRAB) domain-containing KAP1 (dCas9-KAP1) and the red fluorescent protein mScarlet (pX-dCas9-KRAB-Scarlet), together with another plasmid expressing GFP and three single guide RNAs targeting the transcription start site of LENT (pcDNA3-sgRNA-GFP) or a control plasmid expressing GFP only (pCMV-GFP). Double Scarlet-GFP positive cells were sorted 24 hours after co-transfection, stained with Cell Trace Violet and cultured for additional 96 hours.

Plasmid cloning and lentiviral transduction

For the ectopic expression experiment, LENT cDNA was cloned into the pCW57-GFP-P2A-MCS vector (a gift from Adam Karpf; Addgene plasmid #71783; http://n2t.net/addgene:71783; RRID: Addgene_71783). LENT shRNA (shLENT) or a scrambled control (shCTRL) were cloned in LT3GEPIR (a gift from Johannes Zuber; Addgene plasmid #111177; http://n2t.net/addgene:111177; RRID: Addgene_111177). Lentiviral particles were produced in HEK293T cells, purified by ultracentrifugation, and resuspended in PBS. Lentiviruses were titrated with flow cytometry by measuring the GFP signal intensity in HEK293T infected with different dilutions of viruses. Melanoma cells were eventually infected at a multiplicity of infection (MOI) of 1 and selected by puromycin addition to the media (1 mg/mL) in every following passage.

RNAscope

LENT and MITF RNAs were detected with the RNAscope assay (Advanced Cell Diagnostics, ACD) according to the manufacturer's protocol. Patient sections were deparaffinized, incubated with hydrogen peroxide at room temperature for 10 minutes, boiled with target retrieval reagent for 15 minutes, and then treated with protease plus reagent at 40 °C for 30 minutes. Sections were hybridized with Hs-MITF probe (ACD, catalog no. 310951) and hs-LENT at 40 °C for 2 hours. Probes for Hs-LENT were custom designed by ACD. Hybridization signals were

amplified and visualized with RNAscope Multiplex Fluorescent Reagent Kit v2 (ACD, catalog no. 323100). For co-detection of DHX36 with LENT, cells were fixed for 30 minutes with formaldehyde 3.7 %, washed with PBS and incubated 10 minutes at room temperature with H_2O_2. After one wash in distilled water, primary antibody for DHX36 diluted in co-detection diluent (1/200) was added o/n at 4 °C. Slides were washed in PBS + tween 0.1 % (PBST), fixed in formaldehyde 3.7 % for 30 minutes, and washed again in PBST. Slides were treated with protease III and washed with PBS. LENT hybridization signals were amplified following the Multiplex Fluorescent Kit. Finally, DHX36 signal was developed by secondary antibody incubation (diluted 1/2,000 in co-detection diluent), followed by tyramide signal amplification (TSA Plus Kit, NEL760001KT, Perkin Elmer). Images were captured with a confocal (Leica DMI6000) microscope. Mander's and Pearson's coefficients were calculated with the Fiji software using the JACoP plugin.

Analysis of oxygen consumption rate in living cells

Oxygen consumption rate (OCR) was measured in an XF96 extracellular analyzer (Seahorse Bioscience). 20,000 transfected cells per well were seeded 48 hours prior the experiment. The cells were incubated at 37°C and the medium was changed to XF base medium supplemented with 1 mM pyruvate, 2 mM glutamine, and 10 mM glucose for 1 hour before OCR profiling with the Mitostress Test Kit sequentially exposed to 2 µM oligomycin, 1 µM carbonyl cyanide-p-trifluoromethoxyphenylhydrazone (FCCP), and 0.5 µM rotenone and antimycin A. Cells were washed with PBS, fixed with 3 % PFA and permeabilized with 0.2 % triton. Nuclei were counterstained with DAPI (1:500) and number of cells per well was determined with a Celomics Cell Insight CX7 (Thermofisher Scientific).

Flow cytometry

To assess cell viability and proliferation, cells were stained with Cell Trace Violet (Invitrogen) on the day of transfection, harvested after 72 hours and stained with Annexin V (BioLegend)

and TOPRO-3 (Invitrogen) or the active caspase-3 Kit (BD Biosciences). Cells were analyzed on a LSRII Fortessa (BD Biosciences) and data were analyzed with FlowJo software (TreeStar). To analyse intracellular ROS, cells were stained in adherent conditions with CellRox Deep Red (Thermo Fisher Scientific) at final concentration of 500 nM following manufacturer instructions. After harvesting, cells were stained for active caspase-3 (BD Biosciences) and analyzed on a LSRII Fortessa (BD Biosciences). To induce reactive oxygen species (ROS), cells were treated with THBP (200 μM) for 30 minutes. To induce apoptosis, cells were treated with staurosporine (500 nM) for 16 hours.

LENT pulldown and LC/MS-MS analysis

501mel cells were grown in 15 cm petri dishes, harvested by trypsinization, washed, pelleted, resuspended in lysis buffer (TrisHCl 20 mM pH8, NaCl 200 mM, MgCl2 2.5 mM, Triton 0.05%, DEPC water) supplemented with fresh DTT (1 mM), protease and phosphatase inhibitor cocktail (Thermo Fisher Scientific) and RNAsin (Thermo Fisher Scientific) and kept 20 minutes on ice. For crosslinked pulldown, petri dishes were exposed to 400 mJ/cm^2 of UV radiation with a CL-1000 crosslinker (254 nm lamp) and the concentration of NaCl in lysis buffer was adjusted to 300 mM. Membranes were pelleted at 3,000 g for 3 minutes at 4°C and supernatant precleared for 1 hour at 4 °C with 100 µg of streptavidin-coated sepharose beads (Cytiva). The lysate was incubated 2 hours with streptavidin coated beads and 400 pmol anti-PCA3 or LENT-specific DNA biotinylated oligonucleotides (listed in Supplementary Table S4). Beads were pelleted for 3 minutes at 3,000 g and washed five times with lysis buffer. After final wash beads were divided for RNA and protein extraction. RNA was purified by TRI Reagent and isopropanol precipitation, digested with DNAse, reverse transcribed and analyzed by qPCR. Proteins were eluted by boiling beads in Laemmli sample buffer and separated on NuPAGE Novex 4% to 12% gradient gels. For mass spectrometry analysis, three independent experiments were performed and the entire lane was excised after staining with Simply blue

safe stain solution (Invitrogen). Analysis was performed at the Harvard Medical School Taplin Mass Spectrometry Facility, as described previously (9, 15).

IGR37 xenograft model and ASO treatment

Swiss nude mice were purchased from Charles River Laboratories (France) and housed under specific pathogen-free conditions. Animal care, use, and experimental procedures were conducted in accordance with recommendations of the European Community (86/609/EEC), European Union (2010/63/UE) and the French National Committee (87/848). The ethics committee of IGBCM in compliance with institutional guidelines approved animal care and use (APAFIS#2023010611181767). Mice were injected on the rear flank with $3x10^6$ IGR37 cells resuspended in 100 µL of 1x PBS + Cultrex Basement Membrane Extract (ref. 3432–005–01; R&D Systems) with a 1:1 ratio. Tumor growth was monitored by caliper measurement every two days and volume was calculated with the formula: (4/3 π) * (length/2) * (width/2) * (height/2). After tumors reached 100 mm^3, mice were injected subcutaneously every two days with 15 mg/kg of ASO closed to the tumor for the LENT group, or not injected for the control group. After two weeks of treatment, mice were sacrificed and primary tumors were dissected and mechanically lysed in TRI Reagent for RNA extraction or LSDB for protein extraction.

Immunofluorescence of fixed and live cells

Cells grown on Millicell EZ slides (Millipore) were fixed with 4 % paraformaldehyde for 15 minutes. After two washes with PBS buffer, they were permeabilized in PBS + Triton X-100 0.1 % for 5 minutes and blocked with PBS + 10 % FCS for 20 minutes. Primary antibodies were incubated overnight at 4 °C and after three washes with PBS + Triton 0.1%, cells were stained for 1 hour at room temperature with Alexa Fluor-488 conjugated secondary antibodies (Life technologies) diluted 1/500 in PBS + 10 % FCS. After three washes with PBS + Triton 0.1%, cells were stained with DAPI (final concentration 1 µg/mL) and mounted on microscopy slides with Prolong Gold antifade reagent (Invitrogen). Anti-DHX36 (13159-1-AP) and anti-

HSP60 were diluted 1/200 in PBS + 10 % FCS. Images were captured with a confocal (Leica DMI6000) microscope. DHX36 enrichment at mitochondria was calculated with the ratio of the DHX36 signal overlapping with HSP60 signal over the total DHX36 signal for each cell on field.

For lysotracker + mitotracker experiment, live cells were incubated in medium complemented with Lysotracker Deep Red 1/20 000, Mitotracker Green FM 1/10 000 and Hoechst 33342 1/10 000 for 1 h, washed and then observed with a confocal microscope inside a chamber at 37 °C with 5 % CO_2.

Transmission Electron Microscopy

Samples were fixed by immersion in 2.5 % glutaraldehyde and 2.5 % paraformaldehyde in cacodylate buffer (0.1 M, pH 7.4), washed in cacodylate buffer for further 30 minutes. The samples were postfixed in 1% osmium tetroxide in 0.1 M cacodylate buffer for 1 hour at 4 °C and dehydrated through graded alcohol (50, 70, 90, and 100%) and propylene oxide for 30 minutes each. Samples were oriented and embedded in Epon 812. Semithin sections were cut at 2 µm and ultrathin sections were cut at 70 nm (Leica Ultracut UCT) and contrasted with uranyl acetate and lead citrate and examined at 70 kv with a Morgagni 268D electron microscope (FEI Electron Optics, Eindhoven, the Netherlands). Images were captured digitally by Mega View III camera (Soft Imaging System).

RNA extraction and RT-qPCR

Total RNA isolation was performed using TRI Reagent (MRC) and isopropanol precipitation, according to the manufacturer protocol. Pelleted RNAs were resuspended in water and DNA was depleted using the TurboDnase Free Kit (Thermo Fisher Scientific). RNA was then reverse transcribed with the Superscript IV reverse transcriptase (Thermo Fisher Scientific) following manufacturer instructions. qPCR was carried out with SYBR Green I (Roche) and monitored by a LightCycler 480 (Roche). Target gene expression was normalized using TBP, HBMS, and

RPL13A as reference genes. For polysome profiling normalization was performed using mRNAs encoding GAPDH, TBP and HMBS. Primers for RT-qPCR are listed in the Supplementary Table S2.

Protein extraction and Western blotting

Whole cell extracts were prepared by freeze–thaw technique using LSDB 500 buffer [500 mM KCl, 25 mM Tris at pH 7.9, 10 % glycerol (v/v), 0.05 % NP-40 (v/v), 16 mL DTT, and protease inhibitor cocktail]. Lysates were subjected to SDS-PAGE and proteins were transferred onto a nitrocellulose membrane. Membranes were incubated with primary antibodies in PBS + 5 % BSA + 0.01 % Tween-20 o/n at 4°C. The membrane was then incubated with HRP-conjugated secondary antibody (Jackson ImmunoResearch, 1/2000) for 1 hour at room temperature, and visualized using the ECL detection system (GE Healthcare). Antibodies used are listed in Supplementary Table S3.

Mitochondria fractionation

Mitochondria were isolated with the Mitochondria Isolation Kit (Thermo Fisher Scientific) following manufacturer instructions. Briefly, harvested cells were washed and pelleted, resuspended in buffer A, and incubated 2 minutes on ice. Buffer B was added for 5 minutes, vortexing every minute, and diluted with buffer C. Nuclei were pelleted 10 minutes at 700 × g and supernatant centrifuged for 15 minutes at 3,000 × g. Purified mitochondria were washed in buffer C and lysed in CHAPS 2%.

For mitochondrial protein content analysis after digitonin treatment, mitochondria were purified as described above. Then, purified mitochondria were digested on ice for 15 minutes with increasing concentrations of Digitonin (Invitrogen) in mitochondria isolation buffer (210 mM Mannitol, 70 mM Sucrose, 1 mM EDTA, 10 mM HEPES and protease inhibitors cocktail). Digested mitochondria were pelleted by centrifugation (13 000 g for 10 minutes) and the

supernatant was removed. Digested mitochondria were then lysed and the protein content was analyzed by SDS-PAGE.

For the trypsin treatment, mitochondria were also prepared following the Mitochondria Isolation kit. Mitochondria were subsequently digested 20 minutes on ice with 50 µg/mL Trypsin diluted in mitochondria isolation buffer. Digestion was stopped by adding 120 µg / mL of soybean trypsin inhibitor. Pellets were centrifuged 10 minutes at 13 000 g at 4 °C, and lysed as described above. For cell swelling conditions, mitochondria were digested with trypsin diluted in HEPES-KOH 20 mM at pH 7.5 instead of mitochondria isolation buffer.

Identification of RNAs associated with DHX36

Cells were grown in 15 cm petri dishes, harvested by scraping, resuspended in lysis buffer (20 mM Tris-HCl pH 8, 200 mM NaCl, 2.5 mM MgCl2, 0.05 % Triton, DEPC water) supplemented with DTT (1 mM), protease/phosphatase inhibitor cocktail (Thermo Fisher Scientific) and RNAsin (Thermo Fisher Scientific) and kept on ice for 15 minutes, pipetting every 3 minutes. Membranes were pelleted 10 minutes at 10,000 g at 4 °C and the supernatant precleared 1 hour at 4 °C with protein G magnetic beads (Invitrogen). Lysate was quantified by Bradford protein quantification assay (Bio-Rad) and incubated overnight on a rotating wheel at 4 °C with 5 µg of the indicated antibodies. Then, 50 µL of resuspended Protein G magnetic beads were added for 3 hours to isolate RNA – protein complexes and washed five times in lysis buffer. After final wash, RNA was purified by TRI Reagent + isopropanol precipitation and proteins eluted by boiling beads at 95 °C for 15 minutes in Laemmli buffer.

For RNA sequencing, RNAs were obtained from 501mel cells expressing a control shRNA or an shRNA targeting LENT following the method described above. RNA profiles were determined by using a 2100 Bioanalyser. rRNAs were depleted with the Ribo-Zero Plus rRNA depletion kit (Illumina) and the libraries were prepared with the library prep mRNA ultralow

Smarter kit (Takara). Sequencing was performed on a NextSeq 2000 high throughput sequencer (Illumina). Analyses were performed as described in the next paragraph.

Bulk RNA sequencing data

Gene expression in 501mel cells transfected with control or LENT-targeting ASO was analyzed by RNA-seq. After sequencing raw reads were pre-processed in order to remove adapter and low-quality sequences (Phred quality score below 20) using cutadapt version 1.10. and reads shorter than 40 bases were discarded. Reads were mapping to rRNA sequences using bowtie version 2.2.8, were also removed. Reads were mapped onto the hg19 assembly of Homo sapiens genome using STAR version 2.5.3a. Gene expression quantification was performed from uniquely aligned reads using htseq-count version 0.6.1p1, with annotations from Ensembl version 75 and "union" mode. Only non-ambiguously assigned reads were retained for further analyses. Read counts were normalized across samples with the median-of-ratios method. Comparisons of interest were performed using the Wald test for differential expression and implemented in the Bioconductor package DESeq2 version 1.16.1. Genes with high Cook's distance were filtered out and independent filtering based on the mean of normalized counts was performed. P-values were adjusted for multiple testing using the Benjamini and Hochberg method.

Motif enrichment analysis

To identify RIP-seq genes containing G-quadruplex regions, we first retrieved publicly available RIP-seq data for DHX36 reported by Varshney et al (16) from GEO (accession: GSE154570). We then performed de-novo motif analysis using the MEME-ChIP algorithm on DHX36 RIP retained RNAs identifying the G-quadruplex motif "CCGCCGCY" and generating the associated probability matrix. Lastly, we used the FIMO algorithm to find this motif in the 5'UTR regions of genes identified in the DHX36 RIP-seq analysis.

RNA in-vitro transcription and purification

Double-stranded DNA molecules (gBlocks) containing LENT WT or LENT ΔG sequences were ordered from Integrated DNA Technologies and used as PCR templates to generate RNA by in-vitro transcription. DNA constructs were designed with the T7 RNA polymerase promoter sequence. After run-off transcription, RNAs were purified by denaturing polyacrylamide gel electrophoresis (PAGE) and extracted by the "crush and soak method" as described (17). For EMSA experiments, the purified RNA transcripts were labelled at their 5′ end by addition of a radioactive cap using the Vaccine Capping Enzyme with the ScriptCap m7G capping system from CELLSCRIPT in the presence of [32P] αGTP (>6000 Ci/mmol). The 5′-radiolabelled transcript was separated from enzyme and free nucleotides by Bio-Spin 6 Columns (Biorad).

DHX36 Protein Expression and Purification

The inducible expression plasmid (pDHX36 54-989), encompassing the sequence of the human protein (aa 54 to 898) with a His6-SUMO N-terminus tag (18), was generously provided by Rick Russell. Recombinant protein expression was carried out in BL-21 DE3 Rosetta2 pLysS cells (Merck). Cells were grown at 37 °C until reaching an OD-600 nm value of 0.9. Subsequently, the temperature was lowered to 18 °C, and protein expression was induced by adding 0.5 mM IPTG for 16 hours. Cells were harvested by centrifugation, and the pellet was resuspended in lysis buffer [50 mM Tris-HCl pH 8.0, 1 M NaCl, 10 % Glycerol, 10 mM 2-Mercaptoethanol, 1 mM CaCl2, 10 µg/mL Dnase I (Merck), 1x Halt-Protease (Pierce)] before being lysed by sonication at 4 °C. The lysate was clarified by centrifugation, and nucleic acids were precipitated using 0.1 % polyethyleneimine and removed by centrifugation. The supernatant was then passed through a Ni-NTA column (Protino, Macherey-Nagel) equilibrated with buffer A (50 mM Tris-HCl pH 8.0, 1 M NaCl, 10 % Glycerol). After extensive washing with the equilibration buffer, the protein was eluted with buffer A supplemented with 300 mM Imidazole. Factions containing the protein were treated with ULP Protease at a ratio of 1:500 (W/W) and digested/dialyzed overnight at 4 °C in buffer B (50 mM Tris-HCl pH 8.0, 350 mM

NaCl, 10 mM 2-Mercaptoethanol, 10 % Glycerol). The DHX36 54-989 protein, liberated from its N-terminal tag, was separated on a NiNTA column equilibrated with buffer B, with the protein predominantly found in the flow-through fractions. Further purification was accomplished via chromatography on a Heparin affinity column (Hitrap HP, Cytiva). The DHX36 54-989 protein was eluted by a linear gradient from 10 to 500 mM NaCl in buffer (50 mM Tris-HCl pH 8.0, 10 % Glycerol, 10 mM 2-Mercaptoethanol), resulting in a symmetrical peak, and isolated at the end of the gradient. The protein fractions were concentrated in the elution buffer to a final concentration of 7 µM (concentration determined using OD at 280 nm, extinction coefficient 108180 M-1cm-1), snap-frozen in liquid nitrogen, and stored at -80°C. The identity of the final purified protein was confirmed by mass spectrometry analysis (performed at Strasbourg-Esplanade Proteomics Facility). Dynamic light scattering (DLS) was used to investigate the solubility of the purified protein.

EMSA

5' end labeled LENT WT and/or LENT ΔG RNA (10,000 cpm; < 3 nM) and a molar excess of oligo(dT) in 8 µl of milli-Q (Millipore) water were heated for 2 min at 90 °C and chilled on ice for 2 min. After addition of tenfold concentrated refolding buffer (50 mM MES/NaOH pH 6, 100 mM KOAc, 1 mM $Mg(Oac)_2$ and one unit of Rnasin (Promega), RNA was renatured for 15 min at RT. 10 µl of RNA was finally incubated 30min in ice with increasing concentrations of DHX36 (0-125nM) in twofold concentrated binding buffer (final concentration: 50mM Tris/HCl pH8, 150mM NaCl, 10% glycerol, 10mM β-mercaptoethanol). Electrophoresis was performed in TBM (89 mM Tris base, 89 mM boric acid and 1 mM $Mg(Oac)_2$) buffer at 120 V for 5 h at 4 °C. Results were analyzed by phosphorimaging. Quantitative analysis was performed using ImageLab software (Biorad).

Sucrose density gradient ribosome profiling

Cells were washed twice with PBS, harvested by scraping and rapidly centrifuged at 300 g for 5 min at 10 °C. The resulting pellet was resuspended in lysis buffer (20 mM HEPES/NaOH pH 7.4, 100 mM KOAc, 1 mM DTT, 0.5 mM Mg(Oac)2, 100 U of Recombinant Rnasin (Promega) and Halt Protease inhibitor cocktail (ThermoFisher)). Cell lysis was performed by nitrogen cavitation with 4639 Cell Disruption Vessel (Parr Instrument Company) at 350 psi for 50 min, stirring with a small magnet at 500 rpm in a cold room (4 °C). Lysate was then centrifuged at 1000 g for 5 minutes and the supernatant was recovered avoiding the foam (membranes) and the pellet (nuclei). After an incubation of 5 min at 30 °C, 20–30 OD260 of cell extracts from both cell lines were gently layered over 7–47 % sucrose gradients in buffer T (25 mM Hepes/NaOH pH 7.4, 79 mM KOAc, 2.5 mM Mg(Oac)2, 1 mM DTT, 3 U/µl Rnasin (Promega). Gradients were centrifuged at 37000 rpm (Beckman, SW41Ti) for 2 h and 30 min at 4 °C. After centrifugation, 45 fractions (0.25 ml/ fraction) were collected on a BIOCOMP gradient fractionator equipped with an UV detector. For RNA-seq, biological triplicate ribosome profiling was performed from control or shLENT-silenced cells as described above. RNA from three fractions corresponding to the 80S, light or heavy polysome fractions was pooled and sequenced using the Ribo-Zero Plus rRNA depletion kit (Illumina) as described above.

Statistics

All tests used for statistical significance were calculated using GraphPad Prism10 and indicated in the figure legends along with p values (****$p < 0.0001$, ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, ns: $p > 0.05$).

Resources.

All oligonucleotides and antibodies used are listed in Supplementary Tables 1-4.

Data availability.

The datasets generated during and/or analyzed in this study are available from the corresponding author I. Davidson upon request. The RNA-seq data described in this paper have been deposited with the GEO data base under the accession number GSE270716.

Table 1

GapmeR/siRNA/shRNA	Sequence
ASO NEG	AACACGTCTATACGC
ASO LENT 1	TTTGATGAGTGAGTCG
ASO LENT 2	GAGTCGCTGAGAATTA
ASO LENOX	GTAGAGGCTAGAACTG
ASO SAMMSON	GTGTGAACTTGGCT
shSCR	ATTACGTCTGTCATGAACCTC
shLENT	CCTTCCAAGCATTGCCTTTAT
siCTRL	UGGUUUACAUGUUGUGUGA
siDHX36	CGGCAUGUGGUACGCGAAA

Table 2

RT-qPCR Primers	Sequence
GAPDH_F	ACAACTTTGGTATCGTGGAAGG
GAPDH_R	GCCATCACGCCACAGTTTC
RPL13a_F	TTGAGGACCTCTGTGTATTTGTCAA
RPL13a_R	CCTGGAGGAGAAGAGGAAAGAGA
HMBS_F	GGCAATGCGGCTGCAA
HMBS_R	GGGTACCCACGCGAATCAC
TBP_F	CGGCTGTTTAACTTCGCTTC
TBP_R	CACACGCCAAGAAACAGTGA
LENOX _F	ACCTAACCTGCGAATGCTGT
LENOX _R	GCCTAAACATTTGCTGCCCC
LENT_F	CAATGCTTGGAAGGCGTGAT
LENT_R	AAACGTATGGCCACCTCTGA
MALAT1_F	GGATTCCAGGAAGGAGCGAG
MALAT1_R	ATTGCCGACCTCACGGATTT
UBE4A_F	GAGAGCCAAGGAAGAGATTACCA

UBE4A_R	CTTGTTCATGTACTCACGGGC
RBPJ_F	GGAAAGAGCAAAGGAGGGGA
RBPJ_R	TCACCAAATTTCCCAGGCGA
NOX4_F	CACCAGATGTTGGGGCTAGG
NOX4_R	CTCCTGGTTCTCCTGCTTGG
SDHB_F	AGGATCTTGTTCCCGATTTGAG
SDHB_R	CGTAGAGCCCGTCCAGTTTC
ATP5A_F	CTGCAAAGATGCTGTCCGTG
ATP5A_R	GCATTTCTGGAGACCAGTCC
UQCRC2_F	CCAAGCTGCCAAGAACAAGC
UQCRC2_R	CAGCAACTAGAGCCTGGGAC

Table 3

Primary Antibodies	Host	Application	Dilution	Lot number
DHX36 (13159-1-AP)	Rabbit	WB IF IP	1/1000 1/200 5 µg/mL	00004147
CARF (BE-A303-861A-M)	Rabbit	WB	1/1000	1
RAP2 (sc-515711)	Mouse	WB	1/1000	B4397
HSP60 (in house)	Mouse	WB IF	1/500 1/100	4MTE-2H7
VINCULIN (V4505)	Mouse	WB	1/5000	099M4850V
ACTIN (in house)	Mouse	WB	1/1000	1ACT-2D7
COX IV (ab202554)	Rabbit	WB	1/1000	GR3342068-1
UBE4A (sc-365904)	Mouse	WB IF	1/1000 1/200	C3117
NOX4 (NB110-58849)	Rabbit	WB IF	1/1000 1/200	D134519
EIF4A2 (ab31218)	Rabbit	WB	1/1000	GR3383097-1
RPL36 (PA5-117106)	Rabbit	WB	1/2000	WG3327290E
C1QBP (A302-863A) = p32	Rabbit	WB IF	1/1000 1/200	1

TOMM20 (H00009804-M01)	Mouse	WB	1/1000	HC5202153D
RBP-Jk (sc-271128)	Mouse	WB IF	1/500 1/100	A2220
OXPHOS (45-8199)	Mouse	WB	1/1000	VB2939036
LC3B (ab51520)	Rabbit	WB	1/5000	GR3374012-2
gH2AX (ab22551)	Mouse	WB IF	1/1000 1/100	GR3358071-2
FUNDC1 (NBP1-81063)	Rabbit	WB	1/500	000042870
CTSD (2284)	Rabbit	WB	1/1000	2
WFS1 (26995-1-AP)	Rabbit	WB	1/500	00096359
HSPA5 (HPA038845)	Rabbit	WB	1/500	A83196
Normal Rabbit IgG (12-370)	Rabbit	IP	5 µg/mL	3493998

1

Table 4

RNA pulldown Probes	Sequence
PCA3-1	GCACTTGCTATTTCTTCTGT
PCA3-2	CTCTGTTTTTCTGATGCCAG
PCA3-3	TGTTTGTTGCATGTCTTGTG
PCA3-4	ATTCTTTATTGCCAGGAGTG
PCA3-5	TATGCATATTGTGGTTGTCC
PCA3-6	TGTCTGAATCCTCTCCAAAC
PCA3-7	GCTAGCATCCATAATAGGAG
PCA3-8	TTGCATGCATGTACCACAAG
LENT-1	GAAATGTACACCATGCTGGG
LENT-2	TTATTTTGCTCCTTGCTGTT
LENT-3	TGAGACCCCAAAGAGGGAAA
LENT-4	CCTTGCTGTTCTCGAAAGAT

LENT-5	CCTGGCTTTGATGATTCAGT
LENT-6	GGCAATGCTTGGAAGGCG
LENT-7	TTGCCACCAATCTCTCTG
LENT-8	GCGTGATAAGCTACCCAG
LENT-9	CAGAGGTGGCCATACGTTTG
LENT-10	GCTTGATGGGGAGAAGGAAG

Acknowledgements

We thank, I. Michel for excellent technical assistance, N. Messaddeq for electron microscopy, R Tomaino and the Harvard mass-spectrometry platform, J-C. Marine and G. Ghanem for the MM-series primary melanoma cells, D. Lipkser and the Dermatology Clinic of Strasbourg University Hospital for patient melanoma sections, the GenomEast high throughput sequencing platform, the staff of the IGBMC common facilities. . This work was supported by grants from the ITMO Cancer, the Ligue Nationale contre le Cancer, SATT Conectus Alsace, the ANR-10-LABX-0030 and ANR-10-IDEX-0002-02. ID is an ‘équipe labellisée’ of the Ligue Nationale contre le Cancer. AH was supported by a fellowship from the Fondation pour la Recherche Médicale, ref: FDT202304016609.

Author Contributions

AH, GG and GM performed all the ASO targeting cell culture and xenograft experiments, GD, AH and GG performed bioinformatics analyses, AH performed the LENT RNA pulldown and subsequent functional experiments, Md’A, GB, AS and EE generated in vitro transcribed LENT, recombinant DHX36, performed the EMSA and the polysome profiling, AH, GG, EE and ID conceived the experiments, analysed the data and wrote the paper.

References

1. Tirosh I, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science*. 2016;352(6282):189–96.

2. Tsoi J, et al. Multi-stage Differentiation Defines Melanoma Subtypes with Differential Vulnerability to Drug-Induced Iron-Dependent Oxidative Stress. *Cancer Cell*. 2018;33(5):890-904 e5.

3. Rambow F, et al. Toward Minimal Residual Disease-Directed Therapy in Melanoma. *Cell*. 2018;174(4):843-855 e19.

4. Rambow F, Marine JC, Goding CR. Melanoma plasticity and phenotypic diversity: therapeutic barriers and opportunities. *Genes Dev*. 2019;33(19–20):1295–1318.

5. Ennen M, et al. MITF-High and MITF-Low Cells and a Novel Subpopulation Expressing Genes of Both Cell States Contribute to Intra- and Intertumoral Heterogeneity of Primary Melanoma. *Clin Cancer Res*. 2017;23(22):7097–7107.

6. Karras P, et al. A cellular hierarchy in melanoma uncouples growth and metastasis. *Nature*. 2022;610(7930):190–198.

7. Pozniak J, et al. A TCF4-dependent gene regulatory network confers resistance to immunotherapy in melanoma. *Cell*. 2024;187(1):166-183.e25.

8. Strub T, et al. Essential role of microphthalmia transcription factor for DNA replication, mitosis and genomic stability in melanoma. *Oncogene*. 2011;30(20):2319–32.

9. Laurette P, et al. Transcription factor MITF and remodeller BRG1 define chromatin organisation at regulatory elements in melanoma cells. *Elife*. 2015;10.7554/eLife.06857. http://dx.doi.org/10.7554/eLife.06857.

10. Verfaillie A, et al. Decoding the regulatory landscape of melanoma reveals TEADS as regulators of the invasive cell state. *Nat Commun*. 2015;6:6683.

11. Mauduit D, et al. Analysis of long and short enhancers in melanoma cell states. *Elife*. 2021;10:e71735.

12. Marin-Bejar O, et al. Evolutionary predictability of genetic versus nongenetic resistance to anticancer drugs in melanoma. *Cancer Cell*. 2021;39(8):1135-1149.e8.

13. Vendramin R, et al. SAMMSON fosters cancer cell fitness by concertedly enhancing mitochondrial and cytosolic translation. *Nat Struct Mol Biol*. 2018;25(11):1035–1046.

14. Leucci E, et al. Melanoma addiction to the long non-coding RNA SAMMSON. *Nature*. 2016;531(7595):518–22.

15. Gambi G, et al. The lncRNA LENOX interacts with RAP2C to regulate metabolism and promote resistance to MAPK inhibition in melanoma. *Cancer Res*. 2022;CAN-22-0959.

16. Varshney D, et al. RNA G-quadruplex structures control ribosomal protein production. *Sci Rep*. 2021;11(1):22735.

17. Petrov A, et al. RNA Purification by Preparative Polyacrylamide Gel Electrophoresis. *Methods in Enzymology*. Elsevier; 2013:315–330.

18. Yangyuoru PM, et al. The G-quadruplex (G4) resolvase DHX36 efficiently and specifically disrupts DNA G4s via a translocation-based helicase mechanism. *Journal of Biological Chemistry*. 2018;293(6):1924–1932.

19. Chen MC, et al. Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36. *Nature*. 2018;558(7710):465–469.

20. Kikin O, D'Antonio L, Bagga PS. QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res*. 2006;34(Web Server issue):W676-682.

21. Antcliff A, McCullough LD, Tsvetkov AS. G-Quadruplexes and the DNA/RNA helicase DHX36 in health, disease, and aging. *Aging*. 2021;13(23):25578–25587.

22. Chen X, et al. Translational control by DHX36 binding to 5'UTR G-quadruplex is essential for muscle stem-cell regenerative functions. *Nat Commun*. 2021;12(1):5043.

23. Chen X, et al. Translational control by DHX36 binding to 5'UTR G-quadruplex is essential for muscle stem-cell regenerative functions. *Nat Commun*. 2021;12(1):5043.

24. Chen X, et al. Lockd promotes myoblast proliferation and muscle regeneration via binding with DHX36 to facilitate 5' UTR rG4 unwinding and Anp32e translation. *Cell Rep*. 2022;39(10):110927.

25. Herdy B, et al. Analysis of NRAS RNA G-quadruplex binding proteins reveals DDX3X as a novel interactor of cellular G-quadruplex containing transcripts. *Nucleic Acids Research*. 2018;46(21):11592–11604.

26. Guo JU, Bartel DP. RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science*. 2016;353(6306):aaf5371.

27. Kwok CK, et al. rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nat Methods*. 2016;13(10):841–844.

28. Sauer M, et al. DHX36 prevents the accumulation of translationally inactive mRNAs with G4-structures in untranslated regions. *Nat Commun*. 2019;10(1):2421.

29. Arnoult D, et al. An N-terminal addressing sequence targets NLRX1 to the mitochondrial matrix. *Journal of Cell Science*. 2009;122(17):3161–3168.

30. Bykov YS, et al. Cytosolic Events in the Biogenesis of Mitochondrial Proteins. *Trends in Biochemical Sciences*. 2020;45(8):650–667.

31. Béthune J, et al. Membrane-Associated RNA-Binding Proteins Orchestrate Organelle-Coupled Translation. *Trends Cell Biol*. 2019;29(2):178–188.

32. Lesnik C, Golani-Armon A, Arava Y. Localized translation near the mitochondrial outer membrane: An update. *RNA Biol*. 2015;12(8):801–809.

33. Wang X, Chen XJ. A cytosolic network suppressing mitochondria-mediated proteostatic stress and cell death. *Nature*. 2015;524(7566):481–484.

34. Coyne LP, Chen XJ. MPOS is a novel mitochondrial trigger of cell death – implications for neurodegeneration. *FEBS Letters*. 2018;592(5):759–775.

35. Karbowski M, Oshima Y, Verhoeven N. Mitochondrial proteotoxicity: implications and ubiquitin-dependent quality control mechanisms. *Cell Mol Life Sci*. 2022;79(11):574.

36. Haq R, et al. Oncogenic BRAF regulates oxidative metabolism via PGC1alpha and MITF. *Cancer Cell*. 2013;23(3):302–15.

37. Vazquez F, et al. PGC1alpha expression defines a subset of human melanoma tumors with increased mitochondrial capacity and resistance to oxidative stress. *Cancer Cell*. 2013;23(3):287–301.

38. Stanicek L, et al. Long non-coding RNA LASSIE regulates shear stress sensing and endothelial barrier function. *Commun Biol*. 2020;3(1):265.

39. Miao Y, et al. Enhancer-associated long non-coding RNA LEENE regulates endothelial nitric oxide synthase and endothelial function. *Nat Commun*. 2018;9(1):292.

40. Tang X, et al. Long noncoding RNA LEENE promotes angiogenesis and ischemic recovery in diabetes models. *Journal of Clinical Investigation*. 2023;133(3):e161759.

41. Varshney D, et al. The regulation and functions of DNA and RNA G-quadruplexes. *Nat Rev Mol Cell Biol*. 2020;21(8):459–474.

42. Lejault P, et al. How to untie G-quadruplex knots and why? *Cell Chemical Biology*. 2021;28(4):436–455.

43. Murat P, et al. RNA G-quadruplexes at upstream open reading frames cause DHX36- and DHX9-dependent translation of human mRNAs. *Genome Biol*. 2018;19(1):229.

44. Wu G, et al. DDX5 helicase resolves G-quadruplex and is involved in *MYC* gene transcriptional activation. *Proc Natl Acad Sci USA*. 2019;116(41):20453–20461.

45. Matsumura K, et al. The novel G-quadruplex-containing long non-coding RNA GSEC antagonizes DHX36 and modulates colon cancer cell migration. *Oncogene*. 2017;36(9):1191–1199.

46. Martone J, et al. SMaRT lncRNA controls translation of a G-quadruplex-containing mRNA antagonizing the DHX36 helicase. *EMBO Reports*. 2020;21(6):e49942.

47. Cagalinec M, et al. Role of Mitochondrial Dynamics in Neuronal Development: Mechanism for Wolfram Syndrome. *PLoS Biol*. 2016;14(7):e1002511.

48. Patergnani S, et al. The Wolfram-like variant WFS1E864K destabilizes MAM and compromises autophagy and mitophagy in human and mice. *Autophagy*. 2024;1–12.

49. Obrador E, et al. Oxidative stress and antioxidants in the pathophysiology of malignant melanoma. *Biol Chem*. 2019;400(5):589–612.

50. Denat L, et al. Melanocytes as Instigators and Victims of Oxidative Stress. *Journal of Investigative Dermatology*. 2014;134(6):1512–1518.

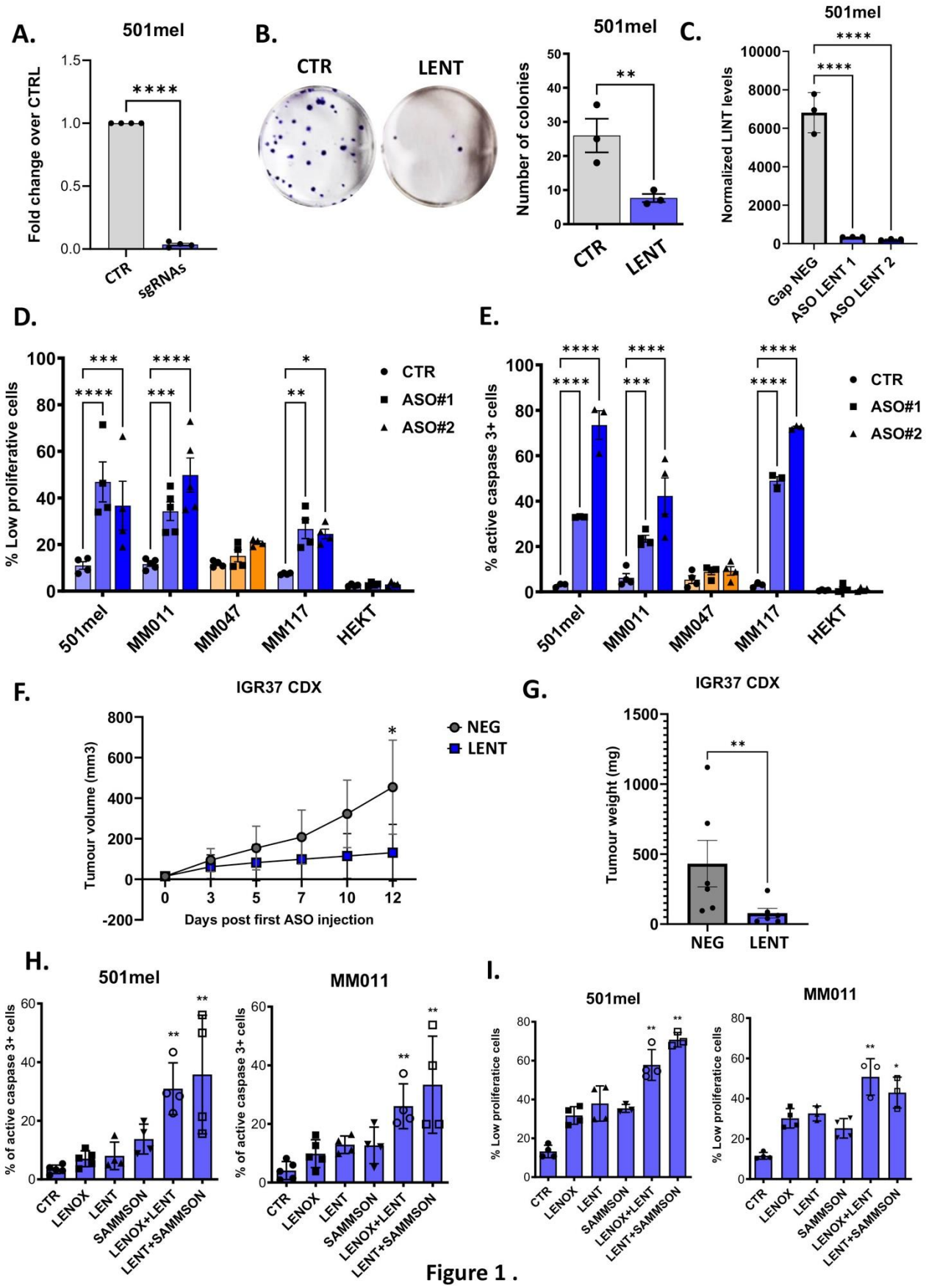
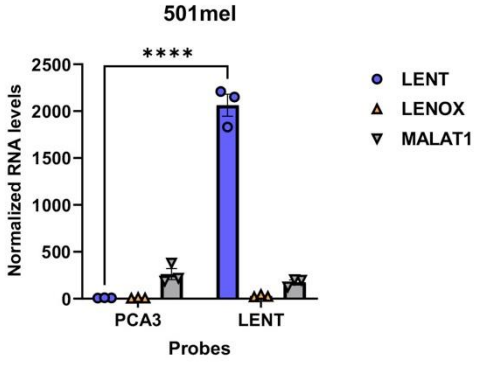

A.
501mel
Fold change over CTRL
CTR
sgRNAs
B.
CTR
LENT
501mel
Number of colonies
CTR
LENT
C.
501mel
Normalized LINT levels
Gap NEG
ASO LENT 1
ASO LENT 2
D.
% Low proliferative cells
CTR
ASO#1
ASO#2
501mel
MM011
MM047
MM117
HEKT
E.
% active caspase 3+ cells
CTR
ASO#1
ASO#2
501mel
MM011
MM047
MM117
HEKT
F.
IGR37 CDX
Tumour volume (mm3)
Days post first ASO injection
NEG
LENT
G.
IGR37 CDX
Tumour weight (mg)
NEG
LENT
H.
501mel
MM011
% of active caspase 3+ cells
CTR
LENOX
LENT
SAMMSON
LENOX+LENT
LENT+SAMMSON
I.
501mel
MM011
% Low proliferatice cells
CTR
LENOX
LENT
SAMMSON
LENOX+LENT
LENT+SAMMSON

Figure 1 .

Figure 1: LENT depletion impairs melanoma cell proliferation and survival. **A.** LENT levels measured by RT-qPCR after dCas9-KAP1-mediated LENT silencing compared by paired t-test. **B.** Colony formation assay upon dCas9-KAP1-mediated LENT silencing, compared by one-way ANOVA (Dunnett test). **C.** LENT levels measured by RT-qPCR after ASO-mediated depletion with two independent ASOs. **D-E.** Measurement of slow proliferating or apoptotic cells by flow cytometry upon ASO-mediated LENT depletion compared by one-way ANOVA (Dunnett test). Melanocytic cell lines are represented in blue and mesenchymal cells are colored in orange. **F-G.** Tumour volumes in mice with IGR37 CDX tumours were measured at the indicated number of days following initial injections of LENT-targeting ASO. Tumours were weighed following sacrifice at day 14. Following the first ASO injection. Volumes were compared by two-way ANOVA and weights by Mann-Whitney test. **H-I.** Same measurements as for D-E but after ASO-mediated depletion of the indicated lncRNAs at sub-optimal ASO doses, compared by one-way ANOVA (Dunnett test). *, $P < 0.033$; **, $P < 0.0021$; ***, $P < 0.0002$; ****, $P < 0.0001$.

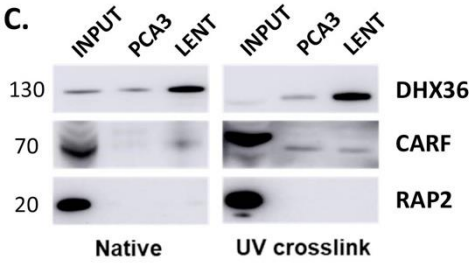
A.



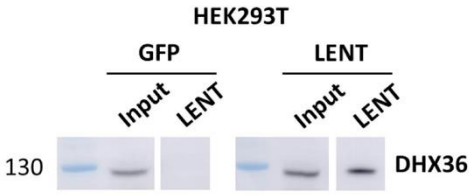
B.

Gene Symbol	Mean PCA3 peptides	Mean LENT peptides	LINT over PCA3 peptides
DHX36	0	17,5	17,50
XRN1	0	8	8,00
EEF1A1	21	51	2,43
DYNC1H1	121	100	0,82
MYH9	74	32	0,43

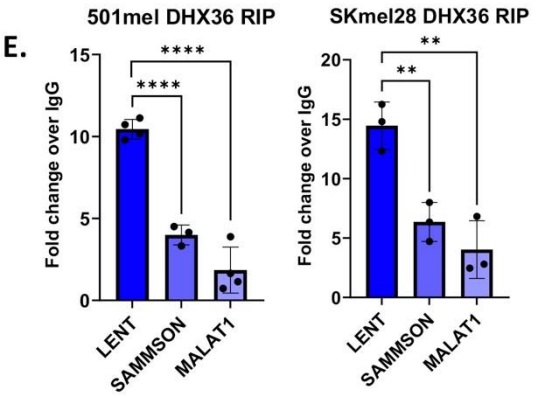
C.



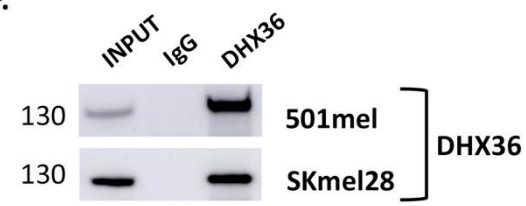
D.



E.



F.



G.

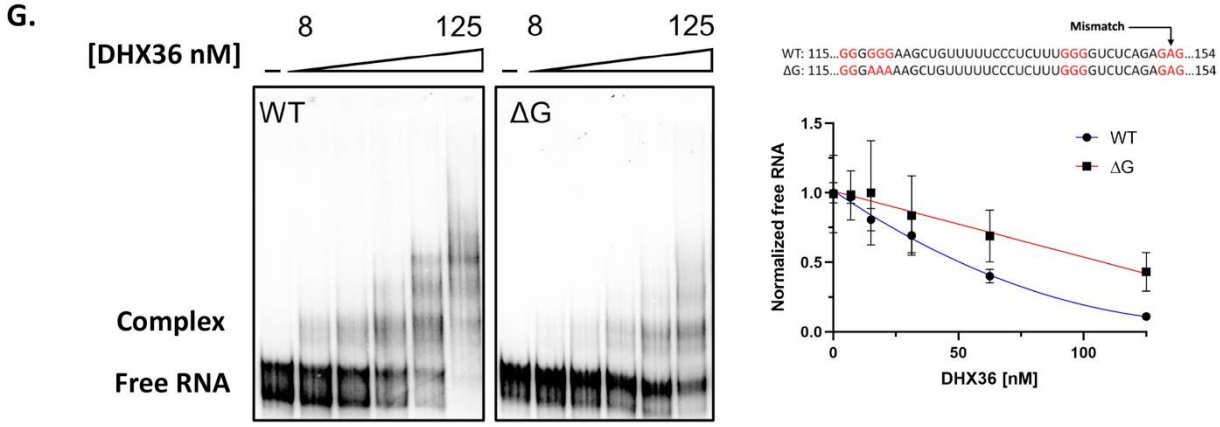


Figure 2.

Figure 2: LENT interacts specifically and directly with the G4 resolvase DHX36. **A.** Measurement of lncRNA levels by RT-qPCR after RNA pulldown with control probes (PCA3) or LENT-targeting probes by two-way ANOVA. **B.** LC/LC Mass spectrometry analysis of peptides retrieved after biological triplicate pulldown of PCA3 or LENT. **C.** Immunoblot showing protein enrichment after the indicated RNA pulldowns, in native or UV crosslinked conditions. **D.** Native LENT pulldown upon ectopic expression of GFP or LENT in HEK293T cells. **E.** Enriched lncRNAs analysed by RT-qPCR after DHX36 immunoprecipitation compared with a control IgG by unpaired t-test. **F.** DHX36 immunoprecipitation using a specific antibody showing DHX36 enrichment by immunoblot. **G.** EMSA assay performed with T7 in vitro transcribed WT or mutated LENT in presence of increasing concentrations of purified truncated DHX36. The potential G4-forming structure predicted by PQSfinder in LENT isoform 5 sequence is indicated in red. The mutated LENT sequence used for the EMSA with the three guanines mutated in adenines underlined. *, $P < 0.033$; **, $P < 0.0021$; ***, $P < 0.0002$; ****, $P < 0.0001$.

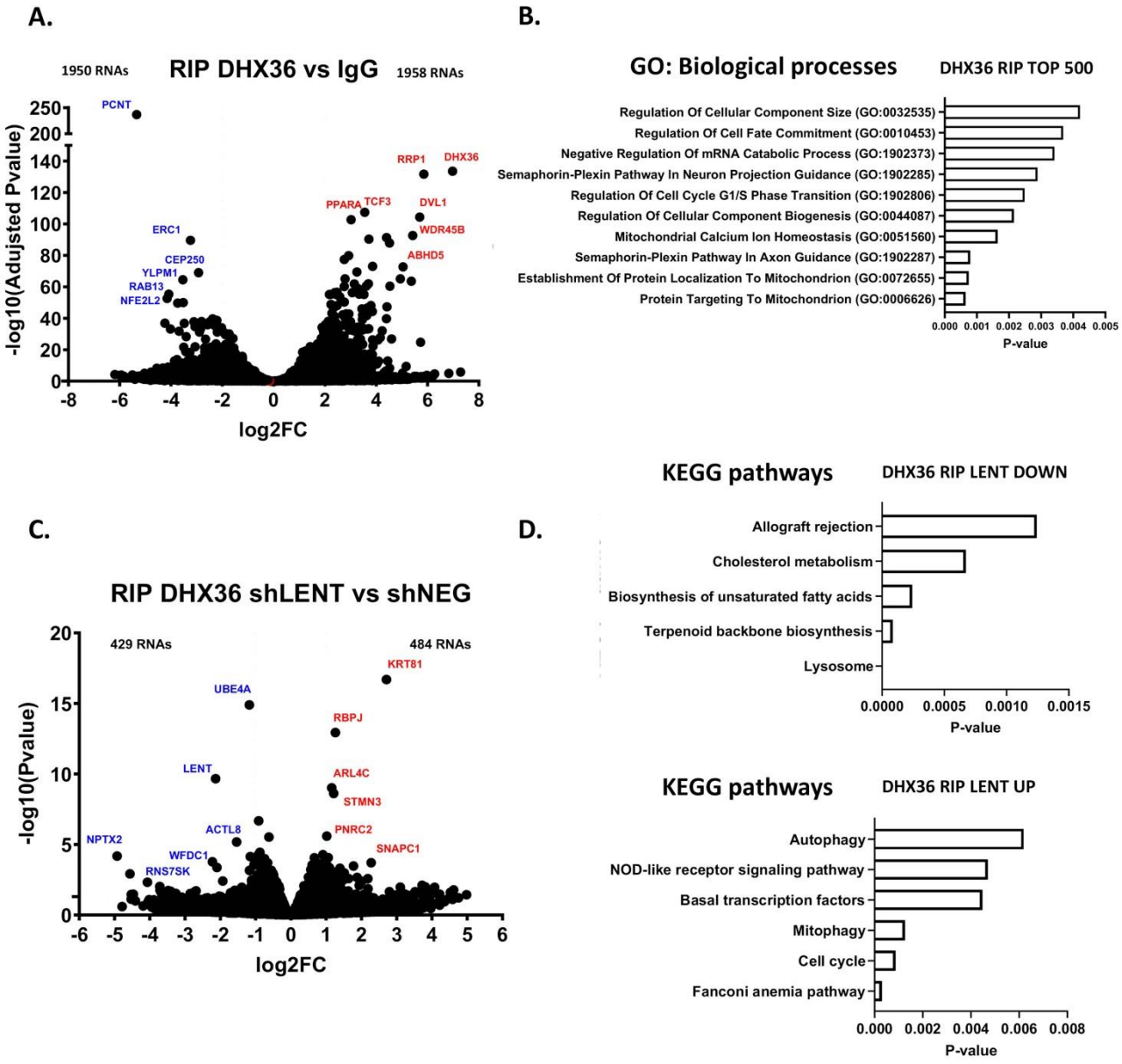

A.
1950 RNAs
RIP DHX36 vs IgG
1958 RNAs
-log10(Adujsted Pvalue)
log2FC
PCNT
RRP1
DHX36
PPARA
TCF3
DVL1
WDR45B
ABHD5
ERC1
CEP250
YLPM1
RAB13
NFE2L2
B.
GO: Biological processes
DHX36 RIP TOP 500
Regulation Of Cellular Component Size (GO:0032535)
Regulation Of Cell Fate Commitment (GO:0010453)
Negative Regulation Of mRNA Catabolic Process (GO:1902373)
Semaphorin-Plexin Pathway In Neuron Projection Guidance (GO:1902285)
Regulation Of Cell Cycle G1/S Phase Transition (GO:1902806)
Regulation Of Cellular Component Biogenesis (GO:0044087)
Mitochondrial Calcium Ion Homeostasis (GO:0051560)
Semaphorin-Plexin Pathway In Axon Guidance (GO:1902287)
Establishment Of Protein Localization To Mitochondrion (GO:0072655)
Protein Targeting To Mitochondrion (GO:0006626)
P-value
C.
RIP DHX36 shLENT vs shNEG
429 RNAs
484 RNAs
-log10(Pvalue)
log2FC
KRT81
UBE4A
RBPJ
LENT
ARL4C
STMN3
ACTL8
PNRC2
SNAPC1
NPTX2
WFDC1
RNS7SK
D.
KEGG pathways
DHX36 RIP LENT DOWN
Allograft rejection
Cholesterol metabolism
Biosynthesis of unsaturated fatty acids
Terpenoid backbone biosynthesis
Lysosome
P-value
KEGG pathways
DHX36 RIP LENT UP
Autophagy
NOD-like receptor signaling pathway
Basal transcription factors
Mitophagy
Cell cycle
Fanconi anemia pathway
P-value

Figure 3.

Figure 3: LENT modulates association of RNA with DHX36. A. Volcano blot showing RNAs enriched or depleted in the DHX36 IP vs control IgG IP. P-values were derived using the Wald test and adjusted using Benjamini-Hochberg FDR correction. **B.** Gene ontology analysis by EnrichR software of the 500 most enriched RNAs in the DHX36 IP. **C.** Volcano blot showing RNAs enriched or depleted in the DHX36 IP upon shRNA-mediated LENT silencing. P-values were derived using the Wald test. **D.** KEGG pathways by EnrichR of RNAs enriched or depleted in DHX36 IP upon LENT depletion.

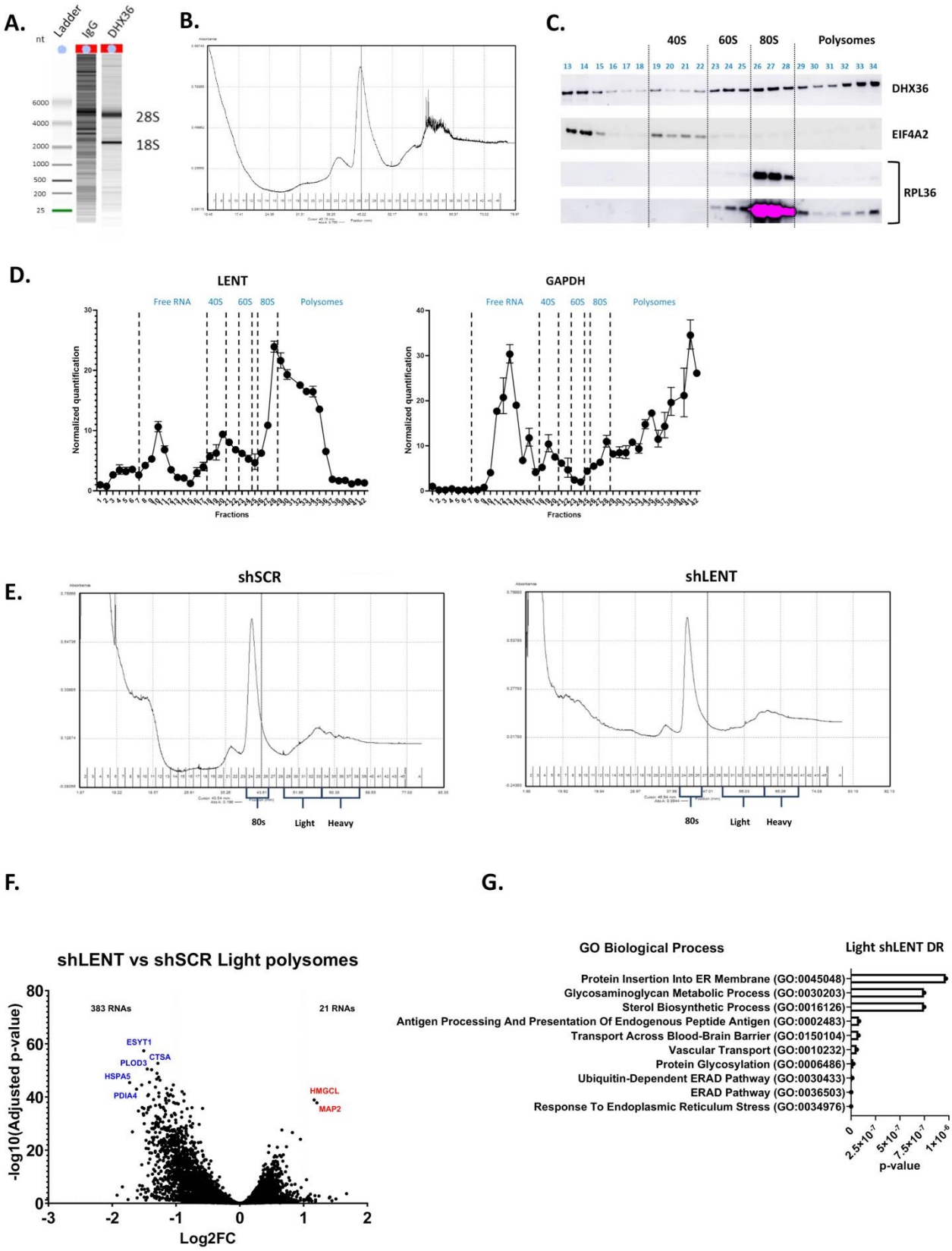

A.
Ladder
IgG
DHX36
nt
6000
4000
2000
1000
500
200
25
28S
18S
B.
C.
40S
60S
80S
Polysomes
13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34
DHX36
EIF4A2
RPL36
D.
LENT
GAPDH
Free RNA
40S
60S
80S
Polysomes
Normalized quantification
Fractions
E.
shSCR
shLENT
80s
Light
Heavy
F.
shLENT vs shSCR Light polysomes
383 RNAs
21 RNAs
ESYT1
CTSA
PLOD3
HSPA5
PDIA4
HMGCL
MAP2
-log10(Adjusted p-value)
Log2FC
G.
GO Biological Process
Light shLENT DR
Protein Insertion Into ER Membrane (GO:0045048)
Glycosaminoglycan Metabolic Process (GO:0030203)
Sterol Biosynthetic Process (GO:0016126)
Antigen Processing And Presentation Of Endogenous Peptide Antigen (GO:0002483)
Transport Across Blood-Brain Barrier (GO:0150104)
Vascular Transport (GO:0010232)
Protein Glycosylation (GO:0006486)
Ubiquitin-Dependent ERAD Pathway (GO:0030433)
ERAD Pathway (GO:0036503)
Response To Endoplasmic Reticulum Stress (GO:0034976)
p-value

Figure 4

Figure 4: LENT regulates polysome engagement of a subset of mRNAs. A. Bio-analyser analyses of precipitated RNA show strong enrichment of 18S and 28S rRNAs in the DHX36 IP but not the control IgG IP. **B.** Sucrose gradient separation of ribosomes with rRNAs measured by UV-light spectrometry. **C.** Immunoblots for DHX36, the initiation factor EIF4A2 and the large ribosomal subunit RPL36 after polysome separation. EIF4A2 is enriched in the 40S and RPL36 in the 80S and heavier polysome fractions. **D.** Presence of LENT or GAPDH mRNA in fractionated ribosomes as measured by RT-qPCR. **E.** Representative sucrose gradient separation of ribosomes with rRNAs measured by UV-light spectrometry. Fractions pooled for the 80S, LP and HP are indicated. **F**. Volcano blot showing RNAs enriched or depleted in the LP fraction from control or shLENT cells. **G.** Gene ontology analysis by EnrichR software of the 383 depleted RNAs in the shLENT LP fraction.

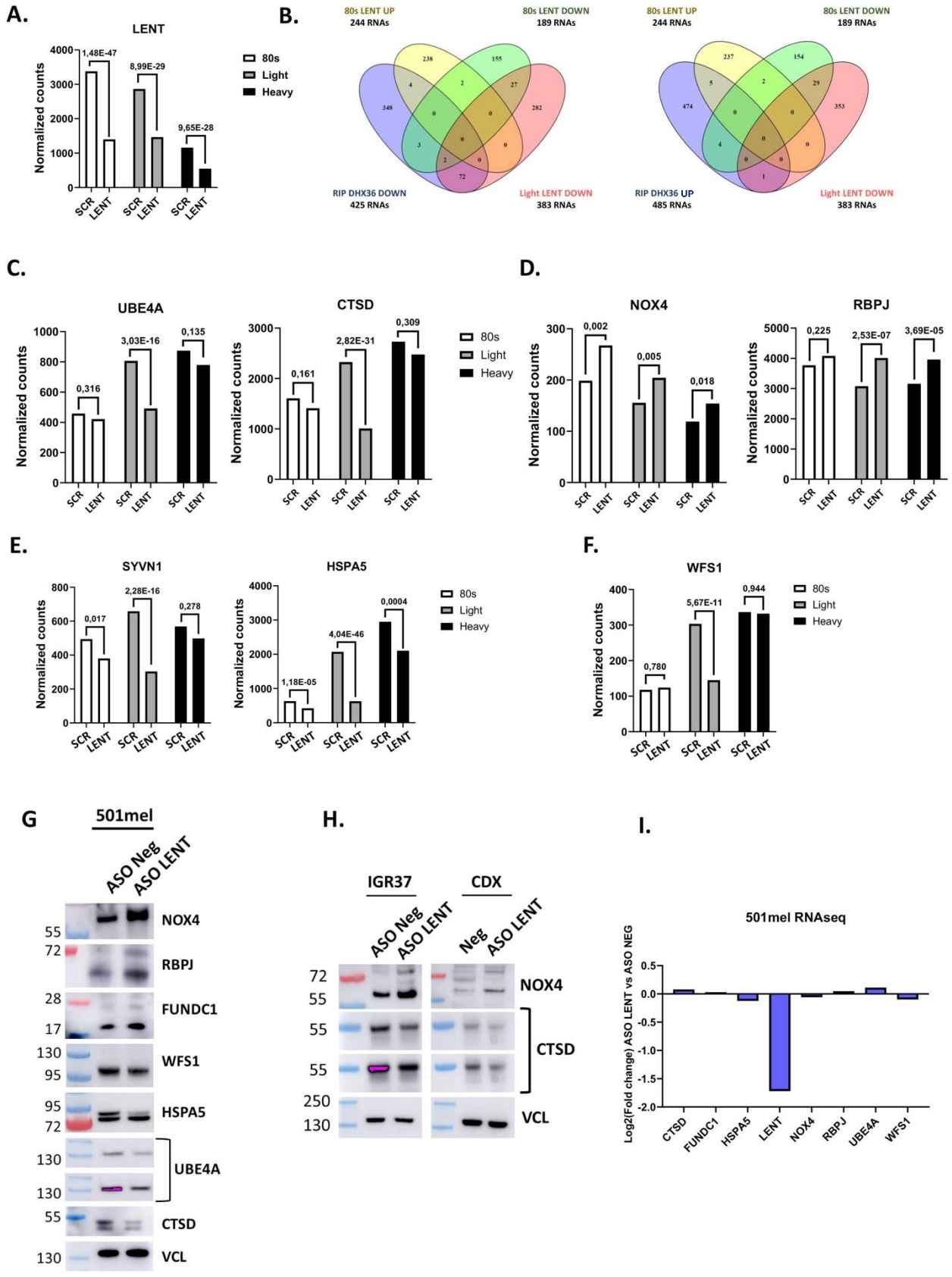

A.
LENT
Normalized counts
1,48E-47
8,99E-29
9,65E-28
SCR
LENT
80s
Light
Heavy
B.
80s LENT UP
244 RNAs
80s LENT DOWN
189 RNAs
RIP DHX36 DOWN
425 RNAs
Light LENT DOWN
383 RNAs
RIP DHX36 UP
485 RNAs
C.
UBE4A
CTSD
0,316
3,03E-16
0,135
0,161
2,82E-31
0,309
D.
NOX4
RBPJ
0,002
0,005
0,018
0,225
2,53E-07
3,69E-05
E.
SYVN1
HSPA5
0,017
2,28E-16
0,278
1,18E-05
4,04E-46
0,0004
F.
WFS1
0,780
5,67E-11
0,944
G
501mel
ASO Neg
ASO LENT
NOX4
RBPJ
FUNDC1
WFS1
HSPA5
UBE4A
CTSD
VCL
H.
IGR37
CDX
Neg
I.
501mel RNAseq
Log2(Fold change) ASO LENT vs ASO NEG
CTSD
FUNDC1
HSPA5
LENT
NOX4
RBPJ
UBE4A
WFS1

Figure 5

Figure 5: LENT fine-tunes translation of mRNAs differentially associated with polysomes. **A. A**. RNA-seq data showing the representation of LENT in the 80S, LP and HP fractions. The normalized number of reads are shown along with the adjusted p-value between the control and shLENT conditions. **B**. Venn diagrams comparing RNAs modulated in the DHX39 RIP in presence or absence of LENT silencing with those differentially present in the 80S and LP fractions. **C-F**. RNA-seq data showing the representation of the indicated RNAs in the 80S, LP and HP fractions. The normalized number of reads are shown along with the adjusted p-value between the control and shLENT conditions. **G.** Immunoblots of the indicated proteins 48 hours following ASO-mediated LENT silencing in 501Mel cells. **H.** Immunoblots of the indicated proteins 48 hours following ASO-mediated LENT silencing in the IGR37 cell line or extracts from IGR37 CDX tumours from control or LENT ASO-injected mice. **I.** Expression of the indicated RNAs in ASO control or ASO LENT transfected cells represented as fold-change.

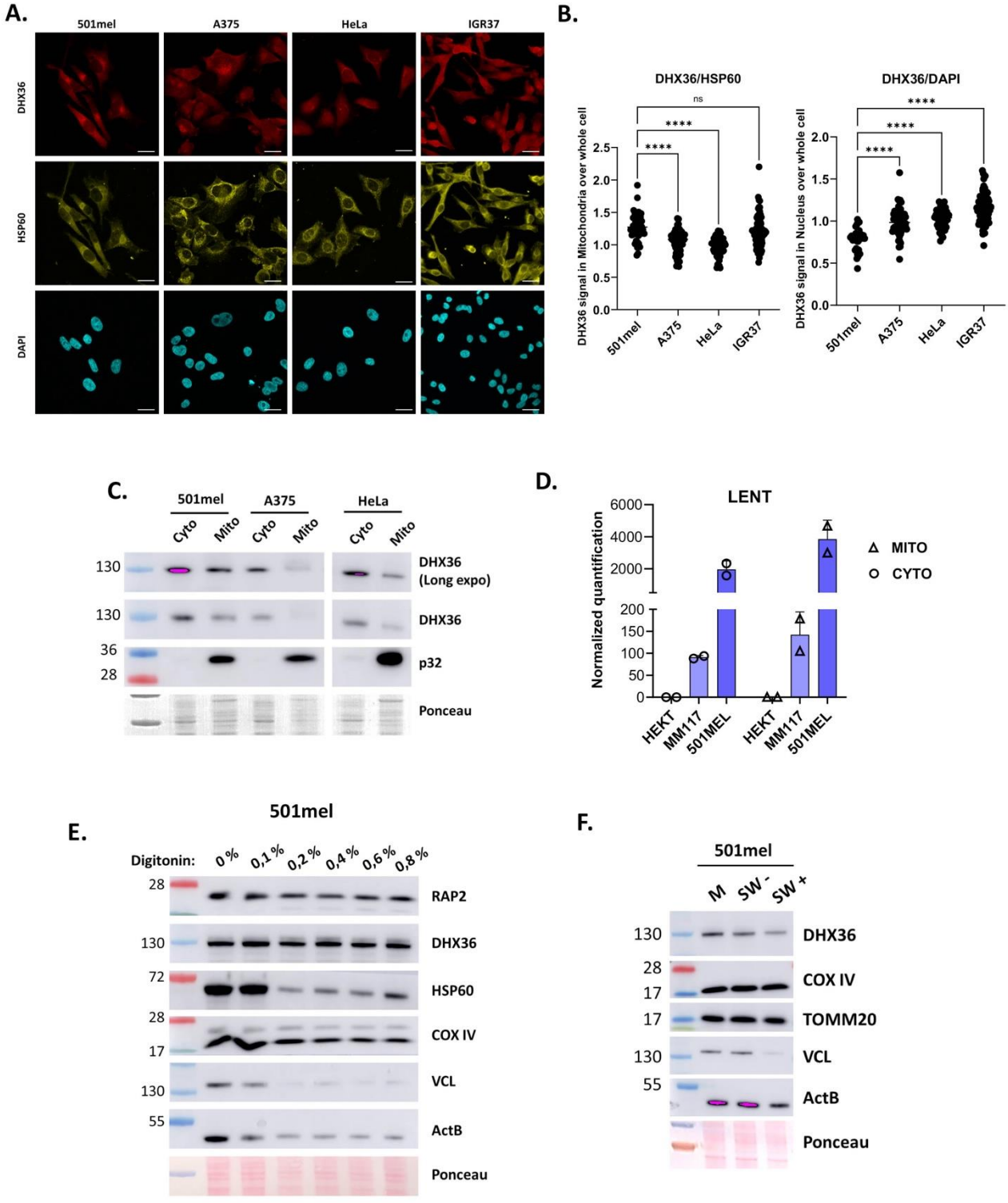
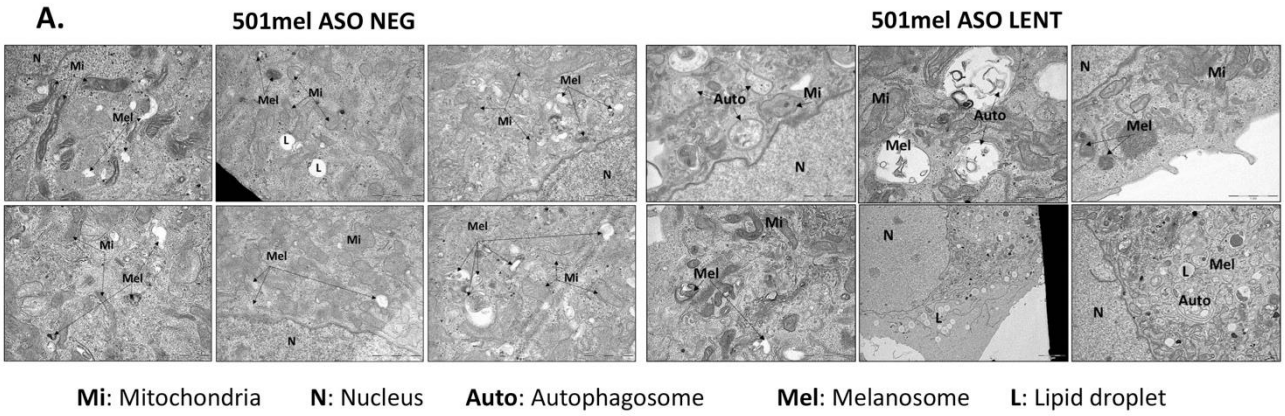
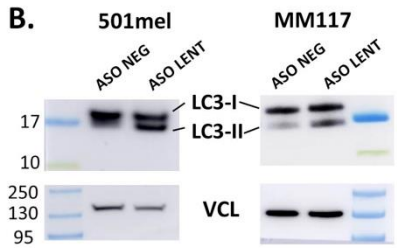
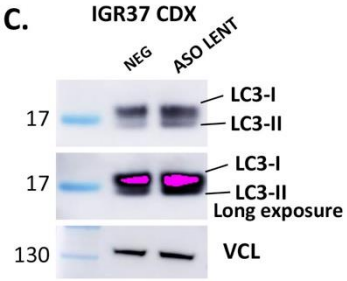

A.
501mel
A375
HeLa
IGR37
DHX36
HSP60
DAPI
B.
DHX36/HSP60
DHX36/DAPI
DHX36 signal in Mitochondria over whole cell
DHX36 signal in Nucleus over whole cell
ns
****
501mel
A375
HeLa
IGR37
C.
501mel
A375
HeLa
Cyto
Mito
130
DHX36 (Long expo)
DHX36
36
28
p32
Ponceau
D.
LENT
Normalized quantification
MITO
CYTO
HEKT
MM117
501MEL
E.
501mel
Digitonin:
0 %
0,1 %
0,2 %
0,4 %
0,6 %
0,8 %
28
RAP2
130
DHX36
72
HSP60
28
17
COX IV
130
VCL
55
ActB
Ponceau
F.
501mel
M
SW -
SW +
130
DHX36
28
17
COX IV
17
TOMM20
130
VCL
55
ActB
Ponceau

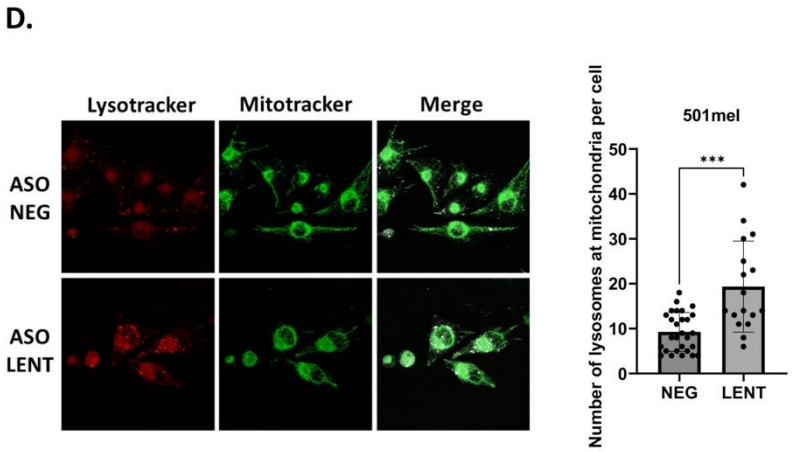
Figure 6.

Figure 6: LENT and DHX36 are mainly localized at mitochondria in melanoma cells. A. Immunofluorescence by confocal microscopy showing DHX36 localization in different cell lines. HSP60 is used as a mitochondria marker and DAPI to stain the nucleus. Scale bars = 10 µM. **B.** DHX36 signal was quantified in the whole cell and divided by the signal in mitochondria or in the nucleus. Each measured cell is represented by one point and groups compared by one-way ANOVA (Dunnett test). *, $P < 0.033$; **, $P < 0.0021$; ***, $P < 0.0002$; ****, $P < 0.0001$. **C-D.** DHX36 and LENT levels quantified by immunoblot or RT-qPCR in cytosolic or mitochondrial fractions of different cell lines. **E.** Purified mitochondria were digested with increasing concentrations of digitonin and retained proteins were analyzed by western blot. **F.** Purified mitochondria were digested with trypsin with or without cell swelling buffer and remaining proteins were analyzed by western blot.


A.
501mel ASO NEG
501mel ASO LENT
Mi: Mitochondria
N: Nucleus
Auto: Autophagosome
Mel: Melanosome
L: Lipid droplet


B.
501mel
MM117
ASO NEG
ASO LENT
LC3-I
LC3-II
VCL


C.
IGR37 CDX
NEG
ASO LENT
LC3-I
LC3-II
Long exposure
VCL


D.
Lysotracker
Mitotracker
Merge
ASO NEG
ASO LENT
501mel
Number of lysosomes at mitochondria per cell
NEG
LENT

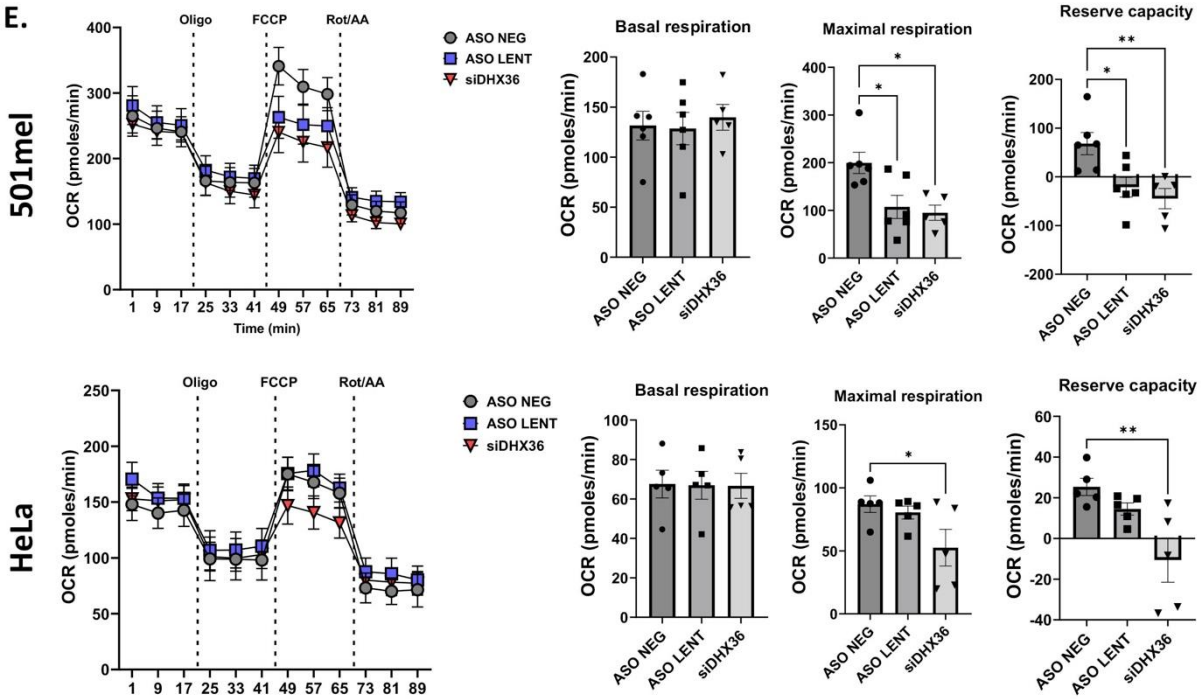

E.
501mel
HeLa
Oligo
FCCP
Rot/AA
ASO NEG
ASO LENT
siDHX36
OCR (pmoles/min)
Time (min)
Basal respiration
Maximal respiration
Reserve capacity

Figure 7.

Figure 7: LENT depletion induces mitophagy and reduces oxygen consumption rate. A. Transmission electron microscopy of 501Mel cells 48 hours following transfection of control or LENT-targeting ASO. Autophagosomes and degraded melanosomes were observed in LENT silenced cells. Scale bars are indicated on the images. **B.** Quantification of cytosolic LC3B (LC3-I) and lipid-associated LC3B (LC3 II) by western blot in 501Mel or MM117 cells. Vinculin is used as a loading control. **C.** LC3B immunoblot in extracts of IGR37 CDX tumours. **D.** Confocal microscopy of unfixed 501Mel cells stained with lysotracker and mitotracker in control or LENT silenced conditions. The numbers of co-localizing lysosomes and mitochondria was determined and compared between the two conditions by Welch's test. Scale bars = 10 µM. **E.** Oxygen consumption rate was determined upon LENT or DHX36 silencing. Reserve capacity was obtained by subtracting the maximal capacity with the basal capacity. Comparisons were done by one-way ANOVA (Dunnett test). *, $P < 0.033$; **, $P < 0.0021$; ***, $P < 0.0002$; ****, $P < 0.0001$.

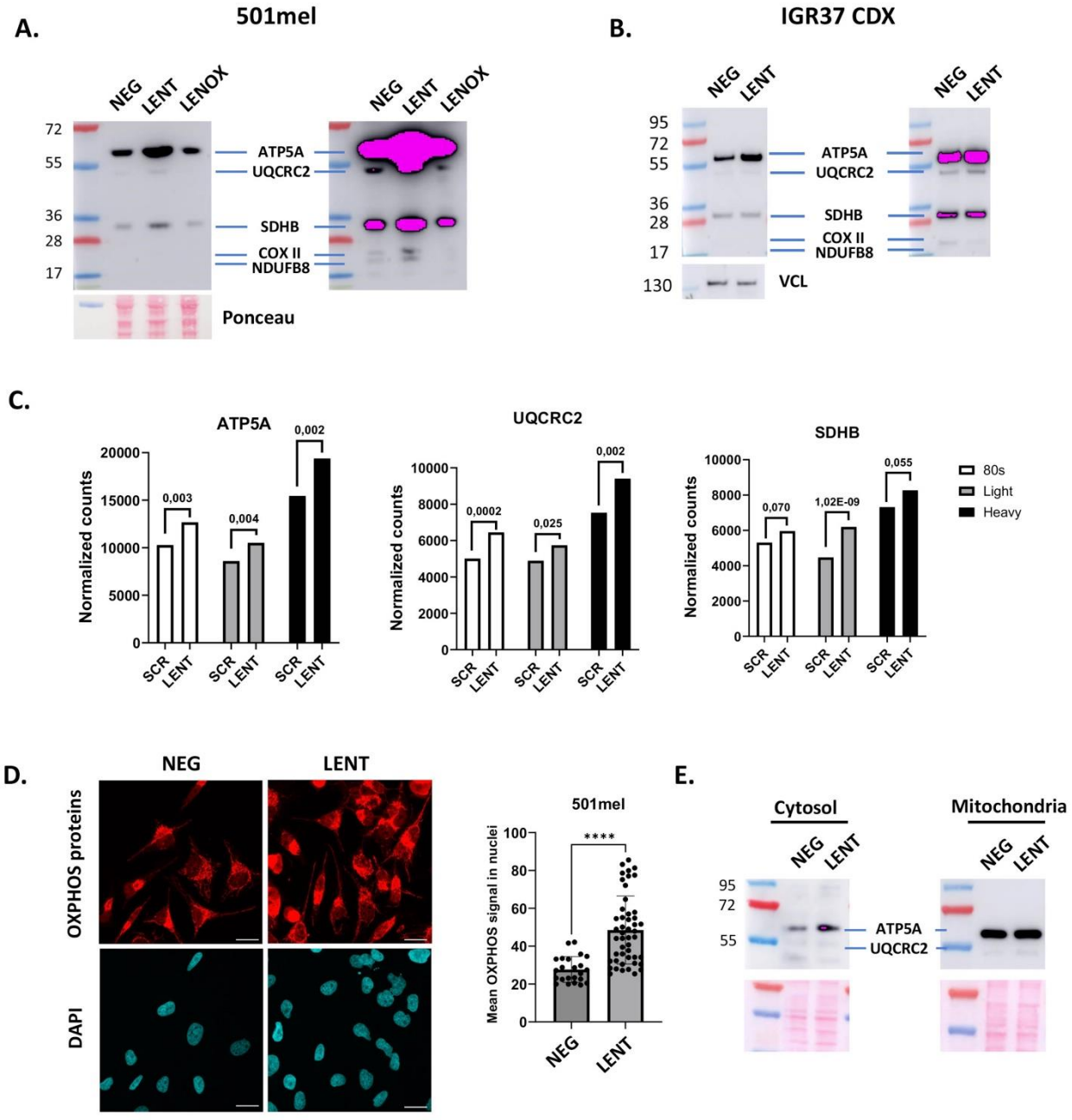

A.
501mel
NEG
LENT
LENOX
72
55
36
28
17
ATP5A
UQCRC2
SDHB
COX II
NDUFB8
Ponceau
B.
IGR37 CDX
95
72
55
36
28
17
130
VCL
C.
ATP5A
UQCRC2
SDHB
Normalized counts
0,003
0,004
0,002
0,0002
0,025
0,002
0,070
1,02E-09
0,055
SCR
LENT
80s
Light
Heavy
D.
NEG
LENT
OXPHOS proteins
DAPI
501mel
Mean OXPHOS signal in nuclei
****
E.
Cytosol
Mitochondria
ATP5A
UQCRC2

Figure 8.

Figure 8: LENT depletion triggers DNA damage and accumulation of OxPhos proteins. **A.** Immunoblots detecting the indicated proteins in control or LENT depleted cells. Panels show different exposures of the same immuoblot. **B.** Immunoblots detecting mitochondrial electron transport chain proteins in extracts from IGR-37 CDX tumours. **C.** RNA-seq data showing the representation of the indicated RNAs in the 80S, LP and HP fractions. The normalized number of reads are shown along with the adjusted p-value between the control and shLENT conditions. **D.** OxPhos proteins were detected by confocal microscopy immunofluorescence in control or LENT-silenced cells. Proteins were quantified in the nucleus using the DAPI signal and comparisons were made by Mann-Whitney test. Scale bars = 10 μM. **E.** Immunoblot of ATP5A1 in cytosolic or mitochondrial fractions upon LENT depletion. Ponceau is shown as loading control. *, $P < 0.033$; **, $P < 0.0021$; ***, $P < 0.0002$; ****, $P < 0.0001$.

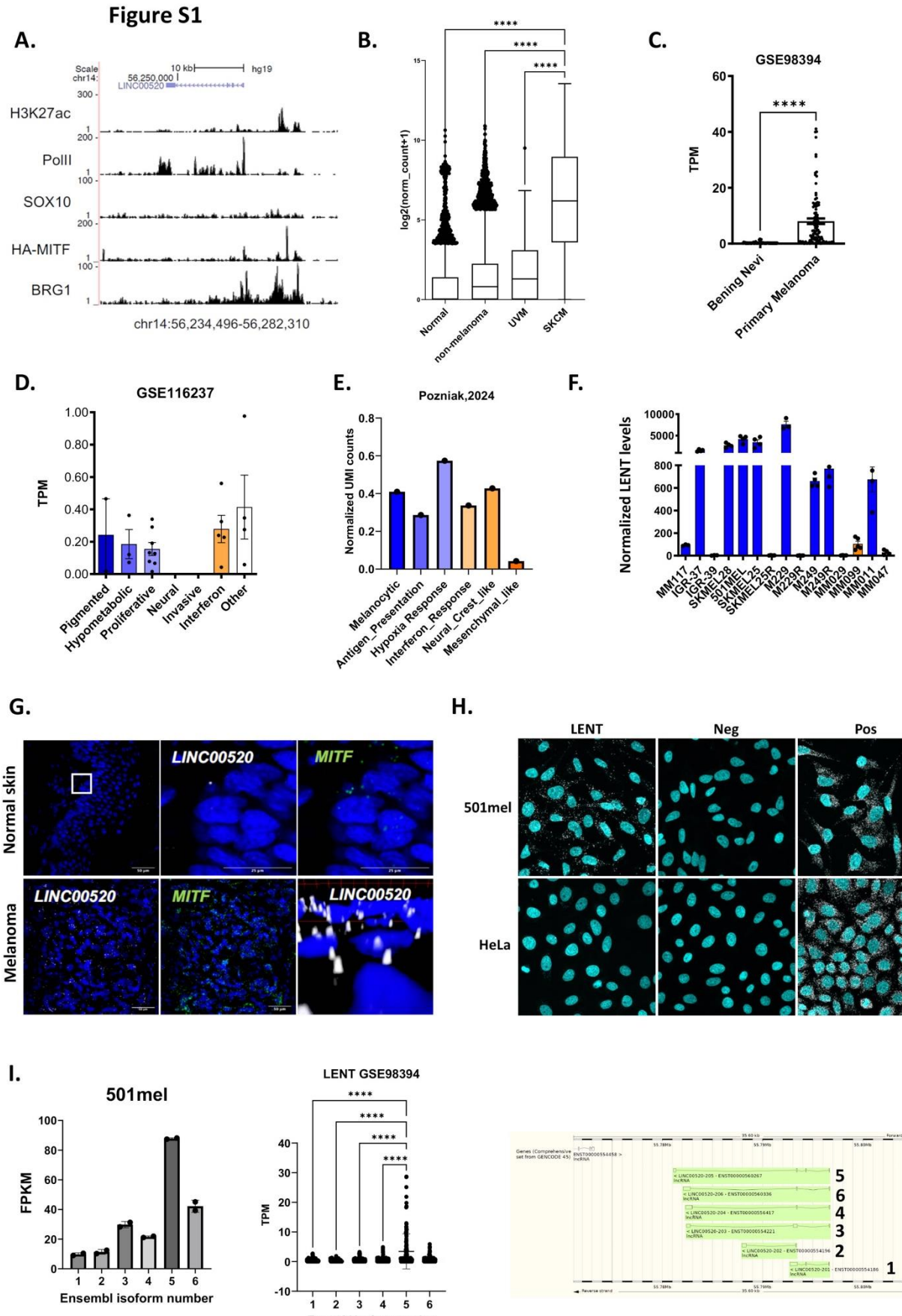


Figure S1
A.
B.
C.
H3K27ac
PolII
SOX10
HA-MITF
BRG1
chr14:56,234,496-56,282,310
log2(norm_count+1)
Normal
non-melanoma
UVM
SKCM
GSE98394
TPM
Bening Nevi
Primary Melanoma
D.
GSE116237
TPM
Pigmented
Hypometabolic
Proliferative
Neural
Invasive
Interferon
Other
E.
Pozniak,2024
Normalized UMI counts
Melanocytic
Antigen_Presentation
Hypoxia Response
Interferon_Response
Neural_Crest_like
Mesenchymal_like
F.
Normalized LENT levels
MM117
IGR-37
IGR-39
SKMEL28
501MEL
SKMEL25
SKMEL25R
M229
M229R
M249
M249R
MM029
MM099
MM011
MM047
G.
Normal skin
Melanoma
LINC00520
MITF
H.
LENT
Neg
Pos
501mel
HeLa
I.
501mel
FPKM
Ensembl isoform number
LENT GSE98394
TPM
Ensembl isoform number

Figure S1: LENT expression is the highest in melanocytic cells of skin cutaneous melanoma. A. UCSC screenshot of the indicated ChIP-seq tracks from Laurette et al (9) at the LENT locus illustrating binding of MITF and BRG1. **B.** LENT expression in tumors and normal tissues were retrieved from the GTEX and TCGA databases and compared by one-way ANOVA (Dunnett test). **C.** RNA-seq data from GSE98394 shows LENT levels in benign nevi compared to melanoma by Mann-Whitney test. **D.** scRNA-seq data from GSE116237 shows LENT expression in the different cell populations. Melanocytic subtypes are shown in blue and mesenchymal subtypes in orange. **E.** scRNA-seq data from EGAD00001009291 from Pozniak et al (7) shows LENT expression in the different cell populations. **F.** LENT levels were determined by RT-qPCR in a collection of melanoma cell lines. Melanocytic are in blue and mesenchymal in orange. **G.** RNAscope coupled with confocal microscopy to detect LENT and MITF mRNA in sections from normal skin and melanoma. Nuclei are stained with DAPI. Scale bars are shown on each image. **H.** RNAscope coupled with confocal microscopy to detect LENT RNA in 501Mel or HeLa cells along with the negative and positive control reactions provided by the supplier. Nuclei are stained with DAPI. **I.** LENT isoform expression in melanoma cell lines and in RNA-seq from patients (GSE98394). Isoforms are represented with the Ensembl genome browser interface. Comparisons were made by one-way ANOVA (Dunn's test). *, $P < 0.033$; **, $P < 0.0021$; ***, $P < 0.0002$; ****, $P < 0.0001$.

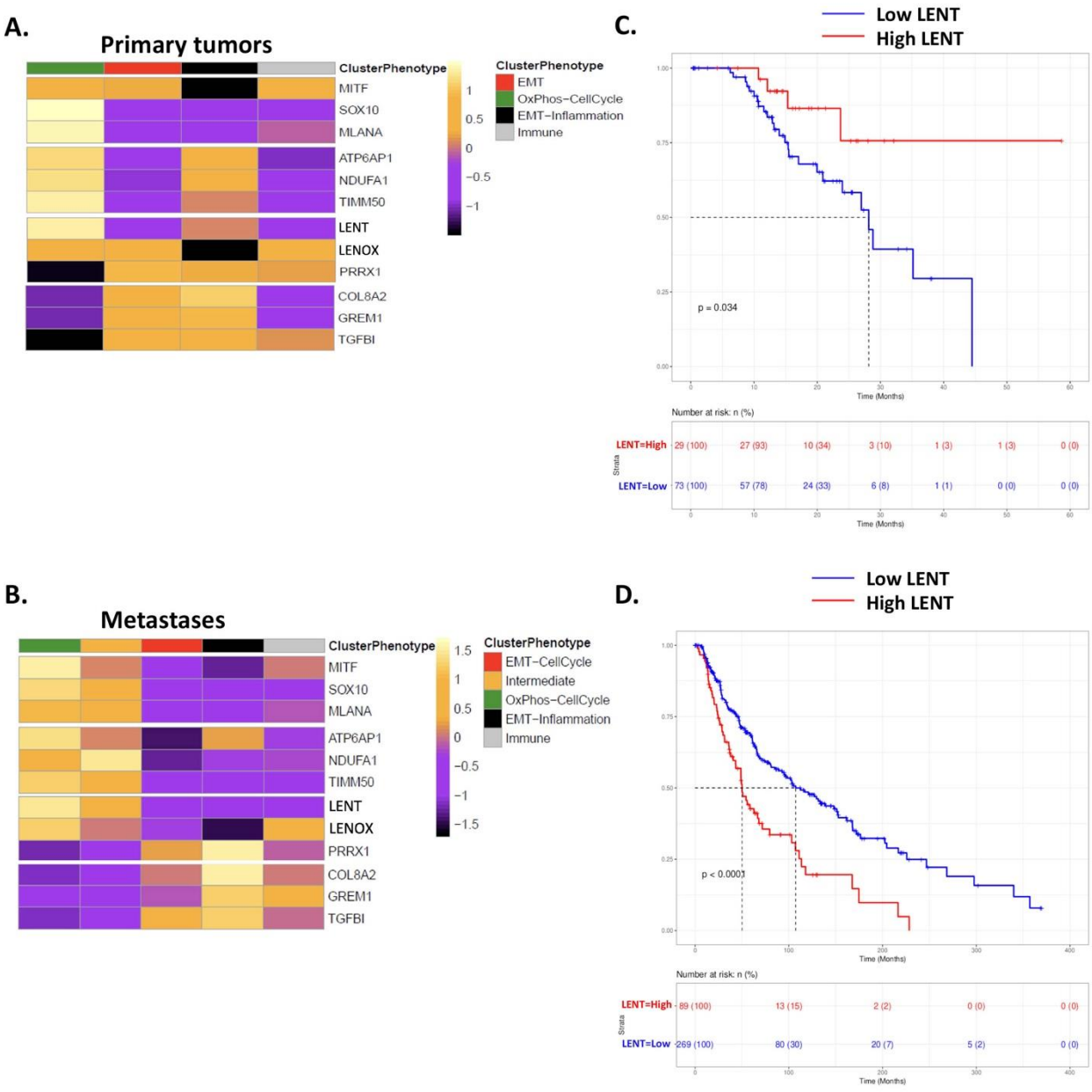

A.
Primary tumors
ClusterPhenotype
MITF
SOX10
MLANA
ATP6AP1
NDUFA1
TIMM50
LENT
LENOX
PRRX1
COL8A2
GREM1
TGFBI
1
0.5
0
−0.5
−1
ClusterPhenotype
EMT
OxPhos−CellCycle
EMT−Inflammation
Immune
C.
Low LENT
High LENT
p = 0.034
Time (Months)
Number at risk: n (%)
LENT=High 29 (100) 27 (93) 10 (34) 3 (10) 1 (3) 1 (3) 0 (0)
LENT=Low 73 (100) 57 (78) 24 (33) 6 (8) 1 (1) 0 (0) 0 (0)
Strata
B.
Metastases
ClusterPhenotype
MITF
SOX10
MLANA
ATP6AP1
NDUFA1
TIMM50
LENT
LENOX
PRRX1
COL8A2
GREM1
TGFBI
1.5
1
0.5
0
−0.5
−1
−1.5
ClusterPhenotype
EMT−CellCycle
Intermediate
OxPhos−CellCycle
EMT−Inflammation
Immune
D.
Low LENT
High LENT
p < 0.0001
Time (Months)
Number at risk: n (%)
LENT=High 89 (100) 13 (15) 2 (2) 0 (0) 0 (0)
LENT=Low 269 (100) 80 (30) 20 (7) 5 (2) 0 (0)
Strata

Figure S2

Figure S2: LENT expression is highest in OxPhos enriched cells and correlates with melanoma patient survival. A and C. Data from the SKCM TCGA were separated into primary tumor and metastases. Unsupervised clustering of the RNA-seq data from the two groups was performed and the identity of the cell populations derived from GSEA Hallmark analyses of the differentially expressed genes. Heatmaps show the expression of the indicated genes as a Z-scored heatmap in the identified cell populations. **B and D.** Kaplan-Meier curves for overall survival in patients according to LENT expression score using the optimal cut point method with the associated log-rank p-value from the univariate Cox proportional-hazard model. The number of patients in each group are indicated.

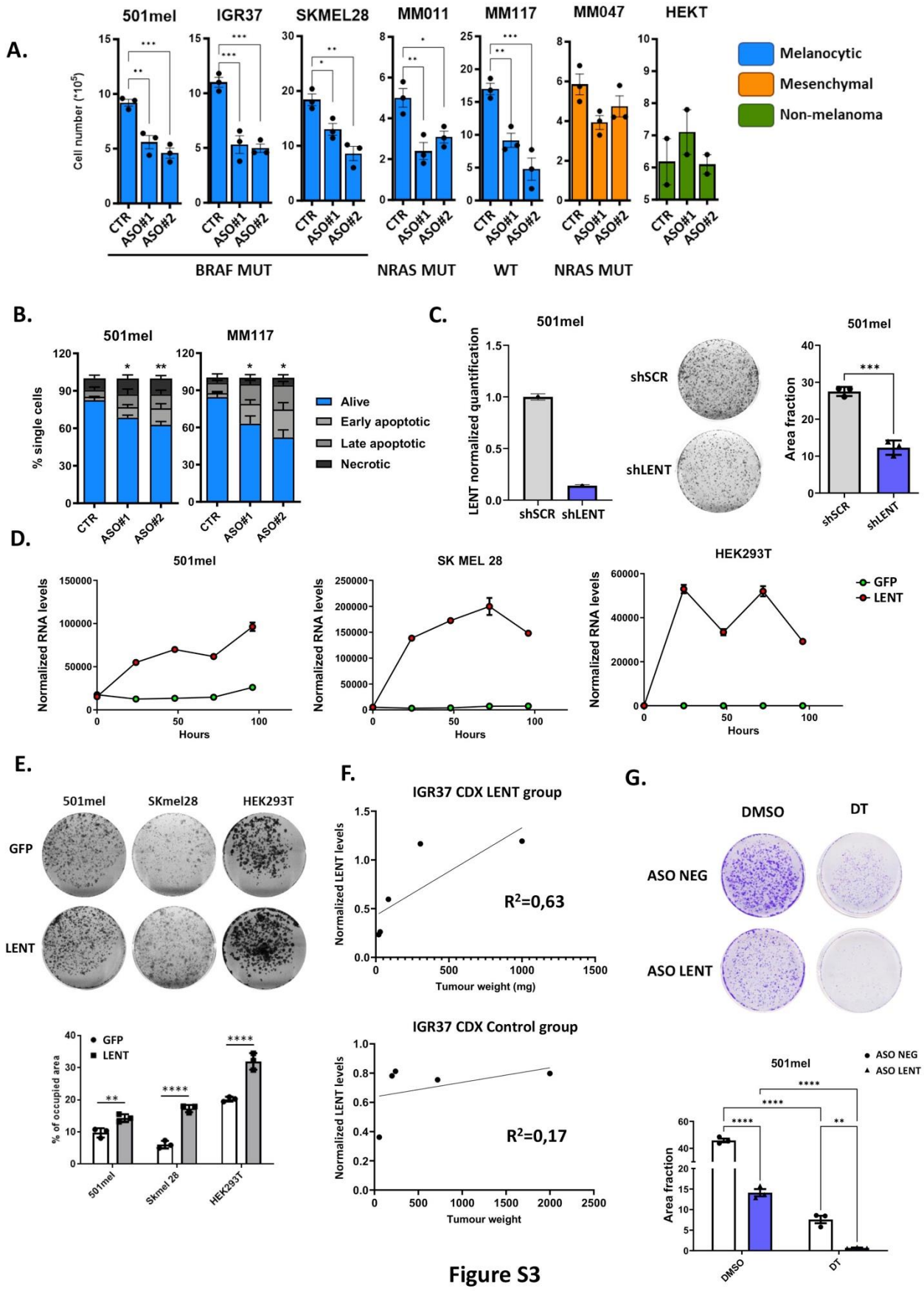

A.
501mel
IGR37
SKMEL28
MM011
MM117
MM047
HEKT
Cell number (*10^5)
CTR
ASO#1
ASO#2
BRAF MUT
NRAS MUT
WT
NRAS MUT
Melanocytic
Mesenchymal
Non-melanoma
B.
501mel
MM117
% single cells
Alive
Early apoptotic
Late apoptotic
Necrotic
C.
501mel
LENT normalized quantification
shSCR
shLENT
shSCR
shLENT
501mel
Area fraction
D.
501mel
SK MEL 28
HEK293T
Normalized RNA levels
Hours
GFP
LENT
E.
501mel
SKmel28
HEK293T
GFP
LENT
% of occupied area
F.
IGR37 CDX LENT group
Normalized LENT levels
R²=0,63
Tumour weight (mg)
IGR37 CDX Control group
R²=0,17
Tumour weight
G.
DMSO
DT
ASO NEG
ASO LENT
501mel
ASO NEG
ASO LENT
Area fraction

Figure S3

Figure S3: Gain and loss of LENT in melanoma cells. A. ASO-mediated depletion of LENT was performed in the indicated cell lines with two independent ASOs and cell numbers were counted after 48h and compared with a control condition by one-way ANOVA (Dunnett test). **B.** ASO-mediated depletion of LENT triggers apoptosis as shown by annexin V + PI staining coupled with flow cytometry and compared with a control condition by one-way ANOVA (Dunnett test). **C.** LENT was silenced by a Dox-inducible shRNA and LENT levels quantified by RT-qPCR. Comparison of colony formation following Dox-induced expression of control or LENT-targeting shRNA by unpaired T test. **D.** Ectopic expression of LENT in the indicated cell lines was measured by RT-qPCR using a cell line with GFP inducible as a control. **E.** Colony formation upon ectopic LENT expression was performed in the indicated cell lines and compared the GFP control by one-way ANOVA (Dunnett test). **F.** LENT levels in the IGR-37 CDX were measured by RT-qPCR and association with tumour weight was performed by linear regression. **G.** Cell proliferation was assessed after ASO-mediated LENT silencing in combination with the indicated drugs. Comparisons were done by two-way ANOVA. *, $P < 0.033$; **, $P < 0.0021$; ***, $P < 0.0002$; ****, $P < 0.0001$.

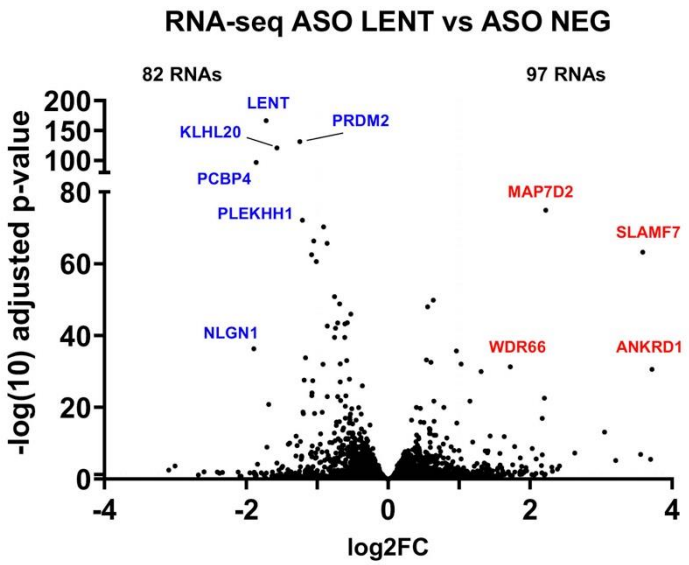


Figure S4

Figure S4: LENT silencing does not regulate gene expression. Volcano blot showing RNAs enriched or depleted in 501Mel cells transfected with control or LENT targeting ASO. P-values were derived using the Wald test.

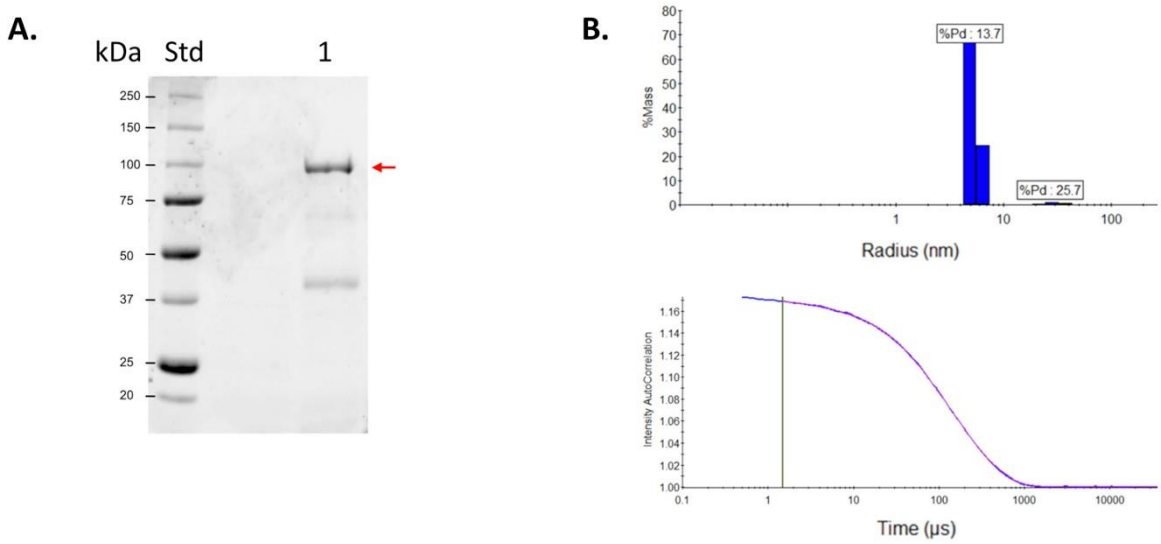
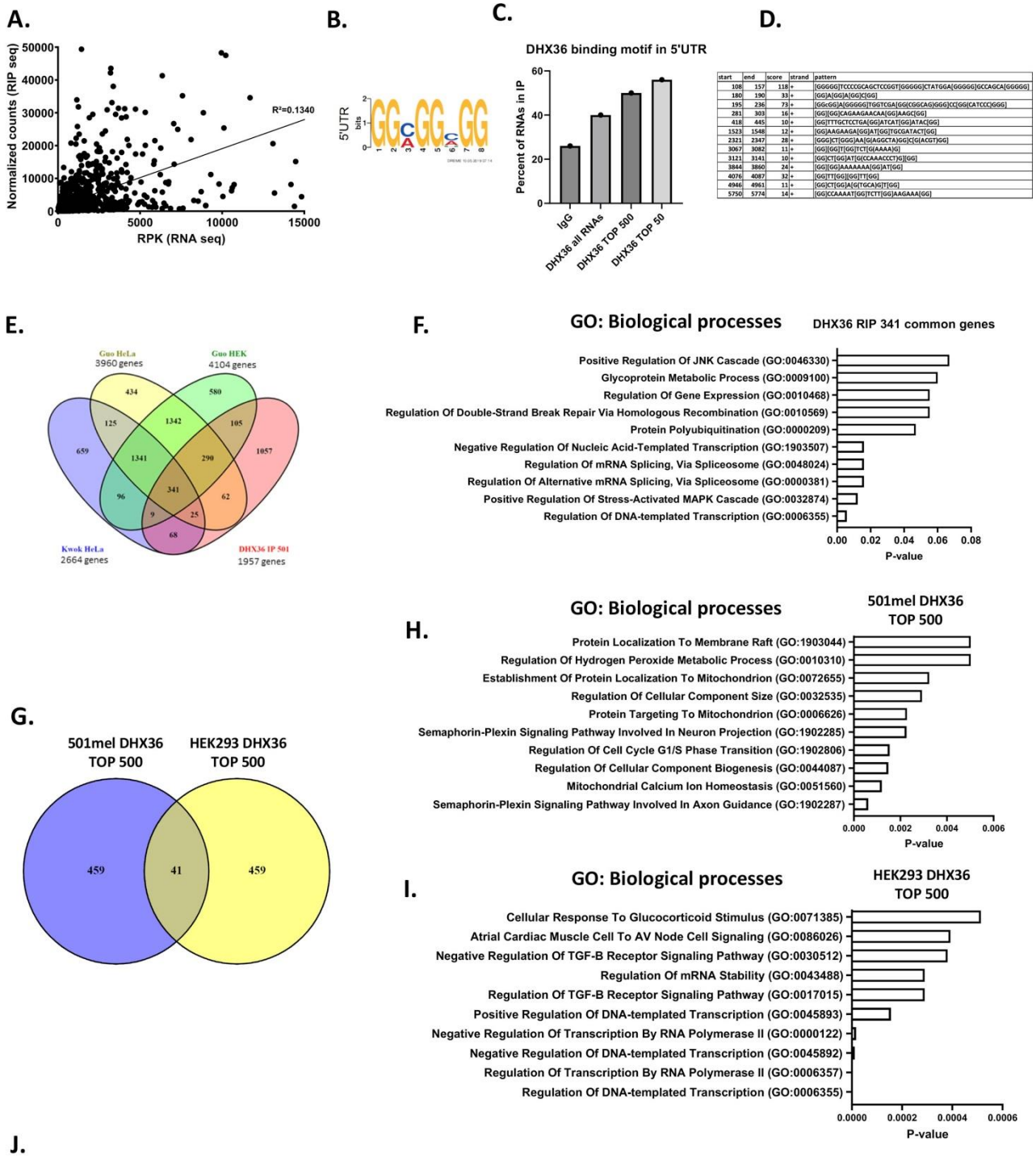


Figure S5

Figure S5: DHX36 purification and LENT sequence prediction. A. Comassie blue staining of an SDS-PAGE of purified DHX36 protein indicated by the red arrow. **B.** Upper panel, dynamic light scattering analysis of protein size distribution with Polydispersity index (%Pd) for purified DHX36; lower panel the autocorrelation function.



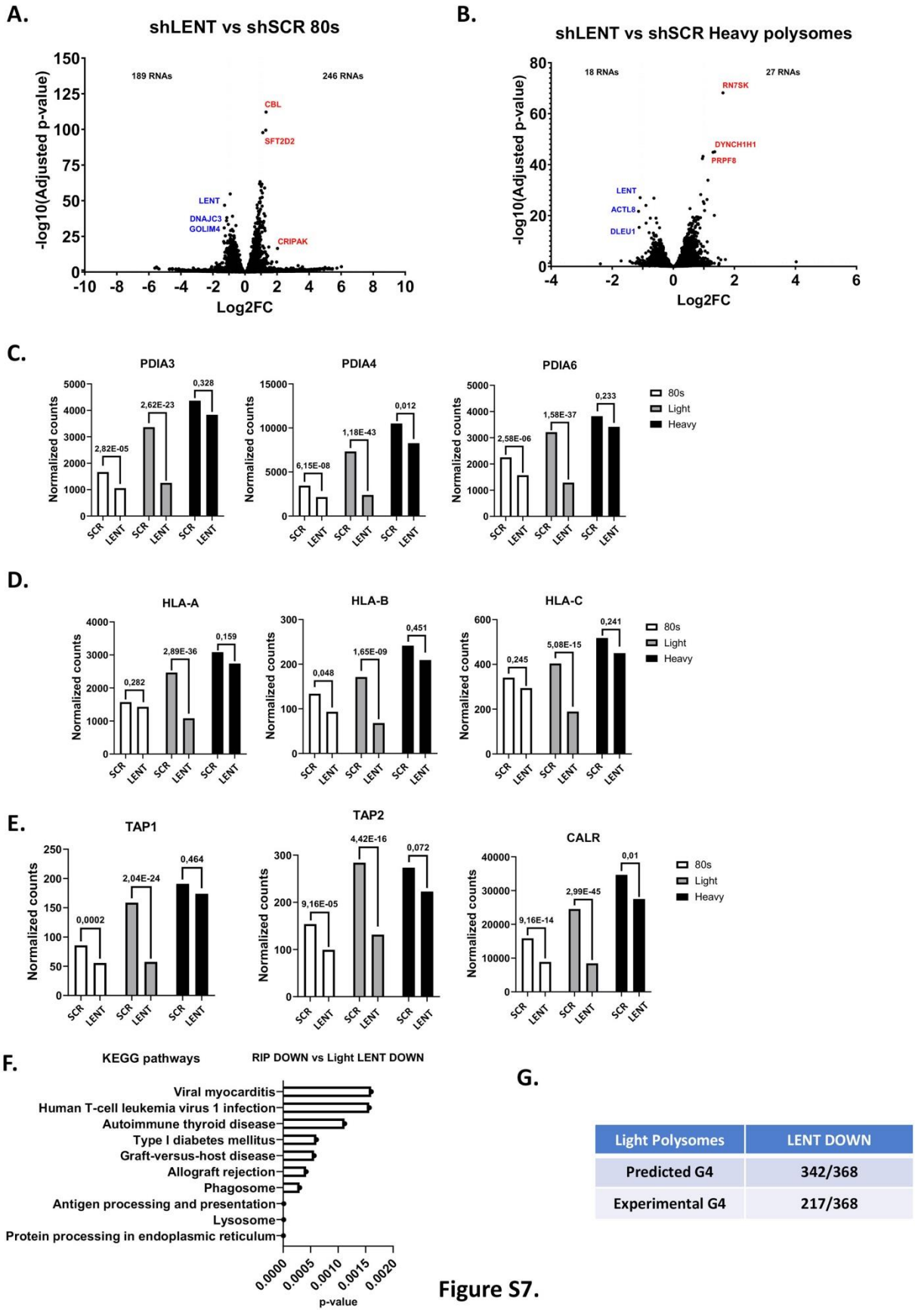
J.

DHX36 RIP	LENT DOWN TOP 100	LENT UP TOP 100
Predicted G4	88/91	76/84
Experimental G4	53/91	43/84
P-value	0	0

DHX36 RIP	DHX36 TOP 100	IgG TOP 100
Predicted G4	83/91	6/54
Experimental G4	71/91	3/54
P-value	0	0,2

Figure S6

Figure S6. Presence of potential G4-forming structures in DHX36 associated RNAs. A. Correlation between RNA expression level and enrichment in the DHX36 RIP calculated by linear regression. **B-C.** The % of RNAs harboring the DHX36-interaction motif shown in B in the 5'UTR of the indicated sets of RNAs was determined by the Fimo algorithm. **D.** G4 prediction by PQSfinder in the DHX36 mRNA sequence. Guanines implicated in potential G4 structures are represented in parenthesis. **E.** Venn diagram representing the overlap between RNAs enriched in the DHX36 RIP dataset reported here and the indicated data sets of potential G4-containing RNAs. 341 genes are shared between the all four datasets. **F.** Gene ontology analysis by EnrichR of the 341 common genes. **G.** Venn diagram showing the overlap between the 500 most enriched RNAs in the 501Mel DHX36 RIP seq compared to the HeLa DHX36 RIP dataset (28). **H-I.** Ontology analysis of the mRNAs specific to the 501Mel or HeLa datasets **J.** Analysis by QUADRatlas of the top 100 RNAs enriched in DHX36 RIP, control IgG RIP, or RNA whose association with DHX36 was modulated by LENT silencing. Not all genes were mapped by the algorithm explaining why the number of predicted or experimental G4s is not represented as a fraction of 100.



Light Polysomes	LENT DOWN
Predicted G4	342/368
Experimental G4	217/368

Figure S7.

Figure S7. Effect on LENT silencing on ribosome engagement of RNAs. A-B. Volcano plots showing RNAs enriched or depleted in the HP and 80S fractions from the control or shLENT cells. **C-E**. RNA-seq data showing the representation of the indicated RNAs in the 80S, LP and HP fractions. The normalized number of reads are shown along with the adjusted p-value between the control and shLENT conditions. **F**. KEGG ontology analyses of the 74 RNAs commonly regulated by LENT in DHX36 IP and polysome profiling. **G.** QUADRatlas of the 383 RNAs depleted in the LP fractions.

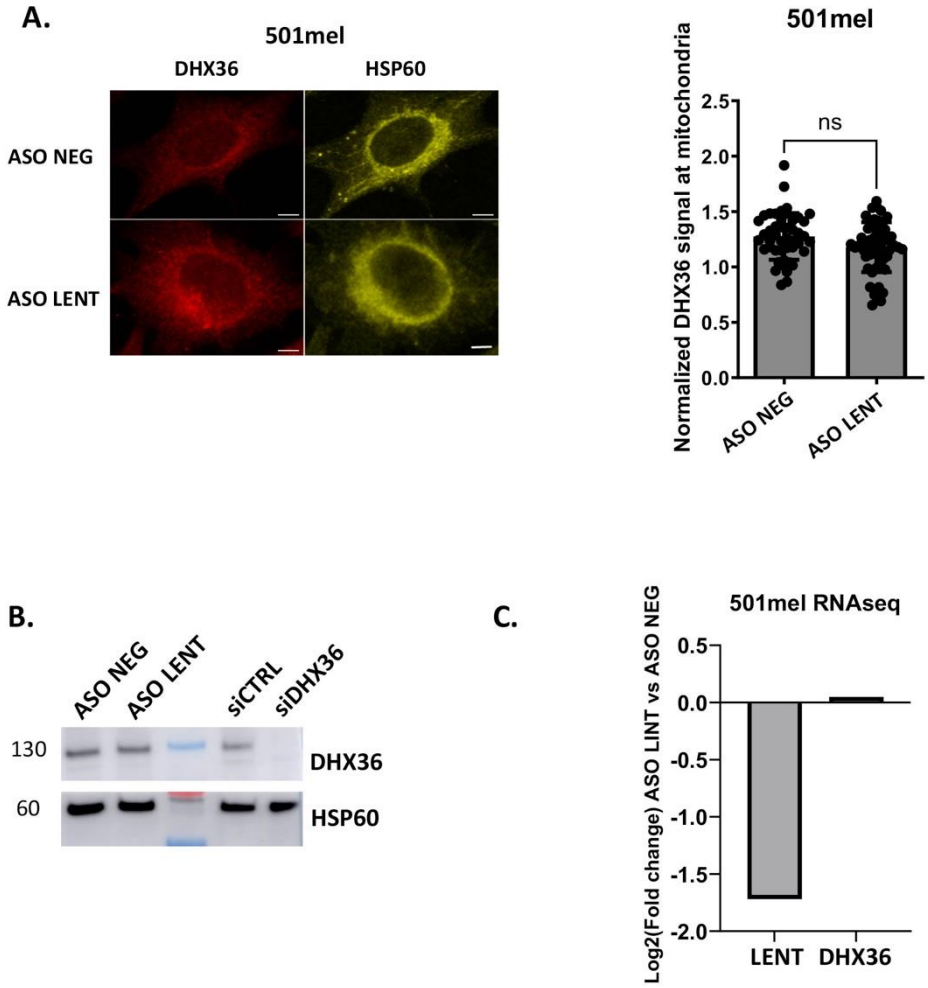


Figure S8

Figure S8: Effects of LENT depletion on DHX36 localization or protein level. A. Immunofluorescence by confocal microscopy showing DHX36 localization in control or LENT depleted cells. HSP60 is shown as mitochondria marker. Quantification was made by measurement of the total DHX36 signal over the DHX36 signal in mitochondria and compared to control by Mann-Whitney test. Scale bars = 2 µM. **B.** Immunoblot showing DHX36 levels upon depletion of LENT or DHX36. HSP60 used as a loading control. **C.** LENT and DHX36 RNA levels shown by RNA-seq upon LENT depletion.

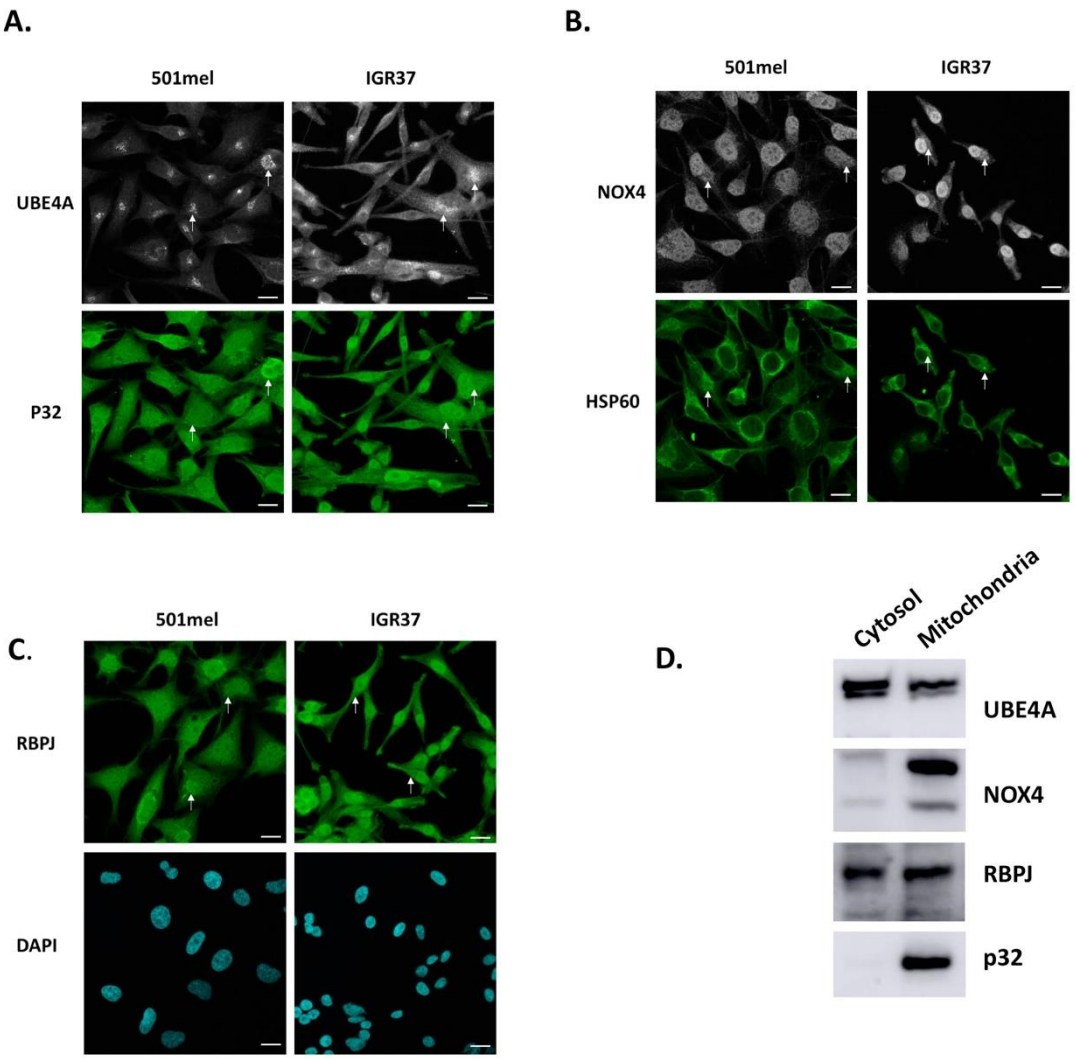


Figure S9

Figure S9. Mitochondrial localization of proteins encoded by several mRNAs differentially bound by DHX36 upon LENT depletion. A. RNA-seq shows quantification of mRNAs differentially bound by DHX36 upon LENT silencing. **B-D.** Immunofluorescence and confocal microscopy of the indicated proteins whose mRNAs were differentially bound by DHX36 upon LENT silencing. Scale bars = 10 μM. **E.** Immunoblot after separation of mitochondria from cytosol.

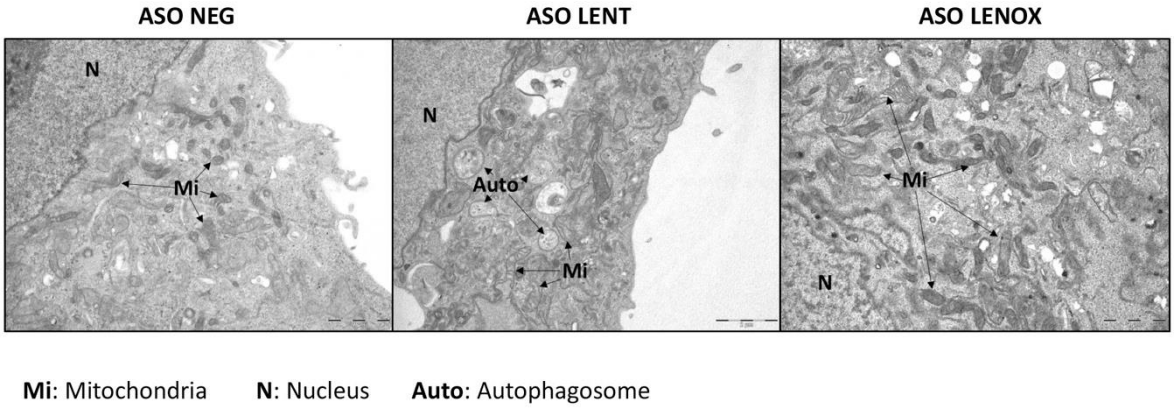


Figure S10

Figure S10. Autophagy/mitophagy is specific to LENT silenced cells. Transmission electron microscopy of 501Mel cells 48 hours following transfection of control, LENT or LENOX-targeting ASO. Scale bars are indicated on the images.

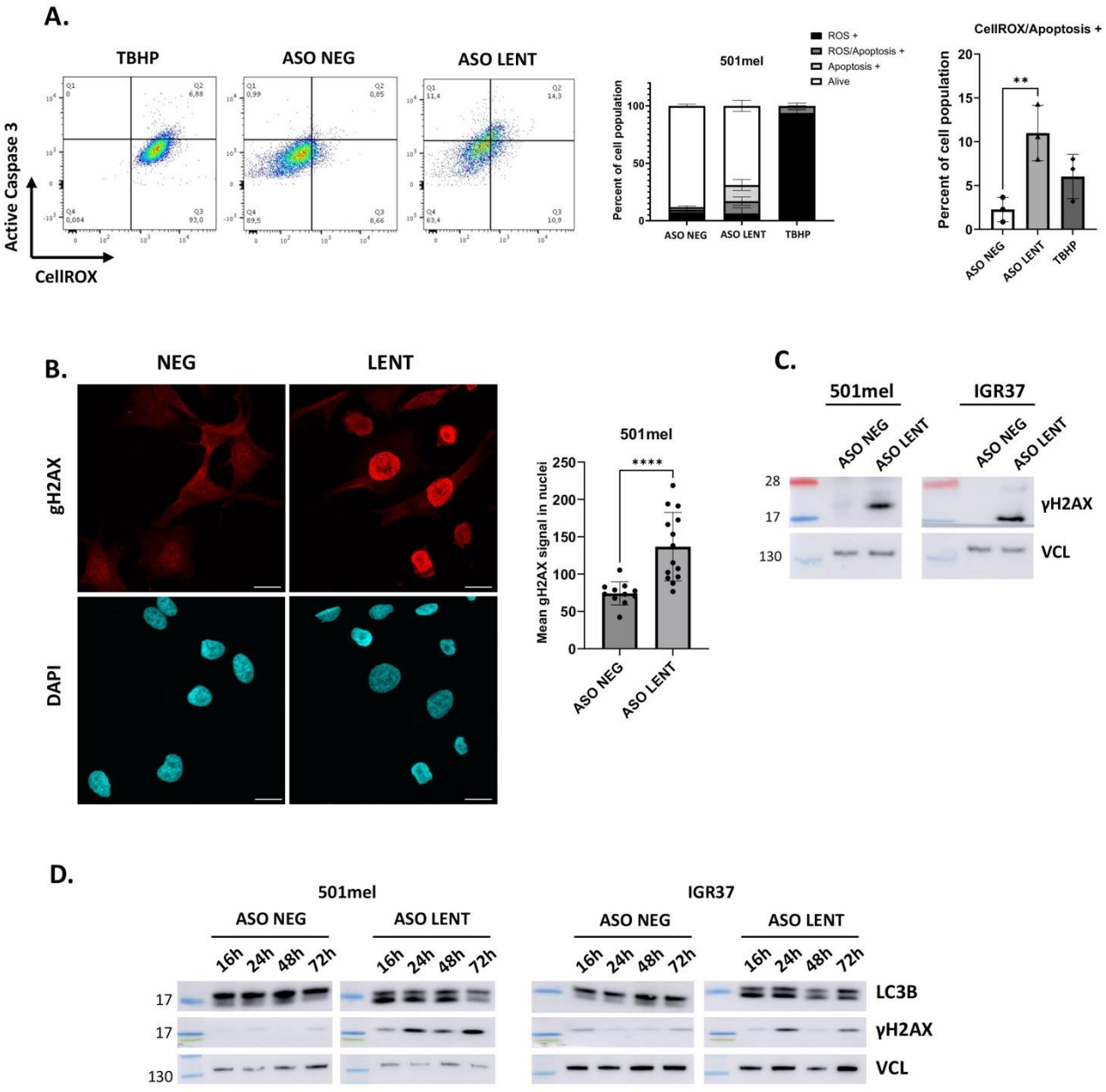

A.
TBHP
ASO NEG
ASO LENT
Active Caspase 3
CellROX
501mel
ROS +
ROS/Apoptosis +
Apoptosis +
Alive
Percent of cell population
ASO NEG
ASO LENT
TBHP
CellROX/Apoptosis +
**
B.
NEG
LENT
gH2AX
DAPI
501mel
****
Mean gH2AX signal in nuclei
C.
501mel
IGR37
ASO NEG
ASO LENT
γH2AX
VCL
D.
501mel
IGR37
16h
24h
48h
72h
LC3B
γH2AX
VCL

Figure S11

Figure S11. ROS production and DNA damage in LENT silenced cells. A. Flow cytometry of control or LENT-silenced 501Mel cells stained by CellROX and cleaved Caspase 3 to visualize ROS production and apoptosis. The ROS inducer TBPH was used as a positive control. Comparisons by one-way ANOVA (Dunnett test). **B.** Immunofluorescence and confocal microscopy of DNA damage marker gH2AX. Quantification of gH2AX signal in nuclei of LENT depleted cells compared to control by Mann-Whitney test. Scale bars = 10 μM. **C.** Immunblot showing gH2AX accumulation in LENT depleted cells compared to control with VCL as a loading control. **D.** Immunoblots showing accumulation of phenotypes markers at different time points upon LENT depletion. Vinculin is used as loading control.

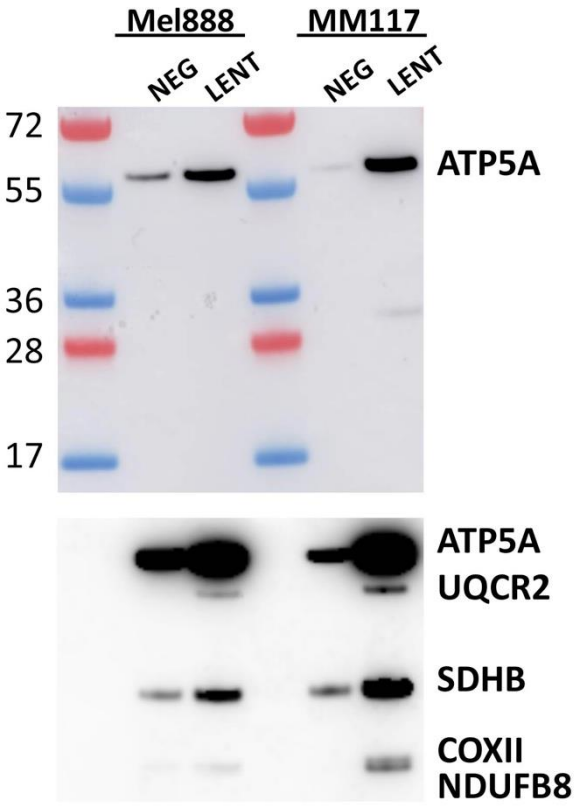


Figure S12

Figure S12. Immunoblots detecting mitochondrial electron transport chain proteins in extracts from Mel88 and MM117 cells. Bottom panel shows longer exposure.

Legends to Supplemental Datasets.

Supplemental Dataset S1. Excel spreadsheet showing genes up and down regulated by ASO silencing of LENT in 501Mel cells. On each page are shown gene symbol, Ensembl ID, description, gene biotype, Log2 fold-change and adjusted p-value.

Supplemental Dataset S2. Excel spreadsheet showing results of mass spectrometry. Shown are gene names, total number of peptides for each control (PCA3) and LENT replicate, mean values and enrichment.

Supplemental Dataset S3. Excel spreadsheet showing results of DHX36 immunoprecipitation. Page 1 shows RNAs enriched the DHX36 IP compared to control IgG, page 2 those depleted, page 3 RNAs that are enriched in the DHX36 IP from LENT silenced cells compared to control. Pages 4-6 show results of gene ontology of DHX36 enriched RNAs, those depleted by LENT silencing and those increased by LENT silencing corresponding to Figs. 3B and D. Pages 7-9 show the ontologies of RNAs common to the DHX36 IP and the G4 enrichment studies the 500 most enriched in the DHX36 IP from Sauer et al and the 500 most enriched in this study corresponding to Figs S6 F, H and I.

Supplemental Dataset S4. Excel spreadsheet showing results of RNA-seq from polysome profiling. Pages 1-6 show RNAs enriched or depleted in each fraction and page 7 the 74 RNAs common to the DHX36 IP and depleted in the LP fraction.

Supplemental Dataset S5. Excel spreadsheet showing on ontologies of RNAs enriched or depleted in polysome profiling. Page 1 shows Biological processes of the RNAs depleted in the LP fraction, (Fig. 4G) page 2 the KEGG pathways, page 3 KEGG ontology of the common genes (Fig. S7F) and page 4 the KEGG ontology of the 80S fraction with relaxed cut off value.

General Discussion and Perspectives

A. LENT is unlikely to be a multifunctional lncRNA

Our study unraveled the essential role of the lncRNA LENT in melanoma cells, specifically expressed in melanocytic cells through the control of MITF. Targeting LENT with ASOs triggered its efficient degradation in melanoma cells, leading to apoptosis, whereas no effect was seen in mesenchymal melanoma cells or non-melanoma cells where LENT is not expressed. Similarly, injection of LENT-targeting ASO in mice reduced growth of IGR37 CDX tumours, with no weight loss or other adverse effects over the short two-week timeline of the experiment. Mechanistically, we showed that LENT is present in the cytoplasm and interacts with DHX36 to regulate translation of a subset mRNAs and promote respiration and downregulate mitophagy and ER stress.

Nevertheless, as mentioned in the discussion of the article, LENT has been previously studied by others in non-melanoma contexts, notably in endothelial cells where MITF-M is not expressed [406–408]. In these cells, it was proposed that LENT expression was driven by KLF2 and KLF4 and was upregulated by sheer or oscillatory stress pathways. However, mining the GTEX data base for LENT expression in normal tissues showed its expression was highest in stomach and esophagus and examination of single cell datasets in the GTEX database did not reveal higher expression in endothelial cells compared to other tissues. Moreover, as shown in the manuscript, expression in normal tissues is much lower than in melanoma.

Nevertheless, it is possible that LENT, even if expressed at very low levels in endothelial cells, may still have a function. It was proposed that transcription of the LENT locus regulates gene expression with LENT acting as an enhancer elncRNA in the nucleus to control eNOS, essential for endothelial homeostasis [407]. This echoes with the property of many poorly conserved nuclear lncRNAs to influence nearby gene expression, even when expressed at low levels[35]. Note however that in the case of LENT, the proposed regulation is not of nearby genes, that are not co-regulated with LENT, but of NOS3 that is located on chromosome 7. It was proposed that LENT promotes inter-

chromosome interactions between the LENT locus on chromosome 14 and NOS3 on chromosome 7 through interactions with the Mediator complex, RNA pol II and KLF4.

Clearly, this proposed function is very different from our observations where LENT is predominant in the cytosol and acts on mRNA translation. Accordingly, we did not observe such NOS3 regulation in melanoma cells by RNA-seq upon LENT depletion and we did not observe enrichment of the transcriptional machinery by mass-spectrometry after LENT pulldown.

One possibility is that different LENT splice isoforms are present in endothelial cells compared to melanoma cells. Indeed, multiple LENT isoforms are reported in the genome databases and we observed that isoform 5 of LENT is the dominant isoform in melanoma cells and in patients. This isoform also seems sufficient for LENT function as its ectopic expression in HEK293T cells was sufficient to interact with DHX36 and enhanced their proliferation. Similarly, *in vitro* transcribed isoform 5 is able to directly bind DHX36, consistent with its capacity to be functional. Unfortunately, it was difficult to assess from the authors description which LENT isoform was acting as an elncRNA in endothelial cells. They also propose that LENT is conserved in mice, but only an overall 20 % sequence homology is observed, although this does not formally exclude that the mouse and human orthologues shared conserved structural features essential for their function.

Thus, it is possible that LENT could have different functions based on the expressed isoform and the cellular context, but more convincing evidence need to be provided. Nevertheless, the isoform 5 of LENT could represent a good therapeutic opportunity as it seems very specific to melanoma and is sufficient promote melanoma cell survival.

B. DHX36 localization is dependent on the cellular context

An observation made in our study was the enrichment of DHX36 at the mitochondria compared to non-melanoma cells or neural-crest like cells. LENT does not seem essential for this mitochondrial localization of DHX36 as its depletion did not influence DHX36 localization in melanocytic melanoma cells. Notably, two DHX36 isoforms have been characterized, one localized to the cytoplasm and one in the nucleus as it contains a nuclear localization signal [409], even if another article described both isoforms present in the cytoplasm [397]. Nevertheless, it has not been yet

reported in the literature that DHX36 could associate with the mitochondria. It is still unclear whether some of the DHX36 localizes inside the mitochondria or is only maintained at the outer mitochondrial membrane in melanocytic cells, considering that DHX36 does not contain a classical mitochondrial targeting signal. We plan to perform an experiment of immunogold antibody labelling coupled with transmission electron microscopy to assess precise DHX36 localization in our cell lines [410]. It would also be interesting to explore DHX36 interactome in melanoma cells through immunoprecipitation of the helicase coupled with mass-spectrometry as a way to find proteins involved in DHX36 localization [411].

Thus, DHX36 mitochondrial localization in melanocytic cells could highlight their dependency on OXPHOS as our RIP-seq data shows that many mRNAs encoding proteins involved in mitochondria function are bound and potentially regulated by the helicase.

C. DHX36 activity is regulated by lncRNAs

We showed that LENT interacts directly with DHX36 through several methods. This feature is not exclusive to LENT as we saw that DHX36 is able to bind hundreds of RNAs, displaying a low selectivity [397]. Nevertheless, it was already reported that lncRNAs binding DHX36 can regulate its function. A first study highlighted DHX36 function in colon cancer in association with a lncRNA named GSEC (G-quadruplex-forming sequence containing lncRNA) [412]. GSEC is an oncogenic lncRNA essential for colon cancer cell motility that binds directly DHX36 with its 5' sequence forming a G4. Overexpression of GSEC resulted in reduced DHX36 activity, whereas overexpression of other G4-containing RNAs showed no effect on the helicase activity. Of note, DHX36 level is not impacted by GSEC which could trigger a compensation in the cells resulting in marginal effects, as shown in other studies [378,397]. Thus, GSEC represents an example of an oncogenic lncRNA able to promote cancer progression by antagonizing DHX36.

It is further interesting to note that DHX36 expression seems correlated with an overall good prognosis of breast cancer patients, suggesting that it could represent a tumor suppressor gene [413]. Consistent with its role of tumor suppressor, overexpression of DHX36 reduced colon cancer cell motility [413]. Interestingly, this is also the case for melanoma as analysis of the TCGA SKCM dataset shows that good prognosis correlates with DHX36 expression.

Another lncRNA modulating DHX36 is SMART, which is able to bind a G4-containing mRNA coding for Mlx-γ to decoy the helicase as a way to promote a correct myogenesis by attenuating Mlx-γ translation [414]. The authors propose that SMART does not directly bind DHX36 as it does not contain a G4, but still influences DHX36 activity as SMART loss promotes translation of DHX36 targets, impairing muscle development. Thus, we are aware of at least two lncRNAs regulating DHX36, either directly or indirectly.

Like GSEC, LENT is a lncRNA which binds DHX36 to modulate its function. Whereas GSEC repress DHX36 activity, LENT seems to fine-tune DHX36 association with RNAs to promote cell survival and proliferation. The common feature of these lncRNAs-DHX36 regulatory axes is the modulation of DHX36 RNA binding profile without impairing its expression level, potentially preventing the cells from triggering compensatory mechanisms. These examples could highlight a propriety of some oncogenic lncRNAs to hijack DHX36 activity as a way to promote cancer cells proliferation and survival.

D. How does LENT regulate RNA association with DHX36?

Even if the regulatory network controlled by DHX36 in presence or absence of LENT has been characterized in our work, it is still unclear how LENT is precisely able to modulate DHX36 function. There are two main possibilities: either LENT forms a G4 and interacts with DHX36 through its described DSM and OB domains [396], or binds the helicase through another interface. If there is a G4-dependent interaction, LENT could compete with other mRNAs also containing G4s. However, this mechanism seems unlikely to happen as LENT is less abundant than other DHX36 RNA targets and has a weak G4 structure predicted. Moreover, LENT can promote DHX36 interaction with RNAs, inconsistent with a competition mechanism. It is possible that LENT interaction with DHX36, G4-dependent or independent, could modulate DHX36 conformation and/or activity in such a way as to discriminate a set of mRNAs displaying common structural features. It is also possible that LENT could also discriminate RNAs bound by DHX36 through base-pairing interaction by sequence complementarity.

We are currently working with the group of Dr Eric Ennifar to perform Cryo-electron microscopy on purified DHX36 complexed with *in vitro* transcribed LENT to characterize their interaction more

precisely and gain insights into the mechanism of action. This is challenging as for now only large ribonucleoprotein complexes such as the ribosome have been well studied with this technique [415]. It may also be possible to perform gel shift assays with truncated versions of DHX36 to assess which domains interact with LENT and to use ATPase dead mutants to assess the role of the catalytic function of DHX36 in interaction with LENT and fine-tuning of translation.

Bibliography

1. Kapranov, P. *et al.* RNA Maps Reveal New RNA Classes and a Possible Function for Pervasive Transcription. *Science* **316**, 1484–1488 (2007).
2. Uszczynska-Ratajczak, B., Lagarde, J., Frankish, A., Guigó, R. & Johnson, R. Towards a complete map of the human long non-coding RNA transcriptome. *Nat. Rev. Genet.* **19**, 535–548 (2018).
3. Li, Z. *et al.* LncBook 2.0: integrating human long non-coding RNAs with multi-omics annotations. *Nucleic Acids Res.* **51**, D186–D191 (2023).
4. The RNAcentral Consortium *et al.* RNAcentral: a comprehensive database of non-coding RNA sequences. *Nucleic Acids Res.* **45**, D128–D134 (2017).
5. Fang, S. *et al.* NONCODEV5: a comprehensive annotation database for long non-coding RNAs. *Nucleic Acids Res.* **46**, D308–D314 (2018).
6. Palazzo, A. F. & Lee, E. S. Non-coding RNA: what is functional and what is junk? *Front. Genet.* **6**, (2015).
7. Mattick, J. S. *et al.* Long non-coding RNAs: definitions, functions, challenges and recommendations. *Nat. Rev. Mol. Cell Biol.* **24**, 430–447 (2023).
8. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775–1789 (2012).
9. Cheng, J. *et al.* Transcriptional Maps of 10 Human Chromosomes at 5-Nucleotide Resolution. *Science* **308**, 1149–1154 (2005).
10. Mukherjee, N. *et al.* Integrative classification of human coding and noncoding genes through RNA metabolism profiles. *Nat. Struct. Mol. Biol.* **24**, 86–96 (2017).

11. Ma, L., Bajic, V. B. & Zhang, Z. On the classification of long non-coding RNAs. *RNA Biol.* **10**, 924–933 (2013).

12. St Laurent, G. *et al.* Intronic RNAs constitute the major fraction of the non-coding RNA in mammalian cells. *BMC Genomics* **13**, 504 (2012).

13. Askarian-Amiri, M. E. *et al.* SNORD-host RNA *Zfas1* is a regulator of mammary development and a potential marker for breast cancer. *RNA* **17**, 878–891 (2011).

14. Johnston, R., Aldrich, A. & Lyons, S. M. Roles of ribosomal RNA in health and disease. *Front. RNA Res.* **1**, 1331185 (2024).

15. Bryzghalov, O., Szcześniak, M. W. & Makałowska, I. SyntDB: defining orthologues of human long noncoding RNAs across primates. *Nucleic Acids Res.* gkz941 (2019) doi:10.1093/nar/gkz941.

16. Cabili, M. N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25**, 1915–1927 (2011).

17. Necsulea, A. *et al.* The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* **505**, 635–640 (2014).

18. Sarropoulos, I., Marin, R., Cardoso-Moreira, M. & Kaessmann, H. Developmental dynamics of lncRNAs across mammalian organs and species. *Nature* **571**, 510–514 (2019).

19. Adnane, S., Marino, A. & Leucci, E. LncRNAs in human cancers: signal from noise. *Trends Cell Biol.* **32**, 565–573 (2022).

20. Kapusta, A. *et al.* Transposable Elements Are Major Contributors to the Origin, Diversification, and Regulation of Vertebrate Long Noncoding RNAs. *PLoS Genet.* **9**, e1003470 (2013).

21. Jacobson, E. C., Pandya-Jones, A. & Plath, K. A lifelong duty: how Xist maintains the inactive X chromosome. *Curr. Opin. Genet. Dev.* **75**, 101927 (2022).

22. Hirose, T., Yamazaki, T. & Nakagawa, S. Molecular anatomy of the architectural NEAT1 noncoding RNA: The domains, interactors, and biogenesis pathway required to build phase-separated nuclear paraspeckles. *WIREs RNA* **10**, e1545 (2019).

23. Zhang, X., Hamblin, M. H. & Yin, K.-J. The long noncoding RNA Malat1: Its physiological and pathophysiological functions. *RNA Biol.* **14**, 1705–1714 (2017).

24. Ganser, L. R., Kelly, M. L., Herschlag, D. & Al-Hashimi, H. M. The roles of structural dynamics in the cellular functions of RNAs. *Nat. Rev. Mol. Cell Biol.* **20**, 474–489 (2019).

25. Loewer, S. *et al.* Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat. Genet.* **42**, 1113–1117 (2010).

26. Bourque, G. *et al.* Ten things you should know about transposable elements. *Genome Biol.* **19**, 199 (2018).

27. Vibert, J. *et al.* Oncogenic chimeric transcription factors drive tumor-specific transcription, processing, and translation of silent genomic regions. *Mol. Cell* **82**, 2458-2471.e9 (2022).

28. Feschotte, C. & Gilbert, C. Endogenous viruses: insights into viral evolution and impact on host biology. *Nat. Rev. Genet.* **13**, 283–296 (2012).

29. Babaian, A. & Mager, D. L. Endogenous retroviral promoter exaptation in human cancer. *Mob. DNA* **7**, 24 (2016).

30. Rands, C. M., Meader, S., Ponting, C. P. & Lunter, G. 8.2% of the Human Genome Is Constrained: Variation in Rates of Turnover across Functional Element Classes in the Human Lineage. *PLoS Genet.* **10**, e1004525 (2014).

31. Haerty, W. & Ponting, C. P. Mutations within lncRNAs are effectively selected against in fruitfly but not in human. *Genome Biol.* **14**, R49 (2013).

32. The ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).

33. Martens, J. A., Laprade, L. & Winston, F. Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene. *Nature* **429**, 571–574 (2004).

34. Li, W. *et al.* Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* **498**, 516–520 (2013).

35. Palazzo, A. F. & Koonin, E. V. Functional Long Non-coding RNAs Evolve from Junk Transcripts. *Cell* **183**, 1151–1161 (2020).

36. Mercer, T. R., Dinger, M. E. & Mattick, J. S. Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.* **10**, 155–159 (2009).

37. Guo, C.-J. *et al.* Distinct Processing of lncRNAs Contributes to Non-conserved Functions in Stem Cells. *Cell* **181**, 621-636.e22 (2020).

38. Herman, A. B., Tsitsipatis, D. & Gorospe, M. Integrated lncRNA function upon genomic and epigenomic regulation. *Mol. Cell* **82**, 2252–2266 (2022).

39. Wang, K. C. *et al.* A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**, 120–124 (2011).

40. Liu, F., Somarowthu, S. & Pyle, A. M. Visualizing the secondary and tertiary architectural domains of lncRNA RepA. *Nat. Chem. Biol.* **13**, 282–289 (2017).

41. Kim, D. N. *et al.* Zinc-finger protein CNBP alters the 3-D structure of lncRNA Braveheart in solution. *Nat. Commun.* **11**, 148 (2020).

42. Jain, A. K. *et al.* LncPRESS1 Is a p53-Regulated LncRNA that Safeguards Pluripotency by Disrupting SIRT6-Mediated De-acetylation of Histone H3K56. *Mol. Cell* **64**, 967–981 (2016).

43. Santos-Pereira, J. M. & Aguilera, A. R loops: new modulators of genome dynamics and function. *Nat. Rev. Genet.* **16**, 583–597 (2015).

44. Arab, K. *et al.* GADD45A binds R-loops and recruits TET1 to CpG island promoters. *Nat. Genet.* **51**, 217–223 (2019).

45. Postepska-Igielska, A. *et al.* LncRNA Khps1 Regulates Expression of the Proto-oncogene SPHK1 via Triplex-Mediated Changes in Chromatin Structure. *Mol. Cell* **60**, 626–636 (2015).

46. Hou, Y., Zhang, R. & Sun, X. Enhancer LncRNAs Influence Chromatin Interactions in Different Ways. *Front. Genet.* **10**, 936 (2019).

47. Tomita, S. *et al.* A cluster of noncoding RNAs activates the ESR1 locus during breast cancer adaptation. *Nat. Commun.* **6**, 6966 (2015).

48. Wang, B. *et al.* Liquid–liquid phase separation in human health and diseases. *Signal Transduct. Target. Ther.* **6**, 290 (2021).

49. Elguindy, M. M. & Mendell, J. T. NORAD-induced Pumilio phase separation is required for genome stability. *Nature* **595**, 303–308 (2021).

50. Smith, K. P., Hall, L. L. & Lawrence, J. B. Nuclear hubs built on RNAs and clustered organization of the genome. *Curr. Opin. Cell Biol.* **64**, 67–76 (2020).

51. Lee, Y. & Rio, D. C. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu. Rev. Biochem.* **84**, 291–323 (2015).

52. Beltran, M. *et al.* A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial–mesenchymal transition. *Genes Dev.* **22**, 756–769 (2008).

53. Engreitz, J. M. *et al.* Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* **539**, 452–455 (2016).

54. Dao, L. T. M. & Spicuglia, S. Transcriptional regulation by promoters with enhancer function. *Transcription* **9**, 307–314 (2018).

55. Gong, C. & Maquat, L. E. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3′ UTRs via Alu elements. *Nature* **470**, 284–288 (2011).

56. Faghihi, M. A. *et al.* Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of b-secretase. *Nat. Med.* **14**, (2008).

57. Aznaourova, M. *et al.* Noncoding RNA *MaIL1* is an integral component of the TLR4–TRIF pathway. *Proc. Natl. Acad. Sci.* **117**, 9042–9053 (2020).

58. Li, J. *et al.* Long noncoding RNA NRON contributes to HIV-1 latency by specifically inducing tat protein degradation. *Nat. Commun.* **7**, 11730 (2016).

59. Gilot, D. *et al.* A non-coding function of TYRP1 mRNA promotes melanoma growth. *Nat. Cell Biol.* **19**, 1348–1357 (2017).

60. O'Brien, J., Hayder, H., Zayed, Y. & Peng, C. Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. *Front. Endocrinol.* **9**, 402 (2018).

61. Hansen, T. B. *et al.* miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA: miRNA mediated cleavage of circular antisense RNA. *EMBO J.* **30**, 4414–4422 (2011).

62. Zhang, X.-O. *et al.* Complementary Sequence-Mediated Exon Circularization. *Cell* **159**, 134–147 (2014).

63. Enuka, Y. *et al.* Circular RNAs are long-lived and display only minimal early alterations in response to a growth factor. *Nucleic Acids Res.* **44**, 1370–1383 (2016).

64. Carlevaro-Fita, J., Rahim, A., Guigó, R., Vardy, L. A. & Johnson, R. Cytoplasmic long noncoding RNAs are frequently bound to and degraded at ribosomes in human cells. *RNA* **22**, 867–882 (2016).

65. van Heesch, S. *et al.* Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. *Genome Biol.* **15**, R6 (2014).

66. Yoon, J.-H. *et al.* LincRNA-p21 Suppresses Target mRNA Translation. *Mol. Cell* **47**, 648–655 (2012).

67. Kesner, J. S. *et al.* Noncoding translation mitigation. *Nature* **617**, 395–402 (2023).

68. Chekulaeva, M. & Rajewsky, N. Roles of Long Noncoding RNAs and Circular RNAs in Translation. *Cold Spring Harb. Perspect. Biol.* **11**, a032680 (2019).

69. Chong, C. *et al.* Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes. *Nat. Commun.* **11**, 1293 (2020).

70. Bartok, O. *et al.* Anti-tumour immunity induces aberrant peptide presentation in melanoma. *Nature* **590**, 332–337 (2021).

71. Rackham, O. *et al.* Long noncoding RNAs are generated from the mitochondrial genome and regulated by nuclear-encoded proteins. *RNA* **17**, 2085–2093 (2011).

72. Georg, J., Honsel, A., Voß, B., Rennenberg, H. & Hess, W. R. A long antisense RNA in plant chloroplasts. *New Phytol.* **186**, 615–622 (2010).

73. Bao, Z. *et al.* LncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases. *Nucleic Acids Res.* **47**, D1034–D1037 (2019).

74. Brown, C. J. *et al.* A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* **349**, 38–44 (1991).

75. Duret, L., Chureau, C., Samain, S., Weissenbach, J. & Avner, P. The *Xist* RNA Gene Evolved in Eutherians by Pseudogenization of a Protein-Coding Gene. *Science* **312**, 1653–1655 (2006).

76. Sun, Z., Fan, J. & Wang, Y. X-Chromosome Inactivation and Related Diseases. *Genet. Res.* **2022**, 1–8 (2022).

77. Markaki, Y. *et al.* Xist nucleates local protein gradients to propagate silencing across the X chromosome. *Cell* **184**, 6174-6192.e32 (2021).

78. McHugh, C. A. *et al.* The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232–236 (2015).

79. Chu, C. *et al.* Systematic Discovery of Xist RNA Binding Proteins. *Cell* **161**, 404–416 (2015).

80. Hutchinson, J. N. *et al.* A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics* **8**, 39 (2007).

81. Clemson, C. M. *et al.* An Architectural Role for a Nuclear Noncoding RNA: NEAT1 RNA Is Essential for the Structure of Paraspeckles. *Mol. Cell* **33**, 717–726 (2009).

82. Pisani, G. & Baron, B. Nuclear paraspeckles function in mediating gene regulatory and apoptotic pathways. *Non-Coding RNA Res.* **4**, 128–134 (2019).

83. Tripathi, V. *et al.* The Nuclear-Retained Noncoding RNA MALAT1 Regulates Alternative Splicing by Modulating SR Splicing Factor Phosphorylation. *Mol. Cell* **39**, 925–938 (2010).

84. West, J. A. *et al.* The Long Noncoding RNAs NEAT1 and MALAT1 Bind Active Chromatin Sites. *Mol. Cell* **55**, 791–802 (2014).

85. Zhang, B. *et al.* The lncRNA Malat1 Is Dispensable for Mouse Development but Its Transcription Plays a cis-Regulatory Role in the Adult. *Cell Rep.* **2**, 111–123 (2012).

86. Azzalin, C. M., Reichenbach, P., Khoriauli, L., Giulotto, E. & Lingner, J. Telomeric Repeat–Containing RNA and RNA Surveillance Factors at Mammalian Chromosome Ends. *Science* **318**, 798–801 (2007).

87. Feretzaki, M. *et al.* RAD51-dependent recruitment of TERRA lncRNA to telomeres through R-loops. *Nature* **587**, 303–308 (2020).

88. Rippe, K. & Luke, B. TERRA and the state of the telomere. *Nat. Struct. Mol. Biol.* **22**, 853–858 (2015).

89. Tufarelli, C. *et al.* Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. *Nat. Genet.* **34**, 157–165 (2003).

90. Maass, P. G. *et al.* A misplaced lncRNA causes brachydactyly in humans. *J. Clin. Invest.* **122**, 3990–4002 (2012).

91. Stamou, M. *et al.* A Balanced Translocation in Kallmann Syndrome Implicates a Long Noncoding RNA, RMST, as a GnRH Neuronal Regulator. *J. Clin. Endocrinol. Metab.* **105**, e231–e244 (2020).

92. Ang, C. E. *et al.* The novel lncRNA lnc-NR2F1 is pro-neurogenic and mutated in human neurodevelopmental disorders. *eLife* **8**, e41770 (2019).

93. Tong, X., Gu, P., Xu, S. & Lin, X. Long non-coding RNA-DANCR in human circulating monocytes: a potential biomarker associated with postmenopausal osteoporosis. *Biosci. Biotechnol. Biochem.* **79**, 732–737 (2015).

94. Morán, I. *et al.* Human β Cell Transcriptome Analysis Uncovers lncRNAs That Are Tissue-Specific, Dynamically Regulated, and Abnormally Expressed in Type 2 Diabetes. *Cell Metab.* **16**, 435–448 (2012).

95. Annilo, T., Kepp, K. & Laan, M. Natural antisense transcript of natriuretic peptide precursor A (NPPA): structural organization and modulation of NPPA expression. *BMC Mol. Biol.* **10**, 81 (2009).

96. Chauviere, L. *et al.* Firalink: A bioinformatics pipeline for long non-coding RNA data analysis. *Non-Coding RNA Res.* **8**, 602–604 (2023).

97. Goodall, G. J. & Wickramasinghe, V. O. RNA in cancer. *Nat. Rev. Cancer* **21**, 22–36 (2021).

98. Jiang, M.-C., Ni, J.-J., Cui, W.-Y., Wang, B.-Y. & Zhuo, W. Emerging roles of lncRNA in cancer and therapeutic opportunities.

99. Hessels, D. *et al.* DD3PCA3-based Molecular Urine Analysis for the Diagnosis of Prostate Cancer. *Eur. Urol.* **44**, 8–16 (2003).

100. Chun, F. K. *et al.* Prostate Cancer Gene 3 (PCA3): Development and Internal Validation of a Novel Biopsy Nomogram. *Eur. Urol.* **56**, 659–668 (2009).

101. Kaewarsa, P., Vilaivan, T. & Laiwattanapaisal, W. An origami paper-based peptide nucleic acid device coupled with label-free DNAzyme probe hybridization chain reaction for prostate cancer molecular screening test. *Anal. Chim. Acta* **1186**, 339130 (2021).

102. Ferreira, L. B. *et al.* PCA3 noncoding RNA is involved in the control of prostate-cancer cell survival and modulates androgen receptor signaling. *BMC Cancer* **12**, 507 (2012).

103. Reebye, V., Frilling, A., Habib, N. A. & Mintz, P. J. Intracellular adaptor molecules and AR signalling in the tumour microenvironment. *Cell. Signal.* **23**, 1017–1021 (2011).

104. Yuan, C., Ning, Y. & Pan, Y. Emerging roles of HOTAIR in human cancer. *J. Cell. Biochem.* **121**, 3235–3247 (2020).

105. Endo, H. *et al.* Enhanced Expression of Long Non-Coding RNA HOTAIR Is Associated with the Development of Gastric Cancer. *PLoS ONE* **8**, e77070 (2013).

106. Song, Y. *et al.* Long non-coding RNA HOTAIR mediates the switching of histone H3 lysine 27 acetylation to methylation to promote epithelial-to-mesenchymal transition in gastric cancer. *Int. J. Oncol.* (2018) doi:10.3892/ijo.2018.4625.

107. Pe, N. Tumor suppressor gene E-cadherin and its role in normal and malignant cells. *Cancer Cell Int.* (2003).

108. Yang, F. *et al.* Repression of the Long Noncoding RNA-LET by Histone Deacetylase 3 Contributes to Hypoxia-Mediated Metastasis. *Mol. Cell* **49**, 1083–1096 (2013).

109. Kuwano, Y. *et al.* NF90 selectively represses the translation of target mRNAs bearing an AU-rich signature motif. *Nucleic Acids Res.* **38**, 225–238 (2010).

110. Lee, J. W., Ko, J., Ju, C. & Eltzschig, H. K. Hypoxia signaling in human diseases and therapeutic targets. *Exp. Mol. Med.* **51**, 1–13 (2019).

111. Gutschner, T. *et al.* The Noncoding RNA *MALAT1* Is a Critical Regulator of the Metastasis Phenotype of Lung Cancer Cells. *Cancer Res.* **73**, 1180–1189 (2013).

112. Ji, P. *et al.* MALAT-1, a novel noncoding RNA, and thymosin β4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene* **22**, 8031–8041 (2003).

113. Robins, H. S. *et al.* Digital Genomic Quantification of Tumor-Infiltrating Lymphocytes. *Sci. Transl. Med.* **5**, (2013).

114. Jiang, R. *et al.* The long noncoding RNA lnc-EGFR stimulates T-regulatory cells differentiation thus promoting hepatocellular carcinoma immune evasion. *Nat. Commun.* **8**, 15129 (2017).

115. Sigismund, S., Avanzato, D. & Lanzetti, L. Emerging functions of the EGFR in cancer. *Mol. Oncol.* **12**, 3–20 (2018).

116. Mercer, T. R., Munro, T. & Mattick, J. S. The potential of long noncoding RNA therapies. *Trends Pharmacol. Sci.* **43**, 269–280 (2022).

117. Obexer, R., Nassir, M., Moody, E. R., Baran, P. S. & Lovelock, S. L. Modern approaches to therapeutic oligonucleotide manufacturing. *Science* **384**, eadl4015 (2024).

118. Quemener, A. M. *et al.* The powerful world of antisense oligonucleotides: From bench to bedside. *WIREs RNA* **11**, e1594 (2020).

119. Hu, B. *et al.* Therapeutic siRNA: state of the art. *Signal Transduct. Target. Ther.* **5**, 101 (2020).

120. Arraiano, C. M. *et al.* The critical role of RNA processing and degradation in the control of gene expression. *FEMS Microbiol. Rev.* **34**, 883–923 (2010).

121. Mello, C. C. & Conte, D. Revealing the world of RNA interference. *Nature* **431**, 338–342 (2004).

122. Berkhout, B. & Haasnoot, J. The interplay between virus infection and the cellular RNA interference machinery. *FEBS Lett.* **580**, 2896–2902 (2006).

123. Hyjek, M., Figiel, M. & Nowotny, M. RNases H: Structure and mechanism. *DNA Repair* **84**, 102672 (2019).

124. Izant, J. G. & Weintraub, H. Inhibition of thymidine kinase gene expression by anti-sense RNA: A molecular approach to genetic analysis. *Cell* **36**, 1007–1015 (1984).

125. Hamann, A., Nguyen, A. & Pannier, A. K. Nucleic acid delivery to mesenchymal stem cells: a review of nonviral methods and applications. *J. Biol. Eng.* **13**, 7 (2019).

126. Krützfeldt, J. *et al.* Silencing of microRNAs in vivo with 'antagomirs'. *Nature* **438**, 685–689 (2005).

127. Dominski, Z. & Kole, R. Restoration of correct splicing in thalassemic pre-mRNA by antisense oligonucleotides. *Proc. Natl. Acad. Sci.* **90**, 8673–8677 (1993).

128. Wheeler, T. M. *et al.* Reversal of RNA Dominance by Displacement of Protein Sequestered on Triplet Repeat RNA. *Science* **325**, 336–339 (2009).

129. Mologni, L., Nielsen, P. E. & Gambacorti-Passerini, C. In Vitro Transcriptional and Translational Block of the bcl-2 Gene Operated by Peptide Nucleic Acid. *Biochem. Biophys. Res. Commun.* **264**, 537–543 (1999).

130. Vickers, T. A. Fully modified 2' MOE oligonucleotides redirect polyadenylation. *Nucleic Acids Res.* **29**, 1293–1299 (2001).

131. Liang, X. *et al.* Translation efficiency of mRNAs is increased by antisense oligonucleotides targeting upstream open reading frames. *Nat. Biotechnol.* **34**, 875–880 (2016).

132. Barbosa, C., Peixeiro, I. & Romão, L. Gene Expression Regulation by Upstream Open Reading Frames and Human Disease. *PLoS Genet.* **9**, e1003529 (2013).

133. Dowdy, S. F. Overcoming cellular barriers for RNA therapeutics. *Nat. Biotechnol.* **35**, 222–229 (2017).

134. Tsui, N. B., Ng, E. K. & Lo, Y. D. Stability of Endogenous and Added RNA in Blood Specimens, Serum, and Plasma. *Clin. Chem.* **48**, 1647–1653 (2002).

135. McNab, F., Mayer-Barber, K., Sher, A., Wack, A. & O'Garra, A. Type I interferons in infectious disease. *Nat. Rev. Immunol.* **15**, 87–103 (2015).

136. Kurreck, J. Antisense technologies: Improvement through novel chemical modifications. *Eur. J. Biochem.* **270**, 1628–1644 (2003).

137. Koziolkiewicz, M., Gendaszewska, E., Maszewska, M., Stein, C. A. & Stec, W. J. The mononucleotide-dependent, nonantisense mechanism of action of phosphodiester and phosphorothioate oligonucleotides depends upon the activity of an ecto-5J-nucleotidase. **98**, (2001).

138. Monia, B. P. *et al.* Evaluation of 2'-modified oligonucleotides containing 2'-deoxy gaps as antisense inhibitors of gene expression. *J. Biol. Chem.* **268**, 14514–14522 (1993).

139. Seth, P. P. *et al.* Design, Synthesis And Evaluation Of Constrained Methoxyethyl (cMOE) and Constrained Ethyl (cEt) Nucleoside Analogs. *Nucleic Acids Symp. Ser.* **52**, 553–554 (2008).

140. Singh, S. K., Koshkin, A. A., Wengel, J. & Nielsen, P. LNA (locked nucleic acids): synthesis and high-affinity nucleic acid recognition. *Chem. Commun.* 455–456 (1998) doi:10.1039/a708608c.

141. Zhang, Y. *et al.* 5-methylcytosine (m^5C) RNA modification controls the innate immune response to virus infection by regulating type I interferons. *Proc. Natl. Acad. Sci.* **119**, e2123338119 (2022).

142. Aboshi, M. *et al.* Safety and immunogenicity of VLPCOV-02, a SARS-CoV-2 self-amplifying RNA vaccine with a modified base, 5-methylcytosine. *iScience* **27**, 108964 (2024).

143. Bege, M. & Borbás, A. Rise and fall of fomivirsen, the first approved gene silencing medicine : A historical review. *Acta Pharm. Hung.* **92**, 38–44 (2022).

144. Geary, R. S., Henry, S. P. & Grillone, L. R. Fomivirsen: Clinical Pharmacology and Potential Drug Interactions. *Clin. Pharmacokinet.* **41**, 255–260 (2002).

145. Chambergo-Michilot, D., Alur, A., Kulkarni, S. & Agarwala, A. Mipomersen in Familial Hypercholesterolemia: An Update on Health-Related Quality of Life and Patient-Reported Outcomes. *Vasc. Health Risk Manag.* **Volume 18**, 73–80 (2022).

146. Fogacci, F. *et al.* Efficacy and Safety of Mipomersen: A Systematic Review and Meta-Analysis of Randomized Clinical Trials. *Drugs* **79**, 751–766 (2019).

147. Burel, S. A. *et al.* Hepatotoxicity of high affinity gapmer antisense oligonucleotides is mediated by RNase H1 dependent promiscuous reduction of very long pre-mRNA transcripts. *Nucleic Acids Res.* **44**, 2093–2109 (2016).

148. Antisense Oligonucleotide-Mediated Exon-skipping Therapies: Precision Medicine Spreading from Duchenne Muscular Dystrophy. *JMA J.* **4**, 232–240 (2021).

149. Kim, J. *et al.* Patient-Customized Oligonucleotide Therapy for a Rare Genetic Disease. *N. Engl. J. Med.* **381**, 1644–1652 (2019).

150. Copolovici, D. M., Langel, K., Eriste, E. & Langel, Ü. Cell-Penetrating Peptides: Design, Synthesis, and Applications. *ACS Nano* **8**, 1972–1994 (2014).

151. Arnold, A. E. *et al.* Antibody-Antisense Oligonucleotide Conjugate Downregulates a Key Gene in Glioblastoma Stem Cells. *Mol. Ther. - Nucleic Acids* **11**, 518–527 (2018).

152. Sicard, G. *et al.* Enhanced Antisense Oligonucleotide Delivery Using Cationic Liposomes Grafted with Trastuzumab: A Proof-of-Concept Study in Prostate Cancer. *Pharmaceutics* **12**, 1166 (2020).

153. Abdo, J. M., Sopko, N. A. & Milner, S. M. The applied anatomy of human skin: A model for regeneration. *Wound Med.* **28**, 100179 (2020).

154. Krispin, S., Nitzan, E., Kassem, Y. & Kalcheim, C. Evidence for a dynamic spatiotemporal fate map and early fate restrictions of premigratory avian neural crest. *Development* **137**, 585–595 (2010).

155. Jin, E.-J., Erickson, C. A., Takada, S. & Burrus, L. W. Wnt and BMP Signaling Govern Lineage Segregation of Melanocytes in the Avian Embryo. *Dev. Biol.* **233**, 22–37 (2001).

156. Renauld, J. M., Davis, W., Cai, T., Cabrera, C. & Basch, M. L. Transcriptomic analysis and *ednrb* expression in cochlear intermediate cells reveal developmental differences between inner ear and skin melanocytes. *Pigment Cell Melanoma Res.* **34**, 585–597 (2021).

157. Adameyko, I. *et al.* Schwann Cell Precursors from Nerve Innervation Are a Cellular Origin of Melanocytes in Skin. *Cell* **139**, 366–379 (2009).

158. Bonaventure, J., Domingues, M. J. & Larue, L. Cellular and molecular mechanisms controlling the migration of melanocytes and melanoma cells. *Pigment Cell Melanoma Res.* **26**, 316–325 (2013).

159. Luciani, F. *et al.* Biological and mathematical modeling of melanocyte development. *Development* **138**, 3943–3954 (2011).

160. Adameyko, I. *et al.* Sox2 and Mitf cross-regulatory interactions consolidate progenitor and melanocyte lineages in the cranial neural crest. *Development* **139**, 397–410 (2012).

161. Goding, C. R. & Arnheiter, H. MITF—the first 25 years. *Genes Dev.* **33**, 983–1007 (2019).

162. Levy, C., Khaled, M. & Fisher, D. E. MITF: master regulator of melanocyte development and melanoma oncogene. *Trends Mol. Med.* **12**, 406–414 (2006).

163. Laurette, P. *et al.* Transcription factor MITF and remodeller BRG1 define chromatin organisation at regulatory elements in melanoma cells. *eLife* **4**, e06857 (2015).

164. Pogenberg, V. *et al.* Restricted leucine zipper dimerization and specificity of DNA recognition of the melanocyte master regulator MITF. *Genes Dev.* **26**, 2647–2658 (2012).

165. Strub, T. *et al.* Essential role of microphthalmia transcription factor for DNA replication, mitosis and genomic stability in melanoma. *Oncogene* **30**, 2319–2332 (2011).

166. Bharti, K., Liu, W., Csermely, T., Bertuzzi, S. & Arnheiter, H. Alternative promoter use in eye development: complex role and regulation of the transcription factor MITF. (2008).

167. Amae, S. *et al.* Identification of a Novel Isoform of Microphthalmia-Associated Transcription Factor That Is Enriched in Retinal Pigment Epithelium. *Biochem. Biophys. Res. Commun.* **247**, 710–715 (1998).

168. Tshori, S. Transcription factor MITF regulates cardiac growth and hypertrophy. *J. Clin. Invest.* **116**, 2673–2681 (2006).

169. George, A. *et al.* Biallelic Mutations in MITF Cause Coloboma, Osteopetrosis, Microphthalmia, Macrocephaly, Albinism, and Deafness. *Am. J. Hum. Genet.* **99**, 1388–1394 (2016).

170. Hughes, A. E., Newton, V. E., Liu, X. Z. & Read, A. P. A gene for Waardenburg Syndrome type 2 maps close to the human homologue of the microphthalmia gene at chromosome 3p12–p14.1. *Nat. Genet.* **7**, 509–512 (1994).

171. Gleason, B. C., Crum, C. P. & Murphy, G. F. Expression patterns of MITF during human cutaneous embryogenesis: evidence for bulge epithelial expression and persistence of dermal melanoblasts. *J. Cutan. Pathol.* **35**, 615–622 (2008).

172. Hirobe, T. Control of Melanocyte Proliferation and Differentiation in the Mouse Epidermis. *Pigment Cell Res.* **5**, 1–11 (1992).

173. Laurent, P. *et al.* Simulation of melanoblast displacements reveals new features of developmental migration. *Development* dev.160200 (2018) doi:10.1242/dev.160200.

174. Nishimura, E. K. *et al.* Dominant role of the niche in melanocyte stem-cell fate determination. *Nature* **416**, 854–860 (2002).

175. Osawa, M. *et al.* Molecular characterization of melanocyte stem cells in their niche. *Development* **132**, 5589–5599 (2005).

176. Mort, R. L., Jackson, I. J. & Patton, E. E. The melanocyte lineage in development and disease. *Development* **142**, 1387–1387 (2015).

177. Tu, S. & Johnson, S. L. Clonal analyses reveal roles of organ founding stem cells, melanocyte stem cells and melanoblasts in establishment, growth and regeneration of the adult zebrafish fin. *Development* **137**, 3931–3939 (2010).

178. Centeno, P. P., Pavet, V. & Marais, R. The journey from melanocytes to melanoma. *Nat. Rev. Cancer* **23**, 372–390 (2023).

179. Sun, Q. *et al.* A novel mouse model demonstrates that oncogenic melanocyte stem cells engender melanoma resembling human disease. *Nat. Commun.* **10**, 5023 (2019).

180. Goding, C. R. Melanocytes: The new Black. *Int. J. Biochem. Cell Biol.* **39**, 275–279 (2007).

181. Domingues, L. *et al.* Coupling of melanocyte signaling and mechanics by caveolae is required for human skin pigmentation. *Nat. Commun.* **11**, 2988 (2020).

182. Lai, X., Wichers, H. J., Soler-Lopez, M. & Dijkstra, B. W. Structure and Function of Human Tyrosinase and Tyrosinase-Related Proteins. *Chem. – Eur. J.* **24**, 47–55 (2018).

183. Liu, Y. *et al.* Comparison of Structural and Chemical Properties of Black and Red Human Hair Melanosomes¶. *Photochem. Photobiol.* **81**, 135 (2005).

184. Hoashi, T. *et al.* The Repeat Domain of the Melanosomal Matrix Protein PMEL17/GP100 Is Required for the Formation of Organellar Fibers. *J. Biol. Chem.* **281**, 21198–21208 (2006).

185. Hoashi, T. *et al.* MART-1 Is Required for the Function of the Melanosomal Matrix Protein PMEL17/GP100 and the Maturation of Melanosomes. *J. Biol. Chem.* **280**, 14006–14016 (2005).

186. Sitaram, A. & Marks, M. S. Mechanisms of Protein Delivery to Melanosomes in Pigment Cells. *Physiology* **27**, 85–99 (2012).

187. D'Alba, L. & Shawkey, M. D. Melanosomes: Biogenesis, Properties, and Evolution of an Ancient Organelle. *Physiol. Rev.* **99**, 1–19 (2019).

188. Wu, X. & Hammer, J. A. Melanosome transfer: it is best to give and receive. *Curr. Opin. Cell Biol.* **29**, 1–7 (2014).

189. Chaplin, G. Geographic distribution of environmental factors influencing human skin coloration. *Am. J. Phys. Anthropol.* **125**, 292–302 (2004).

190. Matsumura, Y. & Ananthaswamy, H. N. Toxic effects of ultraviolet radiation on the skin. *Toxicol. Appl. Pharmacol.* **195**, 298–308 (2004).

191. Tadokoro, T. *et al.* Mechanisms of Skin Tanning in Different Racial/Ethnic Groups in Response to Ultraviolet Radiation. *J. Invest. Dermatol.* **124**, 1326–1332 (2005).

192. Randolph Byers, H., Maheshwary, S., Amodeo, D. M. & Dykstra, S. G. Role of Cytoplasmic Dynein in Perinuclear Aggregation of Phagocytosed Melanosomes and Supranuclear Melanin Cap Formation in Human Keratinocytes. *J. Invest. Dermatol.* **121**, 813–820 (2003).

193. Brenner, M. & Hearing, V. J. The Protective Role of Melanin Against UV Damage in Human Skin †. *Photochem. Photobiol.* **84**, 539–549 (2008).

194. Hoath, S. B. & Leahy, D. G. The Organization of Human Epidermis: Functional Epidermal Units and Phi Proportionality. *J. Invest. Dermatol.* **121**, 1440–1446 (2003).

195. Slominski, A. T. *et al.* Key Role of CRF in the Skin Stress Response System. *Endocr. Rev.* **34**, 827–884 (2013).

196. Kadekaro, A. L. *et al. Melanocortin 1 receptor* genotype: an important determinant of the damage response of melanocytes to ultraviolet radiation. *FASEB J.* **24**, 3850–3860 (2010).

197. Makova, K. & Norton, H. Worldwide polymorphism at the MC1R locus and normal pigmentation variation in humans. *Peptides* **26**, 1901–1908 (2005).

198. Chou, W. C. *et al.* Direct migration of follicular melanocyte stem cells to the epidermis after wounding or UVB irradiation is dependent on Mc1r signaling. *Nat. Med.* **19**, 924–929 (2013).

199. Rabbani, P. *et al.* Coordinated Activation of Wnt in Epithelial and Melanocyte Stem Cells Initiates Pigmented Hair Regeneration. *Cell* **145**, 941–955 (2011).

200. Jones, P., Lucock, M., Veysey, M. & Beckett, E. The Vitamin D–Folate Hypothesis as an Evolutionary Model for Skin Pigmentation: An Update and Integration of Current Ideas. *Nutrients* **10**, 554 (2018).

201. DeLuca, H. F. Overview of general physiologic features and functions of vitamin D. *Am. J. Clin. Nutr.* **80**, 1689S-1696S (2004).

202. Lucock, M. Folic Acid: Nutritional Biochemistry, Molecular Biology, and Role in Disease Processes. *Mol. Genet. Metab.* **71**, 121–138 (2000).

203. Off, M. K. *et al.* Ultraviolet photodegradation of folic acid. *J. Photochem. Photobiol. B* **80**, 47–55 (2005).

204. Belote, R. L. *et al.* Human melanocyte development and melanoma dedifferentiation at single-cell resolution. *Nat. Cell Biol.* **23**, 1035–1047 (2021).

205. Tang, J. *et al.* The genomic landscapes of individual melanocytes from human skin. *Nature* **586**, 600–605 (2020).

206. Australian Pancreatic Cancer Genome Initiative *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).

207. Akbani, R. *et al.* Genomic Classification of Cutaneous Melanoma. *Cell* **161**, 1681–1696 (2015).

208. Cargnello, M. & Roux, P. P. Activation and Function of the MAPKs and Their Substrates, the MAPK-Activated Protein Kinases. *Microbiol. Mol. Biol. Rev.* **75**, 50–83 (2011).

209. Davies, H. *et al.* Mutations of the BRAF gene in human cancer. *Nature* **417**, 949–954 (2002).

210. Ostojić, J. *et al.* Transcriptional co-activator regulates melanocyte differentiation and oncogenesis by integrating cAMP and MAPK/ERK pathways. *Cell Rep.* **35**, 109136 (2021).

211. Trucco, L. D. *et al.* Ultraviolet radiation–induced DNA damage is prognostic for outcome in melanoma. *Nat. Med.* **25**, 221–224 (2019).

212. Daya-Grosjean, L., Dumaz, N. & Sarasin, A. The specificity of p53 mutation spectra in sunlight induced human cancers. *J. Photochem. Photobiol. B* **28**, 115–124 (1995).

213. Wang, H.-T., Choi, B. & Tang, M. Melanocytes are deficient in repair of oxidative DNA damage and UV-induced photoproducts. *Proc. Natl. Acad. Sci.* **107**, 12180–12185 (2010).

214. Denat, L., Kadekaro, A. L., Marrot, L., Leachman, S. A. & Abdel-Malek, Z. A. Melanocytes as Instigators and Victims of Oxidative Stress. *J. Invest. Dermatol.* **134**, 1512–1518 (2014).

215. Newell, F. *et al.* Whole-genome landscape of mucosal melanoma reveals diverse drivers and therapeutic targets. *Nat. Commun.* **10**, 3163 (2019).

216. Liang, W. S. *et al.* Integrated genomic analyses reveal frequent *TERT* aberrations in acral melanoma. *Genome Res.* **27**, 524–532 (2017).

217. Johansson, P. A. *et al.* Whole genome landscapes of uveal melanoma show an ultraviolet radiation signature in iris tumours. *Nat. Commun.* **11**, 2408 (2020).

218. Pandiani, C., Béranger, G. E., Leclerc, J., Ballotti, R. & Bertolotto, C. Focus on cutaneous and uveal melanoma specificities. *Genes Dev.* **31**, 724–743 (2017).

219. Bauer, J. *et al. BRAF* mutations in cutaneous melanoma are independently associated with age, anatomic site of the primary tumor, and the degree of solar elastosis at the primary tumor site. *Pigment Cell Melanoma Res.* **24**, 345–351 (2011).

220. Pollock, P. M. *et al.* High frequency of BRAF mutations in nevi. *Nat. Genet.* **33**, 19–20 (2003).

221. Giuliano, S., Ohanna, M., Ballotti, R. & Bertolotto, C. Advances in melanoma senescence and potential clinical application. *Pigment Cell Melanoma Res.* **24**, 295–308 (2011).

222. The French Familial Melanoma Study Group *et al.* A SUMOylation-defective MITF germline mutation predisposes to melanoma and renal carcinoma. *Nature* **480**, 94–98 (2011).

223. Curtin, J. A., Patel, H. N., Cho, K.-H. & LeBoit, P. E. Distinct Sets of Genetic Alterations in Melanoma. *N. Engl. J. Med.* (2005).

224. Serrano, M. *et al.* Role of the INK4a Locus in Tumor Suppression and Cell Mortality. *Cell* **85**, 27–37 (1996).

225. Hodis, E. *et al.* A Landscape of Driver Mutations in Melanoma. *Cell* **150**, 251–263 (2012).

226. Ozaki, T. & Nakagawara, A. Role of p53 in Cell Death and Human Cancers. *Cancers* **3**, 994–1013 (2011).

227. Lahalle, A. *et al.* The p53 Pathway and Metabolism: The Tree That Hides the Forest. *Cancers* **13**, 133 (2021).

228. Shain, A. H. *et al.* The Genetic Evolution of Melanoma from Precursor Lesions. *N. Engl. J. Med.* **373**, 1926–1936 (2015).

229. Clark, W., From, L., Bernardino, E. & Mihm, M. C. The Histogenesis and Biologic Behavior of Primary Human Malignant Melanomas of the Skin. (1969).

230. Cyclin D1 Is a Candidate Oncogene in Cutaneous Melanoma 1.

231. McDermott, N. C. *et al.* Identification of Vertical Growth Phase in Malignant Melanoma: A Study of Interobserver Agreement. *Am. J. Clin. Pathol.* **110**, 753–757 (1998).

232. Mrozik, K. M., Blaschuk, O. W., Cheong, C. M., Zannettino, A. C. W. & Vandyke, K. N-cadherin in cancer metastasis, its emerging role in haematological malignancies and potential as a therapeutic target in cancer. *BMC Cancer* **18**, 939 (2018).

233. Damsky, W. E., Theodosakis, N. & Bosenberg, M. Melanoma metastasis: new concepts and evolving paradigms. *Oncogene* **33**, 2413–2422 (2014).

234. Jin, X. *et al.* A metastasis map of human cancer cell lines. *Nature* **588**, 331–336 (2020).

235. Martínez-Jiménez, F. *et al.* Pan-cancer whole-genome comparison of primary and metastatic solid tumours. *Nature* **618**, 333–341 (2023).

236. Cymerman, R. M. *et al.* De Novo vs Nevus-Associated Melanomas: Differences in Associations With Prognostic Indicators and Survival. *J. Natl. Cancer Inst.* **108**, djw121 (2016).

237. Walter, F. M. *et al.* Using the 7-point checklist as a diagnostic aid for pigmented skin lesions in general practice: a diagnostic validation study. *Br. J. Gen. Pract.* **63**, e345–e353 (2013).

238. Liao, S. & von der Weid, P. Y. Lymphatic system: An active pathway for immune protection. *Semin. Cell Dev. Biol.* **38**, 83–89 (2015).

239. Saginala, K., Barsouk, A., Aluru, J. S., Rawla, P. & Barsouk, A. Epidemiology of Melanoma. *Med. Sci.* **9**, 63 (2021).

240. Arnold, M. *et al.* Global Burden of Cutaneous Melanoma in 2020 and Projections to 2040. *JAMA Dermatol.* **158**, 495 (2022).

241. Wilson, M. A. & Schuchter, L. M. Chemotherapy for Melanoma. in *Melanoma* (eds. Kaufman, H. L. & Mehnert, J. M.) vol. 167 209–229 (Springer International Publishing, Cham, 2016).

242. Haas, L., Wiesner, T. & Obenauf, A. C. A new era of proactive melanoma therapy: hit hard, hit early. *Br. J. Dermatol.* **178**, 817–820 (2018).

243. Balkwill, F. & Mantovani, A. Inflammation and cancer: back to Virchow? *The Lancet* **357**, 539–545 (2001).

244. Gata, V. A., Lisencu, C. I., Vlad, C. I., Piciu, D. & Irimie, A. Tumor infiltrating lymphocytes as a prognostic factor in malignant melanoma. Review of the literature.

245. Fenton, S. E., Saleiro, D. & Platanias, L. C. Type I and II Interferons in the Anti-Tumor Immune Response. *Cancers* **13**, 1037 (2021).

246. Kirkwood, J. M. *et al.* High-Dose Interferon Alfa-2b Significantly Prolongs Relapse-Free and Overall Survival Compared With the GM2-KLH/QS-21 Vaccine in Patients With Resected Stage IIB-III Melanoma: Results of Intergroup Trial E1694/S9512/C509801. *J. Clin. Oncol.* **19**, 2370–2380 (2001).

247. Eggermont, A. M. M. *et al.* Long term follow up of the EORTC 18952 trial of adjuvant therapy in resected stage IIB–III cutaneous melanoma patients comparing intermediate doses of interferon-alpha-2b (IFN) with observation: Ulceration of primary is key determinant for IFN-sensitivity. *Eur. J. Cancer* **55**, 111–121 (2016).

248. Eggermont, A. M. M. *et al.* Adjuvant therapy with pegylated interferon alfa-2b versus observation alone in resected stage III melanoma: final results of EORTC 18991, a randomised phase III trial. **372**, (2008).

249. Krieg, C., Létourneau, S., Pantaleo, G. & Boyman, O. Improved IL-2 immunotherapy by selective stimulation of IL-2 receptors on lymphocytes and endothelial cells. *Proc. Natl. Acad. Sci.* **107**, 11906–11911 (2010).

250. Han, Y., Guo, Q., Zhang, M., Chen, Z. & Cao, X. CD69+CD4+CD25− T Cells, a New Subset of Regulatory T Cells, Suppress T Cell Proliferation through Membrane-Bound TGF-β1. *J. Immunol.* **182**, 111–120 (2009).

251. Wang, Z. *et al.* Ontak-like human IL-2 fusion toxin. *J. Immunol. Methods* **448**, 51–58 (2017).

252. Attia, P., Maker, A. V., Haworth, L. R., Rogers-Freezer, L. & Rosenberg, S. A. Inability of a Fusion Protein of IL-2 and Diphtheria Toxin (Denileukin Diftitox, DAB389IL-2, ONTAK) to Eliminate Regulatory T Lymphocytes in Patients With Melanoma: *J. Immunother.* **28**, 582–592 (2005).

253. Ribas, A. *et al.* Intratumoral Immune Cell Infiltrates, FoxP3, and Indoleamine 2,3-Dioxygenase in Patients with Melanoma Undergoing CTLA4 Blockade. *Clin. Cancer Res.* **15**, 390–399 (2009).

254. Chambers, C. A., Kuhns, M. S., Egen, J. G. & Allison, J. P. CTLA-4-Mediated Inhibition in Regulation of T Cell Responses: Mechanisms and Manipulation in Tumor Immunotherapy. *Annu. Rev. Immunol.* **19**, 565–594 (2001).

255. Brohl, A. S. *et al.* A phase IB study of ipilimumab with peginterferon alfa-2b in patients with unresectable melanoma. *J. Immunother. Cancer* **4**, 85 (2016).

256. Han, Y., Liu, D. & Li, L. PD-1/PD-L1 pathway: current researches in cancer.

257. Raedler, L. A. Opdivo (Nivolumab): Second PD-1 Inhibitor Receives FDA Approval for Unresectable or Metastatic Melanoma. **8**, (2015).

258. Specenier, P. Nivolumab in melanoma. *Expert Rev. Anticancer Ther.* **16**, 1247–1261 (2016).

259. Robert, C. *et al.* Pembrolizumab versus Ipilimumab in Advanced Melanoma. *N. Engl. J. Med.* **372**, 2521–2532 (2015).

260. Wolchok, J. D. *et al.* Overall Survival with Combined Nivolumab and Ipilimumab in Advanced Melanoma. *N. Engl. J. Med.* **377**, 1345–1356 (2017).

261. Morad, G., Helmink, B. A., Sharma, P. & Wargo, J. A. Hallmarks of response, resistance, and toxicity to immune checkpoint blockade. *Cell* **184**, 5309–5337 (2021).

262. Liu, Y.-T. & Sun, Z.-J. Turning cold tumors into hot tumors by improving T-cell infiltration. *Theranostics* **11**, 5365–5386 (2021).

263. Chapman, P. B. *et al.* Improved Survival with Vemurafenib in Melanoma with BRAF V600E Mutation. *N. Engl. J. Med.* **364**, 2507–2516 (2011).

264. Ballantyne, A. D. & Garnock-Jones, K. P. Dabrafenib: First Global Approval. *Drugs* **73**, 1367–1376 (2013).

265. Koelblinger, P., Thuerigen, O. & Dummer, R. Development of encorafenib for BRAF-mutated advanced melanoma. *Curr. Opin. Oncol.* **30**, 125–133 (2018).

266. Livingstone, E., Zimmer, L., Vaubel, J. & Schadendorf, D. BRAF, MEK and KIT inhibitors for melanoma: adverse events and their management. *Chin. Clin. Oncol.* **3**, (2014).

267. Rizos, H. *et al.* BRAF Inhibitor Resistance Mechanisms in Metastatic Melanoma: Spectrum and Clinical Impact. *Clin. Cancer Res.* **20**, 1965–1977 (2014).

268. Czarnecka, A. M., Bartnik, E., Fiedorowicz, M. & Rutkowski, P. Targeted Therapy in Melanoma and Mechanisms of Resistance. *Int. J. Mol. Sci.* **21**, 4576 (2020).

269. Shi, H. *et al.* Melanoma whole-exome sequencing identifies V600EB-RAF amplification-mediated acquired B-RAF inhibitor resistance. *Nat. Commun.* **3**, 724 (2012).

270. Poulikakos, P. I., Zhang, C., Bollag, G., Shokat, K. M. & Rosen, N. RAF inhibitors transactivate RAF dimers and ERK signalling in cells with wild-type BRAF. *Nature* **464**, 427–430 (2010).

271. Thota, R., Johnson, D. B. & Sosman, J. A. Trametinib in the treatment of melanoma. *Expert Opin. Biol. Ther.* **15**, 735–747 (2015).

272. Flaherty, K. T. *et al.* Combined BRAF and MEK Inhibition in Melanoma with BRAF V600 Mutations. *N. Engl. J. Med.* **367**, 1694–1703 (2012).

273. Larkin, J. *et al.* Combined Vemurafenib and Cobimetinib in *BRAF* -Mutated Melanoma. *N. Engl. J. Med.* **371**, 1867–1876 (2014).

274. Ascierto, P. A. *et al.* Overall survival with first-line atezolizumab in combination with vemurafenib and cobimetinib in BRAFV600 mutation-positive advanced melanoma (IMspire150): second interim analysis of a multicentre, randomised, phase 3 study. *Lancet Oncol.* **24**, 33–44 (2023).

275. Sharma, P., Hu-Lieskovan, S., Wargo, J. A. & Ribas, A. Primary, Adaptive, and Acquired Resistance to Cancer Immunotherapy. *Cell* **168**, 707–723 (2017).

276. Haas, L. *et al.* Acquired resistance to anti-MAPK targeted therapy confers an immune-evasive tumor microenvironment and cross-resistance to immunotherapy in melanoma. *Nat. Cancer* **2**, 693–708 (2021).

277. Dagogo-Jack, I. & Shaw, A. T. Tumour heterogeneity and resistance to cancer therapies. *Nat. Rev. Clin. Oncol.* **15**, 81–94 (2018).

278. Ennen, M. *et al.* Single-cell gene expression signatures reveal melanoma cell heterogeneity. *Oncogene* **34**, 3251–3263 (2015).

279. Haque, A., Engel, J., Teichmann, S. A. & Lönnberg, T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med.* **9**, 75 (2017).

280. Hoek, K. S. *et al.* Metastatic potential of melanomas defined by specific gene expression profiles with no BRAF signature. *Pigment Cell Res.* **19**, 290–302 (2006).

281. Zhang, Z. *et al.* Activation of the AXL kinase causes resistance to EGFR-targeted therapy in lung cancer. *Nat. Genet.* **44**, 852–860 (2012).

282. Boiko, A. D. *et al.* Human melanoma-initiating cells express neural crest nerve growth factor receptor CD271. *Nature* **466**, 133–137 (2010).

283. Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).

284. Tsoi, J. *et al.* Multi-stage Differentiation Defines Melanoma Subtypes with Differential Vulnerability to Drug-Induced Iron-Dependent Oxidative Stress. *Cancer Cell* **33**, 890-904.e5 (2018).

285. Shakhova, O. *et al.* Sox10 promotes the formation and maintenance of giant congenital naevi and melanoma. *Nat. Cell Biol.* **14**, 882–890 (2012).

286. Rambow, F. *et al.* Toward Minimal Residual Disease-Directed Therapy in Melanoma. *Cell* **174**, 843-855.e19 (2018).

287. Kim, P., Cheng, F., Zhao, J. & Zhao, Z. ccmGDB: a database for cancer cell metabolism genes. *Nucleic Acids Res.* **44**, D959–D968 (2016).

288. Wouters, J. *et al.* Robust gene expression programs underlie recurrent cell states and phenotype switching in melanoma. *Nat. Cell Biol.* **22**, 986–998 (2020).

289. Karras, P. *et al.* A cellular hierarchy in melanoma uncouples growth and metastasis. *Nature* **610**, 190–198 (2022).

290. Batlle, E. & Clevers, H. Cancer stem cells revisited. *Nat. Med.* **23**, 1124–1134 (2017).

291. Pozniak, J. *et al.* A TCF4-dependent gene regulatory network confers resistance to immunotherapy in melanoma. *Cell* **187**, 166-183.e25 (2024).

292. Pishesha, N., Harmand, T. J. & Ploegh, H. L. A guide to antigen processing and presentation. *Nat. Rev. Immunol.* **22**, 751–764 (2022).

293. Gavish, A. *et al.* Hallmarks of transcriptional intratumour heterogeneity across a thousand tumours. *Nature* **618**, 598–606 (2023).

294. Bassez, A. *et al.* A single-cell map of intratumoral changes during anti-PD1 treatment of patients with breast cancer. *Nat. Med.* **27**, 820–832 (2021).

295. Chen, Z., Han, F., Du, Y., Shi, H. & Zhou, W. Hypoxic microenvironment in cancer: molecular mechanisms and therapeutic interventions. *Signal Transduct. Target. Ther.* **8**, 70 (2023).

296. Williams, A. B. & Schumacher, B. p53 in the DNA-Damage-Repair Process. *Cold Spring Harb. Perspect. Med.* **6**, a026070 (2016).

297. Ennen, M. *et al. MITF* -High and *MITF* -Low Cells and a Novel Subpopulation Expressing Genes of Both Cell States Contribute to Intra- and Intertumoral Heterogeneity of Primary Melanoma. *Clin. Cancer Res.* **23**, 7097–7107 (2017).

298. Hugo, W. *et al.* Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. *Cell* **165**, 35–44 (2016).

299. Barrett, R. L. & Puré, E. Cancer-associated fibroblasts and their influence on tumor immunity and immunotherapy. *eLife* **9**, e57243 (2020).

300. Corre, S. *et al.* Sustained activation of the Aryl hydrocarbon Receptor transcription factor promotes resistance to BRAF-inhibitors in melanoma. *Nat. Commun.* **9**, 4775 (2018).

301. Falletta, P. *et al.* Translation reprogramming is an evolutionarily conserved driver of phenotypic plasticity and therapeutic resistance in melanoma. *Genes Dev.* **31**, 18–33 (2017).

302. Hoek, K. S. *et al. In vivo* Switching of Human Melanoma Cells between Proliferative and Invasive States. *Cancer Res.* **68**, 650–656 (2008).

303. Chauhan, J. S., Hölzel, M., Lambert, J., Buffa, F. M. & Goding, C. R. The MITF regulatory network in melanoma. *Pigment Cell Melanoma Res.* **35**, 517–533 (2022).

304. Riesenberg, S. *et al.* MITF and c-Jun antagonism interconnects melanoma dedifferentiation with pro-inflammatory cytokine responsiveness and myeloid cell recruitment. *Nat. Commun.* **6**, 8755 (2015).

305. Verfaillie, A. *et al.* Decoding the regulatory landscape of melanoma reveals TEADS as regulators of the invasive cell state. *Nat. Commun.* **6**, 6683 (2015).

306. Pastushenko, I. *et al.* Identification of the tumour transition states occurring during EMT. *Nature* **556**, 463–468 (2018).

307. Heerboth, S. *et al.* EMT and tumor metastasis. *Clin. Transl. Med.* **4**, e6 (2015).

308. Berico, P. *et al.* CDK7 and MITF repress a transcription program involved in survival and drug tolerance in melanoma. *EMBO Rep.* **22**, e51683 (2021).

309. Fontanals-Cirera, B. *et al.* Harnessing BET Inhibitor Sensitivity Reveals AMIGO2 as a Melanoma Survival Gene. *Mol. Cell* **68**, 731-744.e9 (2017).

310. Robey, R. W. *et al.* Revisiting the role of ABC transporters in multidrug-resistant cancer. *Nat. Rev. Cancer* **18**, 452–464 (2018).

311. García-Jiménez, C. & Goding, C. R. Starvation and Pseudo-Starvation as Drivers of Cancer Metastasis through Translation Reprogramming. *Cell Metab.* **29**, 254–267 (2019).

312. Bailey, C. M., Morrison, J. A. & Kulesa, P. M. Melanoma revives an embryonic migration program to promote plasticity and invasion. *Pigment Cell Melanoma Res.* **25**, 573–583 (2012).

313. Kunz, M. *et al.* RNA-seq analysis identifies different transcriptomic types and developmental trajectories of primary melanomas. *Oncogene* **37**, 6136–6151 (2018).

314. Marin-Bejar, O. *et al.* Evolutionary predictability of genetic versus nongenetic resistance to anticancer drugs in melanoma. *Cancer Cell* **39**, 1135-1149.e8 (2021).

315. Warburg, O. The Metabolism of Carcinoma Cells. *J. Cancer Res.* **9**, 148–163 (1925).

316. Pfeiffer, T., Schuster, S. & Bonhoeffer, S. Cooperation and Competition in the Evolution of ATP-Producing Pathways. *Science* **292**, 504–507 (2001).

317. Finley, L. W. S. What is cancer metabolism? *Cell* **186**, 1670–1688 (2023).

318. Scott, D. A. *et al.* Comparative Metabolic Flux Profiling of Melanoma Cell Lines. *J. Biol. Chem.* **286**, 42626–42634 (2011).

319. Liberti, M. V. & Locasale, J. W. The Warburg Effect: How Does it Benefit Cancer Cells? *Trends Biochem. Sci.* **41**, 211–218 (2016).

320. Coassolo, S. *et al.* Citrullination of pyruvate kinase M2 by PADI1 and PADI3 regulates glycolysis and cancer cell proliferation. *Nat. Commun.* **12**, 1718 (2021).

321. Hosseini, M., Kasraian, Z. & Rezvani, H. R. Energy metabolism in skin cancers: A therapeutic perspective. *Biochim. Biophys. Acta BBA - Bioenerg.* **1858**, 712–722 (2017).

322. Hudson, C. C. *et al.* Regulation of Hypoxia-Inducible Factor 1␣ Expression and Function by the Mammalian Target of Rapamycin. *MOL CELL BIOL* **22**, (2002).

323. Davies, M. A. *et al.* A novel AKT3 mutation in melanoma tumours and cell lines. *Br. J. Cancer* **99**, 1265–1268 (2008).

324. Hall, A. *et al.* Dysfunctional oxidative phosphorylation makes malignant melanoma cells addicted to glycolysis driven by the V600EBRAF oncogene. *Oncotarget* **4**, 584–599 (2013).

325. Haq, R. *et al.* Oncogenic BRAF Regulates Oxidative Metabolism via PGC1α and MITF. *Cancer Cell* **23**, 302–315 (2013).

326. Vazquez, F. *et al.* PGC1α Expression Defines a Subset of Human Melanoma Tumors with Increased Mitochondrial Capacity and Resistance to Oxidative Stress. *Cancer Cell* **23**, 287–301 (2013).

327. Shimobayashi, M. & Hall, M. N. Making new contacts: the mTOR network in metabolism and signalling crosstalk. *Nat. Rev. Mol. Cell Biol.* **15**, 155–162 (2014).

328. Jose, C., Bellance, N. & Rossignol, R. Choosing between glycolysis and oxidative phosphorylation: A tumor's dilemma? *Biochim. Biophys. Acta BBA - Bioenerg.* **1807**, 552–561 (2011).

329. Shen, S. *et al.* An epitranscriptomic mechanism underlies selective mRNA translation remodelling in melanoma persister cells. *Nat. Commun.* **10**, 5713 (2019).

330. Ferguson, J., Smith, M., Zudaire, I., Wellbrock, C. & Arozarena, I. Glucose availability controls ATF4-mediated MITF suppression to drive melanoma cell growth. *Oncotarget* **8**, 32946–32959 (2017).

331. Romero-Garcia, S., Moreno-Altamirano, M. M. B., Prado-Garcia, H. & Sánchez-García, F. J. Lactate Contribution to the Tumor Microenvironment: Mechanisms, Effects on Immune Cells and Therapeutic Relevance. *Front. Immunol.* **7**, (2016).

332. Arena, G. *et al.* Mitochondrial MDM2 Regulates Respiratory Complex I Activity Independently of p53. *Mol. Cell* **69**, 594-609.e8 (2018).

333. Kook, S. *et al.* Mdm2 enhances ligase activity of parkin and facilitates mitophagy. *Sci. Rep.* **10**, 5028 (2020).

334. Kumar, P. *et al.* Deciphering oxygen distribution and hypoxia profiles in the tumor microenvironment: a data-driven mechanistic modeling approach. *Phys. Med. Biol.* **69**, 125023 (2024).

335. Ho, J. *et al.* Importance of glycolysis and oxidative phosphorylation in advanced melanoma. *Mol. Cancer* **11**, 76 (2012).

336. Ma, L. & Zong, X. Metabolic Symbiosis in Chemoresistance: Refocusing the Role of Aerobic Glycolysis. *Front. Oncol.* **10**, 5 (2020).

337. Parry, R. V. *et al.* CTLA-4 and PD-1 Receptors Inhibit T-Cell Activation by Distinct Mechanisms. *MOL CELL BIOL* **25**, (2005).

338. Patsoukis, N. *et al.* PD-1 alters T-cell metabolic reprogramming by inhibiting glycolysis and promoting lipolysis and fatty acid oxidation. *Nat. Commun.* **6**, 6692 (2015).

339. Inamdar, S. *et al.* Rescue of dendritic cells from glycolysis inhibition improves cancer immunotherapy in mice. *Nat. Commun.* **14**, 5333 (2023).

340. Parmenter, T. J. *et al.* Response of *BRAF* -Mutant Melanoma to BRAF Inhibition Is Mediated by a Network of Transcriptional Regulators of Glycolysis. *Cancer Discov.* **4**, 423–433 (2014).

341. Gopal, Y. N. V. *et al.* Inhibition of mTORC1/2 Overcomes Resistance to MAPK Pathway Inhibitors Mediated by PGC1α and Oxidative Phosphorylation in Melanoma. *Cancer Res.* **74**, 7037–7047 (2014).

342. Roesch, A. *et al.* Overcoming Intrinsic Multidrug Resistance in Melanoma by Blocking the Mitochondrial Respiratory Chain of Slow-Cycling JARID1Bhigh Cells. *Cancer Cell* **23**, 811–825 (2013).

343. Goguet-Rubio, P. *et al.* E4F1-mediated control of pyruvate dehydrogenase activity is essential for skin homeostasis. *Proc. Natl. Acad. Sci.* **113**, 11004–11009 (2016).

344. Rodier, G. *et al.* The Transcription Factor E4F1 Coordinates CHK1-Dependent Checkpoint and Mitochondrial Functions. *Cell Rep.* **11**, 220–233 (2015).

345. Zhang, G. *et al.* Targeting mitochondrial biogenesis to overcome drug resistance to MAPK inhibitors. *J. Clin. Invest.* **126**, 1834–1856 (2016).

346. Vendramin, R. *et al.* Activation of the integrated stress response confers vulnerability to mitoribosome-targeting antibiotics in melanoma. *J. Exp. Med.* **218**, e20210571 (2021).

347. Cerezo, M., Tomic, T., Ballotti, R. & Rocchi, S. Is it time to test biguanide metformin in the treatment of melanoma? *Pigment Cell Melanoma Res.* **28**, 8–20 (2015).

348. Lehraiki, A. *et al.* Inhibition of Melanogenesis by the Antidiabetic Metformin. *J. Invest. Dermatol.* **134**, 2589–2597 (2014).

349. Niehr, F. *et al.* Combination therapy with vemurafenib (PLX4032/RG7204) and metformin in melanoma cell lines with distinct driver mutations. *J. Transl. Med.* **9**, 76 (2011).

350. Martin, M. J., Hayward, R., Viros, A. & Marais, R. Metformin Accelerates the Growth of BRAFV600E-Driven Melanoma by Upregulating VEGF-A. *Cancer Discov.* **2**, 344–355 (2012).

351. Wahl, D. *et al.* Characterization of the metabolic phenotype of chronically activated lymphocytes. *Lupus* **19**, 1492–1501 (2010).

352. Leucci, E. *et al.* Melanoma addiction to the long non-coding RNA SAMMSON. *Nature* **531**, 518–522 (2016).

353. Brannan, K. *et al.* mRNA Decapping Factors and the Exonuclease Xrn2 Function in Widespread Premature Termination of RNA Polymerase II Transcription. *Mol. Cell* **46**, 311–324 (2012).

354. Fogal, V. *et al.* Mitochondrial p32 Protein Is a Critical Regulator of Tumor Metabolism via Maintenance of Oxidative Phosphorylation. *Mol. Cell. Biol.* **30**, 1303–1318 (2010).

355. Richter-Dennerlein, R., Dennerlein, S. & Rehling, P. Integrating mitochondrial translation into the cellular context. *Nat. Rev. Mol. Cell Biol.* **16**, 586–592 (2015).

356. Wrobel, L. *et al.* Mistargeted mitochondrial proteins activate a proteostatic response in the cytosol. *Nature* **524**, 485–488 (2015).

357. Bykov, Y. S., Rapaport, D., Herrmann, J. M. & Schuldiner, M. Cytosolic Events in the Biogenesis of Mitochondrial Proteins. *Trends Biochem. Sci.* **45**, 650–667 (2020).

358. Coyne, L. P. & Chen, X. J. MPOS is a novel mitochondrial trigger of cell death – implications for neurodegeneration. *FEBS Lett.* **592**, 759–775 (2018).

359. Vendramin, R. *et al.* SAMMSON fosters cancer cell fitness by concertedly enhancing mitochondrial and cytosolic translation. *Nat. Struct. Mol. Biol.* **25**, 1035–1046 (2018).

360. Cheung, C. T., Singh, R., Kalra, R. S., Kaul, S. C. & Wadhwa, R. Collaborator of ARF (CARF) Regulates Proliferative Fate of Human Cells by Dose-dependent Regulation of DNA Damage Signaling. *J. Biol. Chem.* **289**, 18258–18269 (2014).

361. Vadivel Gnanasundram, S. & Fåhraeus, R. Translation Stress Regulates Ribosome Synthesis and Cell Proliferation. *Int. J. Mol. Sci.* **19**, 3757 (2018).

362. Seberg, H. E., Van Otterloo, E. & Cornell, R. A. Beyond MITF : Multiple transcription factors directly regulate the cellular phenotype in melanocytes and melanoma. *Pigment Cell Melanoma Res.* **30**, 454–466 (2017).

363. Gerami, P. *et al.* Development and validation of a noninvasive 2-gene molecular assay for cutaneous melanoma. *J. Am. Acad. Dermatol.* **76**, 114-120.e2 (2017).

364. Kawabe, H. *et al.* Regulation of Rap2A by the Ubiquitin Ligase Nedd4-1 Controls Neurite Development. *Neuron* **65**, 358–372 (2010).

365. Machida, N. *et al.* Mitogen-activated Protein Kinase Kinase Kinase Kinase 4 as a Putative Effector of Rap2 to Activate the c-Jun N-terminal Kinase. *J. Biol. Chem.* **279**, 15711–15714 (2004).

366. Adebayo, M., Singh, S., Singh, A. P. & Dasgupta, S. Mitochondrial fusion and fission: The fine-tune balance for cellular homeostasis. *FASEB J.* **35**, (2021).

367. Kashatus, J. A. *et al.* Erk2 Phosphorylation of Drp1 Promotes Mitochondrial Fission and MAPK-Driven Tumor Growth. *Mol. Cell* **57**, 537–551 (2015).

368. Cribbs, J. T. & Strack, S. Reversible phosphorylation of Drp1 by cyclic AMP-dependent protein kinase and calcineurin regulates mitochondrial fission and cell death. *EMBO Rep.* **8**, 939–944 (2007).

369. Hernandez-Alias, X., Benisty, H., Schaefer, M. H. & Serrano, L. Translational efficiency across healthy and tumor tissues is proliferation-related. *Mol. Syst. Biol.* **16**, e9275 (2020).

370. Melixetian, M. *et al.* Long non-coding RNA TINCR suppresses metastatic melanoma dissemination by preventing ATF4 translation. *EMBO Rep.* **22**, e50852 (2021).

371. Khaitan, D. *et al.* The Melanoma-Upregulated Long Noncoding RNA *SPRY4-IT1* Modulates Apoptosis and Invasion. *Cancer Res.* **71**, 3852–3862 (2011).

372. Spiegel, J., Adhikari, S. & Balasubramanian, S. The Structure and Function of DNA G-Quadruplexes. *Trends Chem.* **2**, 123–136 (2020).

373. Murat, P. & Balasubramanian, S. Existence and consequences of G-quadruplex structures in DNA. *Curr. Opin. Genet. Dev.* **25**, 22–29 (2014).

374. Kwok, C. K., Marsico, G., Sahakyan, A. B., Chambers, V. S. & Balasubramanian, S. rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nat. Methods* **13**, 841–844 (2016).

375. Guo, J. U. & Bartel, D. P. RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science* **353**, aaf5371–aaf5371 (2016).

376. Bedrat, A., Lacroix, L. & Mergny, J.-L. Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.* **44**, 1746–1759 (2016).

377. Chery, M. *et al.* The Arabidopsis TDR Ala forms G-quadruplex structures that can be unwound by the DEXH1 DEA (D/H)-box RNA helicase. *Plant J.* **118**, 124–140 (2024).

378. Dumas, L., Herviou, P., Dassi, E., Cammas, A. & Millevoi, S. G-Quadruplexes in RNA Biology: Recent Advances and Future Directions. *Trends Biochem. Sci.* **46**, 270–283 (2021).

379. Esnault, C. *et al.* G4access identifies G-quadruplexes and their associations with open chromatin and imprinting control regions. *Nat. Genet.* **55**, 1359–1369 (2023).

380. Shen, J. *et al.* Promoter G-quadruplex folding precedes transcription and is controlled by chromatin. *Genome Biol.* **22**, 143 (2021).

381. Falabella, M. *et al.* G-quadruplex dynamics contribute to regulation of mitochondrial gene expression. *Sci. Rep.* **9**, 5605 (2019).

382. Bryan, T. M. G-Quadruplexes at Telomeres: Friend or Foe? *Molecules* **25**, 3686 (2020).

383. Byrd, A. K. *et al.* Evidence That G-quadruplex DNA Accumulates in the Cytoplasm and Participates in Stress Granule Assembly in Response to Oxidative Stress. *J. Biol. Chem.* **291**, 18041–18057 (2016).

384. Protter, D. S. W. & Parker, R. Principles and Properties of Stress Granules. *Trends Cell Biol.* **26**, 668–679 (2016).

385. Kharel, P. *et al.* Stress promotes RNA G-quadruplex folding in human cells. *Nat. Commun.* **14**, 205 (2023).

386. Asamitsu, S. *et al.* RNA G-quadruplex organizes stress granule assembly through DNAPTP6 in neurons. *Sci. Adv.* **9**, eade2035 (2023).

387. Huppert, J. L., Bugaut, A., Kumari, S. & Balasubramanian, S. G-quadruplexes: the beginning and end of UTRs. *Nucleic Acids Res.* **36**, 6260–6268 (2008).

388. Varshney, D. *et al.* RNA G-quadruplex structures control ribosomal protein production. *Sci. Rep.* **11**, 22735 (2021).

389. Cammas, A. & Millevoi, S. RNA G-quadruplexes: emerging mechanisms in disease. *Nucleic Acids Res.* gkw1280 (2016) doi:10.1093/nar/gkw1280.

390. Leonetti, C. *et al.* G-Quadruplex Ligand RHPS4 Potentiates the Antitumor Activity of Camptothecins in Preclinical Models of Solid Tumors. *Clin. Cancer Res.* **14**, 7284–7291 (2008).

391. Wu, T.-Y., Huang, Q., Huang, Z.-S., Hu, M.-H. & Tan, J.-H. A drug-like imidazole-benzothiazole conjugate inhibits malignant melanoma by stabilizing the c-MYC G-quadruplex. *Bioorganic Chem.* **99**, 103866 (2020).

392. Platella, C. *et al.* Selective Targeting of Cancer-Related G-Quadruplex Structures by the Natural Compound Dicentrine. *Int. J. Mol. Sci.* **24**, 4070 (2023).

393. Schult, P. & Paeschke, K. The DEAH helicase *DHX36* and its role in G-quadruplex-dependent processes. *Biol. Chem.* **402**, 581–591 (2021).

394. Vaughn, J. P. *et al.* The DEXH Protein Product of the DHX36 Gene Is the Major Source of Tetramolecular Quadruplex G4-DNA Resolving Activity in HeLa Cell Lysates. *J. Biol. Chem.* **280**, 38117–38120 (2005).

395. Lattmann, S., Giri, B., Vaughn, J. P., Akman, S. A. & Nagamine, Y. Role of the amino terminal RHAU-specific motif in the recognition and resolution of guanine quadruplex-RNA by the DEAH-box RNA helicase RHAU. *Nucleic Acids Res.* **38**, 6219–6233 (2010).

396. Chen, M. C. *et al.* Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36. *Nature* **558**, 465–469 (2018).

397. Sauer, M. *et al.* DHX36 prevents the accumulation of translationally inactive mRNAs with G4-structures in untranslated regions. *Nat. Commun.* **10**, 2421 (2019).

398. Meyer, C. *et al.* The TIA1 RNA-Binding Protein Family Regulates EIF2AK2-Mediated Stress Response and Cell Cycle Progression. *Mol. Cell* **69**, 622-635.e6 (2018).

399. Chen, X. *et al.* Translational control by DHX36 binding to 5′UTR G-quadruplex is essential for muscle stem-cell regenerative functions. *Nat. Commun.* **12**, 5043 (2021).

400. Kaur, E., Agrawal, R. & Sengupta, S. Functions of BLM Helicase in Cells: Is It Acting Like a Double-Edged Sword? *Front. Genet.* **12**, 634789 (2021).

401. Tang, W. *et al.* The Werner syndrome RECQ helicase targets G4 DNA in human cells to modulate transcription. *Hum. Mol. Genet.* **25**, 2060–2069 (2016).

402. Murat, P. *et al.* RNA G-quadruplexes at upstream open reading frames cause DHX36- and DHX9-dependent translation of human mRNAs. *Genome Biol.* **19**, 229 (2018).

403. Kim, H.-N. *et al.* Histone deacetylase inhibitor MS-275 stimulates bone formation in part by enhancing Dhx36-mediated TNAP transcription. *J. Bone Miner. Res.* **26**, 2161–2173 (2011).

404. Huang, W. *et al.* Yin Yang 1 contains G-quadruplex structures in its promoter and 5′-UTR and its expression is modulated by G4 resolvase 1. *Nucleic Acids Res.* **40**, 1033–1049 (2012).

405. Booy, E. P. *et al.* The RNA helicase RHAU (DHX36) unwinds a G4-quadruplex in human telomerase RNA and promotes the formation of the P1 helix template boundary. *Nucleic Acids Res.* **40**, 4110–4124 (2012).

406. Stanicek, L. *et al.* Long non-coding RNA LASSIE regulates shear stress sensing and endothelial barrier function. *Commun. Biol.* **3**, 265 (2020).

407. Miao, Y. *et al.* Enhancer-associated long non-coding RNA LEENE regulates endothelial nitric oxide synthase and endothelial function. *Nat. Commun.* **9**, 292 (2018).

408. Tang, X. *et al.* Long noncoding RNA LEENE promotes angiogenesis and ischemic recovery in diabetes models. *J. Clin. Invest.* **133**, e161759 (2023).

409. Tran, H., Schilling, M., Wirbelauer, C., Hess, D. & Nagamine, Y. Facilitation of mRNA Deadenylation and Decay by the Exosome-Bound, DExH Protein RHAU. *Mol. Cell* **13**, 101–111 (2004).

410. Bergersen, L. H., Storm-Mathisen, J. & Gundersen, V. Immunogold quantification of amino acids and proteins in complex subcellular compartments. *Nat. Protoc.* **3**, 144–152 (2008).

411. Richards, A. L., Eckhardt, M. & Krogan, N. J. Mass spectrometry-based protein–protein interaction networks for the study of human diseases. *Mol. Syst. Biol.* **17**, e8792 (2021).

412. Matsumura, K. *et al.* The novel G-quadruplex-containing long non-coding RNA GSEC antagonizes DHX36 and modulates colon cancer cell migration. *Oncogene* **36**, 1191–1199 (2017).

413. Zeng, Y. *et al.* Identification of DHX36 as a tumour suppressor through modulating the activities of the stress-associated proteins and cyclin-dependent kinases in breast cancer.

414. Martone, J. *et al.* SMaRT lncRNA controls translation of a G-quadruplex-containing mRNA antagonizing the DHX36 helicase. *EMBO Rep.* **21**, e49942 (2020).

415. Dimitrova-Paternoga, L., Jagtap, P. K. A., Chen, P.-C. & Hennig, J. Integrative Structural Biology of Protein-RNA Complexes. *Structure* **28**, 6–28 (2020).

Annexes

Article 1: SMARCB1 regulates a TFCP2L1-MYC transcriptional switch promoting renal medullary carcinoma transformation and ferroptosis resistance

This study was published in Nature Commmunications in 2023. It describes the role of *SMARCB1*, a gene encoding BAF47, a protein part of the SWI/SNF chromatin remodeler complex, lost in cases of renal medullary carcinoma (RMC).

I helped in the assessment of changes in the proteome of RMC cells upon re-expression of *SMARCB1*. Several components of the SWI/SNF complex such as BRM and BAF45c are re-integrated into the complex, showing the essential role of BAF47 as a scaffolder of this complex. Moreover, epithelial markers such as MITF or CDH1 are gained upon *SMARCB1* expression whereas mesenchymal markers such as VIM or MYC are lost, highlighting that SMARC1 re-expression promotes a mesenchymal to epithelial transition in these cells.

I also participated in the characterization of the RMC cells death by ferroptosis upon re-expression of *SMARCB1* by flow cytometry. Indeed, *SMARCB1* expression promotes TFCP2L1 activation, restoring ferroptosis sensitivity in RMC cells as it also downregulates NFE2L2, a known effector of a ferroptosis-resistance program.

Lastly, I characterized BRG1 (*SMARCA4*) interactome in presence or absence of BAF47 to assess which proteins are incorporated into the SWI/SNF complex in these two conditions by immunoprecipitation.

Article 1: SMARCB1 regulates a TFCP2L1-MYC transcriptional switch promoting renal medullary carcinoma transformation and ferroptosis resistance

nature communications

Article

https://doi.org/10.1038/s41467-023-38472-y

SMARCB1 regulates a TFCP2L1-MYC transcriptional switch promoting renal medullary carcinoma transformation and ferroptosis resistance

Received: 16 December 2021

Accepted: 3 May 2023

Published online: 26 May 2023

Check for updates

Bujamin H. Vokshi[1,12], Guillaume Davidson[1,12], Nassim Tawanaie Pour Sedehi[1], Alexandra Helleux[1], Marc Rippinger[1], Alexandre R. Haller[1], Justine Gantzer[1,2], Jonathan Thouvenin[1,2], Philippe Baltzinger[1], Rachida Bouarich[3], Valeria Manriquez[3], Sakina Zaidi[3], Priya Rao[4], Pavlos Msaouel[5], Xiaoping Su[6], Hervé Lang[7], Thibault Tricard[7], Véronique Lindner[8], Didier Surdez[9,10], Jean-Emmanuel Kurtz[2], Franck Bourdeaut[3], Nizar M. Tannir[5], Irwin Davidson[1,11,13] ✉ & Gabriel G. Malouf[1,2,11,13] ✉

Renal medullary carcinoma (RMC) is an aggressive tumour driven by bi-allelic loss of SMARCB1 and tightly associated with sickle cell trait. However, the cell-of-origin and oncogenic mechanism remain poorly understood. Using single-cell sequencing of human RMC, we defined transformation of thick ascending limb (TAL) cells into an epithelial-mesenchymal gradient of RMC cells associated with loss of renal epithelial transcription factors *TFCP2L1*, *HOXB9* and *MITF* and gain of *MYC* and *NFE2L2*-associated oncogenic and ferroptosis resistance programs. We describe the molecular basis for this transcriptional switch that is reversed by SMARCB1 re-expression repressing the oncogenic and ferroptosis resistance programs leading to ferroptotic cell death. Ferroptosis resistance links TAL cell survival with the high extracellular medullar iron concentrations associated with sickle cell trait, an environment propitious to the mutagenic events associated with RMC development. This unique environment may explain why RMC is the only SMARCB1-deficient tumour arising from epithelial cells, differentiating RMC from rhabdoid tumours arising from neural crest cells.

First described in 1995[1], renal medullary carcinoma (RMC) is a lethal malignant neoplasm arising from the kidney medulla region. Despite its relative rarity, RMC is the third most common renal cancer among young adults[2]. It typically afflicts male patients of African descent with sickle cell trait at a median age of 28 years, yet the association is still poorly understood[3,4]. RMC is highly aggressive with most patients presenting metastatic disease at the time of diagnosis and less than 5% survive longer than 36 months[5,6]. In addition, RMC tumours are resistant to targeted therapies used for other renal cancers and the best available cytotoxic chemotherapy regimens produce a brief objective response in less than 30% of cases[7,8]. Alternative treatments such as anti-angiogenics, EZH2 inhibitors and immunotherapy have been tested with varying success[6]. RMC tumour tissue resembles a high-grade carcinoma exhibiting reticular or cribriform patterns and usually stain positive for VIM, MUC1, pankeratins, PAX8, HIF1α and VEGF[8,9]. RMC are also characterized by a strong desmoplasia, a

A full list of affiliations appears at the end of the paper. ✉e-mail: irwin@igbmc.fr; maloufg@igbmc.fr

Article 1: SMARCB1 regulates a TFCP2L1-MYC transcriptional switch promoting renal medullary carcinoma transformation and ferroptosis resistance

prominent inflammatory infiltrate as well as the frequent presence of sickled red blood cells[10,11].

The hallmark of RMC is loss of SMARCB1 expression[12], a core subunit of the SWItch/Sucrose Non-Fermentable (SWI/SNF) chromatin remodelling complex. Several mechanisms lead to SMARCB1 loss in RMC including deletions, point mutations, inactivating translocations and loss-of-heterozygosity[6]. SMARCB1 loss is also the hallmark of malignant rhabdoid tumours (RTs), atypical teratoid/rhabdoid tumours (ATRTs) and epithelioid sarcomas (ESs). The majority of RTs and RMCs share common features such as their renal location and low mutation burden[6]. We recently characterized the molecular characteristics of RMC identifying frequent chromosome 8q gain associated with a copy-number gain of MYC[6]. SMARCB1 loss activates the MYC pathway resulting in increased DNA replication stress and DNA damage response. RMC are thought to arise from the distal region of the nephron, however evidence is limited to correlation inference using bulk RNA-seq data from 8 nephron biopsies with identified renal cell populations[6,13]. Thus, despite the above pathology and molecular characterization, the cell of RMC origin is as yet not fully defined and the molecular mechanisms involved in oncogenic transformation associated with SMARCB1 loss remain poorly characterized.

To address these issues, we integrated data from single-cell (sc) RNA sequencing of human tumours, multi-region RNA sequencing, bulk transcriptomic data from 2 RMC cohorts, and SMARCB1 gain of function experiments in cellular models. This comprehensive approach revealed how the thick ascending limb (TAL) cells are transformed into RMC through a transcriptional switch involving loss of renal master regulator TFCP2L1 and activation of a MYC and NFE2L2-associated transformation and ferroptosis resistance programs.

Results

RMC ontogeny and molecular characterization of tumour cell states

To characterize the molecular features and ontogeny of RMC, we performed scRNA-seq on a post-treatment primary nephrectomy from an RMC patient with lung metastases at diagnosis. The patient showed complete response following 6 cycles of Methotrexate, Vinblastine, Doxorubicin, Cisplatin (MVAC) treatment. A total of 996 cells from the residual tumour site and 1722 cells from normal adjacent renal tissue (NAT) were aggregated and analysed. Seurat UMAP clustering revealed 14 distinct populations amongst which were 7 renal epithelial clusters and 7 stromal and immune clusters (Fig. 1a, b). Epithelial clusters comprised 6 groups of cells from the proximal and distal tubules and 1 group of collecting duct cells each expressing specific markers (Fig. 1c). Amongst these, we identified thick ascending limb (TAL) cells with expression of *SLC12A1*, *EPCAM*, *CDH1* and keratin 7 (*KRT7*), consistent with previous renal scRNA-seq datasets[14–16].

After merging cancer and NAT samples, we identified populations enriched in the tumour sample comprising TAMs (tumour-associated macrophages) and 2 clusters of cells harbouring an epithelial mesenchymal transition (EMT) signature that we identified to be the RMC tumour and CAF (cancer-associated fibroblast) cells (Fig. 1b). All three clusters expressed specific markers (*LYZ*, *MMP7* and *POSTN*, respectively) with cytokeratin expression in RMC cells (Fig. 1c). Further analyses of RMC and CAFs showed that each expressed overlapping as well as distinct sets of EMT markers (Fig. S1a and Supplementary Data 1a).

The UMAP plot revealed that RMC cells were located close to the TAL population, consistent with a putative cell of origin located in the distal part of the nephron. We interrogated all renal epithelial populations for shared transcriptional signatures with RMC cells and found the best correlation with TAL cells of the kidney medulla (Fig. 1d). Differential gene expression analysis of a pseudo-bulk reconstitution of the RMC versus the CAF populations identified about 150 signature genes for RMC and 50 genes for CAF (Fig. 1e). RMC cells showed a specific oncogenic program, but retained many genes associated with TAL and more broadly epithelial identities (Fig. 1e). RMC and CAF cells did however commonly express EMT genes such as *VIM* and *FN1*, in contrast to TAL cells (Figs. 1c, e). Altogether, these observations identified TAL cells to be the normal renal population most related to RMC and hence the likely cell-of-origin.

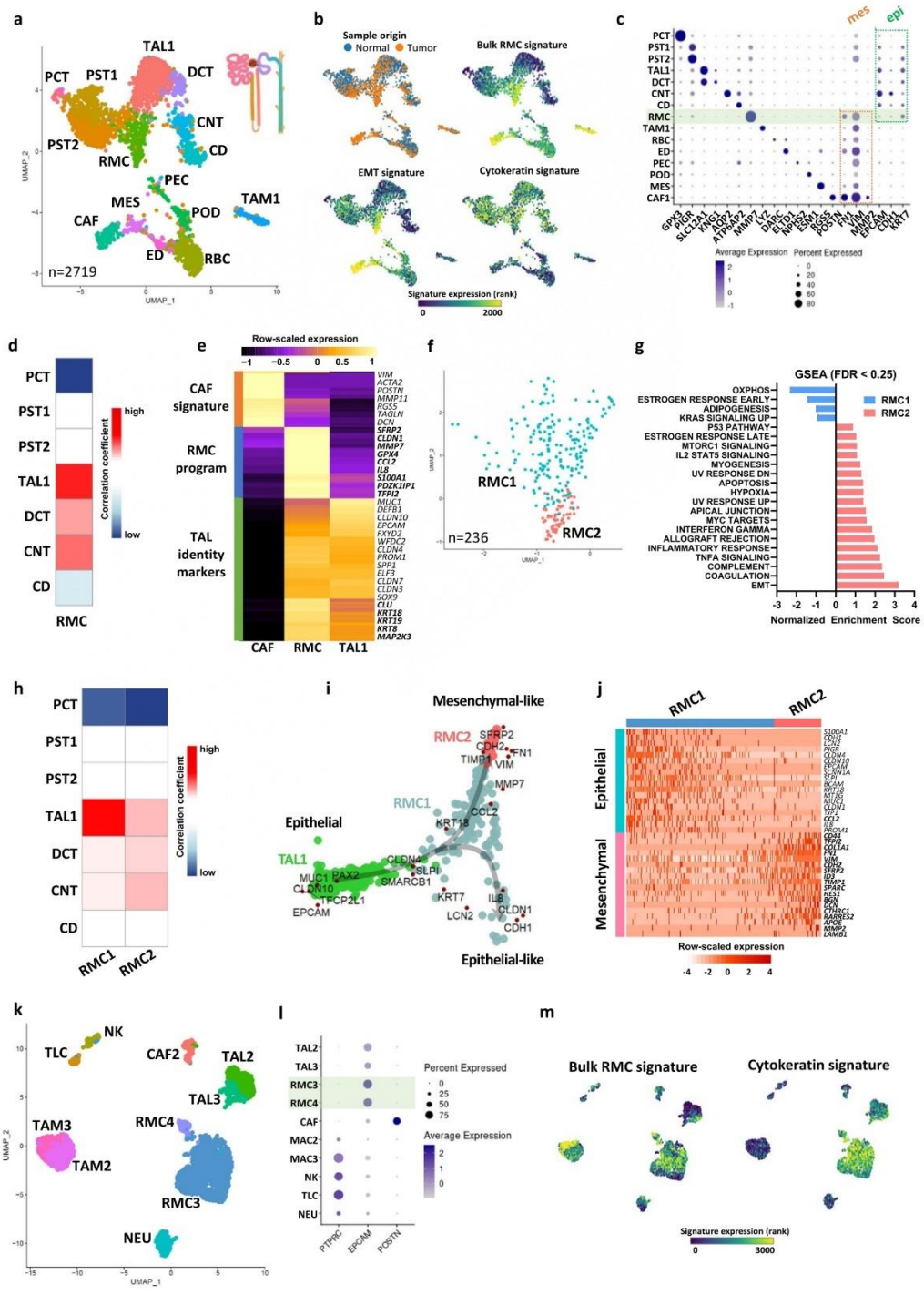
To investigate intra-tumoural heterogeneity, we re-clustered the RMC cells identifying distinct RMC1 and RMC2 subpopulations (Fig. 1f). Gene Set Enrichment Analysis (GSEA) revealed that RMC1 were enriched in oxidative phosphorylation (OXPHOS), whereas RMC2 were enriched in EMT, interferon gamma, inflammatory response and hypoxia (Fig. 1g). Correlation of the RMC1 and RMC2 specific signatures to those of normal tubules revealed that RMC1 partly retained a TAL signature that was reduced in RMC2 (Fig. 1h). These observations were independently confirmed by SWNE trajectory analysis that traced transformation of TAL cells to RMC2 via the RMC1 population with some cells retaining a more epithelial identity (Fig. 1i). This was further supported by separation of RMC cells into a 'stressed' epithelial-like phenotype with higher levels of cytokines (*IL8*, *LCN2*), keratins and epithelial markers such as *CDH1*, *CLDN1* and into RMC2 cells with higher expression of mesenchymal markers such as *SFRP2*, *CDH2* and *FN1*. Thus, this RMC tumour comprised epithelial-like RMC1 cells and mesenchymal-like RMC2 cells (Fig. 1j).

We next analysed a naive RMC sample from a primary nephrectomy of a 16-year-old patient with regional lymph node and adrenal gland metastases (pT4N1M1) at presentation capturing a total of 3372 cells. Following surgery, the patient showed rapid progressive disease under adjuvant MVAC regimen. The patient was also primary resistant to durvalumab-tremelimumab immunotherapy and EZH2 inhibitor Tazemetostat leading to death within one year of diagnosis. Among 3372 captured cells, a large group of RMC cells was identified along with TAMs and other *CD45*-expressing immune cells (Natural killers, neutrophils and T-cells), *POSTN*-expressing CAFs, and an unexpected population of tumour-associated TAL2/3 cells (Fig. 1k–l). Both the RMC and TAL cells, that segregated into two closely located groups on the UMAP plot, expressed *EPCAM* as well as a cytokeratin signature (Fig. 1l–m). The TAL3 population could be distinguished from TAL2 cells by the lowered expression of the *SLC12A1*, *HOXB9* and *PAX8* renal identity markers (Fig. 2a and Supplementary Data 1b). The RMC3 and RMC4 populations were highly similar with the smaller RMC4 cluster displaying an additional G2/M phase cell cycle signature designating them as mitotic RMC3 cells (Fig. 2a). The SWNE trajectory representation of the TAL and RMC populations illustrated the progressive loss of TAL identity markers from the most differentiated TAL2 to TAL3 with some TAL3 cells closely related to the RMC group that retained an epithelial-like signature (Fig. 2b).

Aggregation of the batch corrected data from the two tumours (Fig. 2c), highlighted differences between the TAL and RMC subpopulations illustrated using a collection of epithelial, mesenchymal, endoplasmic reticulum (ER)-related stress genes (Fig. 2d, Fig. S1b). The RMC3/4 cells from the naive tumour had a marked epithelial character compared to intermediate RMC1 cells from the treated tumour, whereas RMC2 cells had the most mesenchymal phenotype (Fig. 2d). GSVA analyses revealed enrichment for cell cycle in RMC4 cells, OXPHOS and apical junction in RMC3 cells, and EMT and interferon gamma response in RMC1 and RMC2 cells (Fig. 2e). SWNE trajectory analyses highlighted the gradient of epithelial to mesenchymal phenotypes of the different populations (Fig. 2f).

We further performed multi-region tumour RNA-seq on a cohort of four patients, for which single region transcriptome sequencing was previously reported along with that of 7 additional cases ([6], designated as the MDACC cohort; and Supplementary Data 2). Overall, we generated an additional 25 bulk RNA-seq from multiple regions of these primary tumours and the corresponding regional lymph nodes as well

Article 1: SMARCB1 regulates a TFCP2L1-MYC transcriptional switch promoting renal medullary carcinoma transformation and ferroptosis resistance



as 3 NATs and analysed intra- and inter-tumour heterogeneity using CIBERSORTx deconvolution to infer their RMC1-3 composition (Fig. S1c, d). For clarity, we did not include the cycling RMC4 signature. Primary tumour sections showed varying composition, some more enriched in the epithelial-like signature, others with epithelial-like and intermediate signatures, and a third group with all 3 signatures. In contrast, the lymph node metastases sections were strongly enriched in the mesenchymal-like signature. These data unravelled intra-tumour heterogeneity in RMC and the importance of tumour cells with a mesenchymal signature to metastatic progression.

We used SCENIC regulon analyses software to identify transcriptional regulatory networks underlying the above signatures[17]. Comparison of the TAL and RMC populations from the treated tumour revealed a transcriptional switch from high *HOXB9* and *TFCP2L1*

Fig. 1 | Single-cell RNA sequencing of treated (A-J) and (K-M) naive RMC tumours. a UMAP plot of the aggregated treated tumour and normal adjacent tissue (NAT) representing the clusters identified by Seurat using a resolution of 1.12. PCT proximal convoluted tubule cells, PST1/2 proximal straight tubule cells 1 and 2, RMC renal medullary carcinoma cells, TAL1 thick ascending tubule cells of Henle's loop, DCT distal convoluted tubule cells, CNT connecting tubule cells, CD collecting duct cells, CAF cancer-associated fibroblasts, MES mesangial cells, ED endothelial cells, RBC red blood cells, PEC parietal epithelial cells, POD podocytes, TAM1 tumour-associated macrophages. **b** UMAP projection of sample origin or selected gene signatures. **c** Dot-plots representing gene markers of each identified cluster in the RMC treated sample. Rectangles regroup clusters according to either mesenchymal or epithelial markers. **d** Clustifyr correlation between RMC cells and renal epithelial tubules transcriptomes. **e** Pseudo-bulk heatmap of 100 RMC-specific and 50 CAF-specific genes using CAF1, RMC and TAL1 clusters as a matrix. **f** UMAP representing RMC subclusters as identified by Seurat using a resolution of 1. **g** GSEA showing enriched "Hallmark gene sets" in RMC1 relative to RMC2 cell clusters. **h** Clustifyr correlation between RMC subclusters and renal epithelial tubules transcriptomes. **i** SWNE trajectory analysis of the treated RMC clusters using a set of selected markers per cluster and assuming TAL1 cells as origin. **j** Heatmap representation of a set of selected EMT genes in the 2 RMC subclusters. **k** UMAP plot of the naive tumour cell clusters as identified by Seurat. RMC3/4: Renal medullary carcinoma cells; TAL2/3: thick ascending tubule cells of Henle's loop; NEU neutrophils, CAF2 cancer-associated fibroblasts, NK natural killers, TLC T-lymphocyte cells, TAM2/3 tumour-associated macrophages. **l** Dot-plots of selected gene markers of immune, epithelial and CAF cells. **m** UMAP projection of the bulk RMC and cytokeratin signatures.

activity in TAL1 cells, to high *MYC*, *HIF1A*, *YY1* and *NFE2L2* activity in RMC cells (Fig. 2g). These data were consistent with the known role of MYC in RMC transformation, whereas *TFCP2L1* is a previously described determinant of the distal portion of the nephron[18]. Top TAL regulons were progressively lost upon transformation into RMC1 and RMC2 populations exemplified by *TFCP2L1*, *PPARGC1A*, perhaps contributing to the OXPHOS signature[19], and *HOXB9*, whereas others like *SOX9* were maintained (Fig. 2h).

Comparable observations were made between the TAL and RMC populations of the naive tumour with loss of *TFCP2L1* activity and gain of *MYC* and *NFE2L2/3* activity (Fig. 2i). Interestingly, while TAL2/3 cells displayed *TFCP2L1* activity they also showed a stress signature with prominent activity of *ATF4*, *XBP1* and *HIF2A*. Moreover, they further showed *YY1* and *MYC* activity, hallmarks of RMC cells. TAL1 cells were derived from NAT, whereas TAL2/3 cells were tightly associated with the RMC cells in the tumour sample and showed a stressed pre-tumoural phenotype with activation of several RMC regulons. Each RMC population displayed a characteristic regulon activity such as cell cycle (*BRCA1*, *E2F4/6*) in RMC4 cells[20,21], epithelial-like (*OVOL2*, *ELF3*) in RMC3 cells[22,23] and mesenchymal-like (*HES1*, *FOSL2*) in RMC2[24,25]. Notably, activity of the *PAX8* renal identity marker was strongly reduced in the RMC1 and RMC2 populations compared to RMC3 (Fig. S1e).

The role of *TFCP2L1* in driving expression of epithelial genes was reinforced by analyses of the Cancer Cell Line Encyclopaedia (CCLE) showing positive correlation between *TFCP2L1* (and also *OVOL2*) and *EPCAM* (Fig. S1f). Similarly, *TFCP2L1* correlated with epithelia markers and anti-correlated with mesenchymal markers (Fig. S1g). In the TCGA chromophobe renal cell carcinoma dataset, originating also from distal tubules, *TFCP2L1* and *MITF* expression correlated with that of *CDH1* (Fig. S1h).

The above data defined an EMT gradient of RMC cells defined by distinct transcriptional signatures also found in patient tumour samples. NAT-derived TAL1 cells were further distinguished from tumour-associated TAL2/3 cells that displayed a stressed, pre-tumoral phenotype in their transcriptional signatures and regulon activities.

Tumour cell state of a patient derived RMC xenograft

We analysed a patient derived xenograft (IC-PDX-132) from an RMC tumour treated with 6 cycles of cisplatin, gemcitabine and bevacizumab that had undergone 4 passages of subcutaneous injections on immunocompromised mice. Around 10,000 cells were captured and the sequences aligned to a human-mouse hybrid genome. A large group of human RMC tumour cells were identified with high expression of *EPCAM* and the bulk RMC signature as well as a group of murine cells corresponding to CAFs and pericytes, TAMs and monocytes, and a smaller number of other immune cells (Fig. S2a–c and Supplementary Data 1c). A third group that we tagged 'LQ' (low-quality) comprised cells with high levels of mitochondrial genes and potential doublets, that were removed from the subsequent analyses.

Re-clustering the RMC cells revealed 4 subpopulations together with some mouse cells of undefined identity that were not further considered (Fig. S2d). The RMC8 cluster showed a strong cell cycle signature and regulon activity designating them as mitotic RMC cells, whereas RMC6 cells displayed high hypoxia and stress-associated regulons such as *ATF4* and *DDIT3* (Fig S2d–f)[26]. RMC5 and RMC7 on the other hand corresponded to epithelial-like and intermediate state cells respectively analogous to the RMC3 and RMC1 cells in the human tumours (Fig. S2e). No distinct highly mesenchymal population was observed, although the mitotic RMC8 cells showed the most dedifferentiated phenotype and highest expression of *FN1* and *CD44*. SCENIC analyses of these populations identified the key *MYC*, *YY1*, and *NFE2L2* regulons in the RMC cells as seen above in primary human tumours (Fig S2f).

These analyses revealed that the RMC PDX comprised principally epithelial-like, intermediate and mitotic RMC cells as well as a subpopulation of hypoxic cells consistent with the idea that angiogenesis could not fully irrigate the rapidly proliferating tumour.

Characterization of the RMC microenvironment

In addition to TAL and RMC cells, scRNA-seq revealed prominent CAF and TAM populations in the RMC tumour microenvironment (TME). Analyses of CAFs from both tumours revealed two populations with either a myofibroblast myCAF signature (CAF1) predominant in the treated tumour or an inflamed iCAF signature (CAF2) in the naive tumour (Fig. S3a). Renal CAFs may arise from pericyte-like mesangial cells[27]. SWNE analyses incorporating NAT-derived mesangial (MES) cells supported the idea they gave rise to the two CAF populations.

Analyses of the TAM population identified TAM1 cells displaying a pro-inflammatory M1 signature (Fig. S3b). In contrast, TAM2 and TAM3 displayed an anti-inflammatory M2 signature with high expression of known M2 markers IL10 and MAF[28], that was strongest in TAM3. SWNE trajectory analysis further confirmed the idea that the TAM2 signature represented an intermediate state between the most polarized TAM1 and TAM3 states.

We then applied the CAF and TAM signatures to the bulk-RNA-seq data from the patient tumour sections as described above. CAF2 cells were detected in all primary and metastases sections, whereas CAF1 were not present in all primary sections and lowly represented in metastases sections (Fig. S3a). Likewise, the TAM2 and TAM3 signatures were detected in a subset of primary and metastases sections, whereas the TAM1 signature was poorly represented in the majority of primary tumour sections, but was highly enriched in the lymph node metastases sections (Fig. S3b).

These analyses showed that the naive tumour and untreated primary patient sections displayed a pro-tumoural, immunosuppressive microenvironment with predominantly iCAFs and M2-type TAMs. However, the MVAC-treated microenvironment was characterized by M1-type TAMs and myCAFs.

Cultured RMC cells recapitulate the EMT gradient

To better define the mechanism by which SMARCB1 loss drives transition from the TFCP2L1-TAL epithelial program to the MYC-driven

Article 1: SMARCB1 regulates a TFCP2L1-MYC transcriptional switch promoting renal medullary carcinoma transformation and ferroptosis resistance

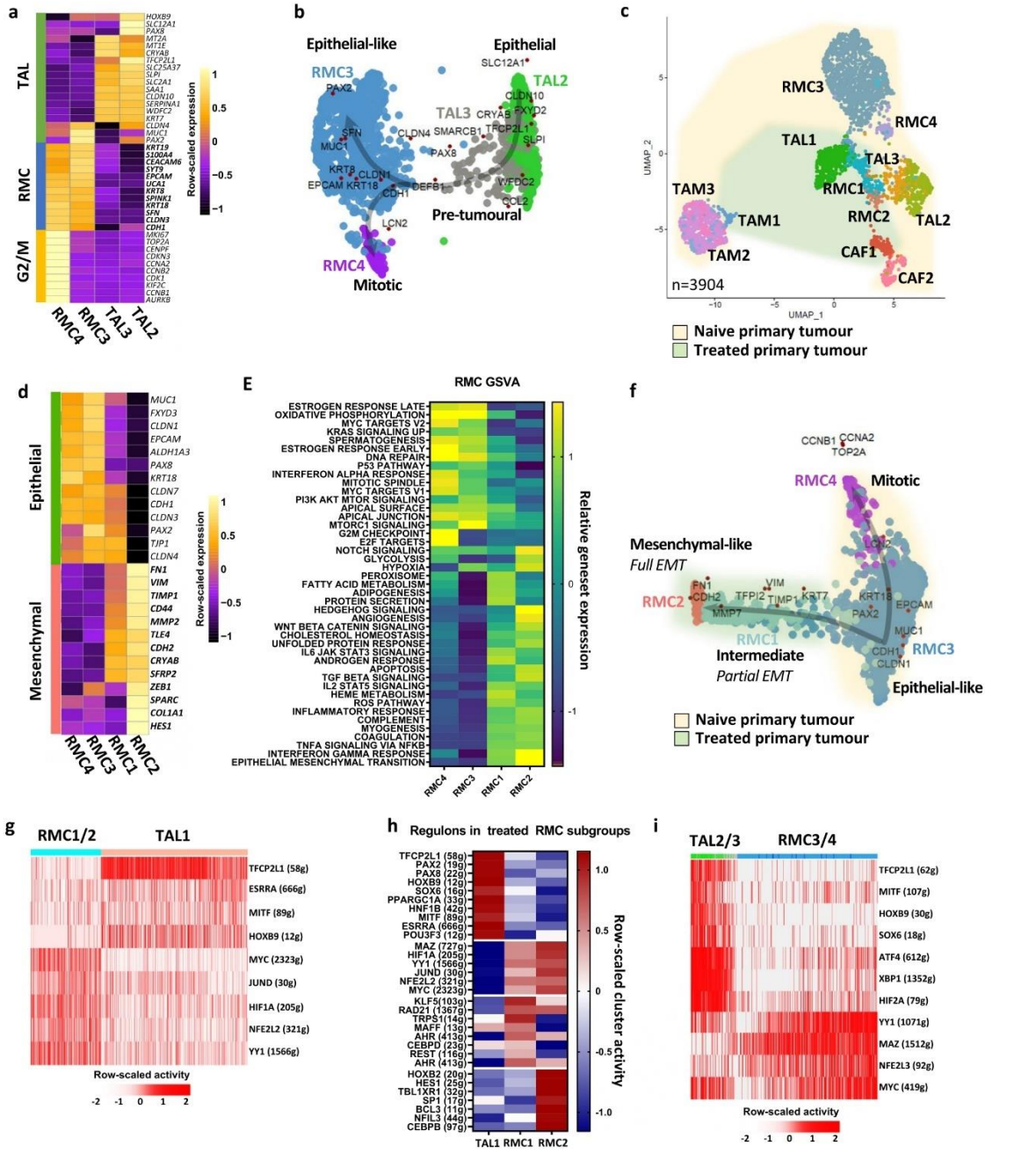


Fig. 2 | Intra-tumoural heterogeneity of RMC. a Pseudo-bulk heatmap of expression of top markers of RMC and TAL clusters. **b** SWNE trajectory analysis of the naive RMC clusters using a set of selected markers per cluster and assuming TAL2 cells as origin. **c** UMAP representing the normalized merge of selected TAL, RMC, CAF and TAM clusters from the treated tumour (green hue) and the naive tumour (yellow hue). **d** Pseudo-bulk heatmap showing a set of known EMT markers in all RMC clusters. Note that as RMC4 were cycling RMC3 cells, they were omitted from the analysis to avoid redundancy. **e** GSVA analysis showing ontologies of indicated RMC clusters. **f** SWNE trajectory analysis of normalized merged RMC clusters from treated and naive tumours using a set of differentially expressed EMT markers. **g** SCENIC analysis of the treated tumour showing top regulons of RMC1/2 and TAL1 cells. Note that in brackets are indicated the number of genes (g) per selected regulon. **h** SCENIC analysis of the treated tumour indicating activities of TAL regulons and RMC1- or RMC2-specific regulons. **i** SCENIC analysis of the naive tumour revealing top regulons of RMC3/4 and TAL2/3 cells.

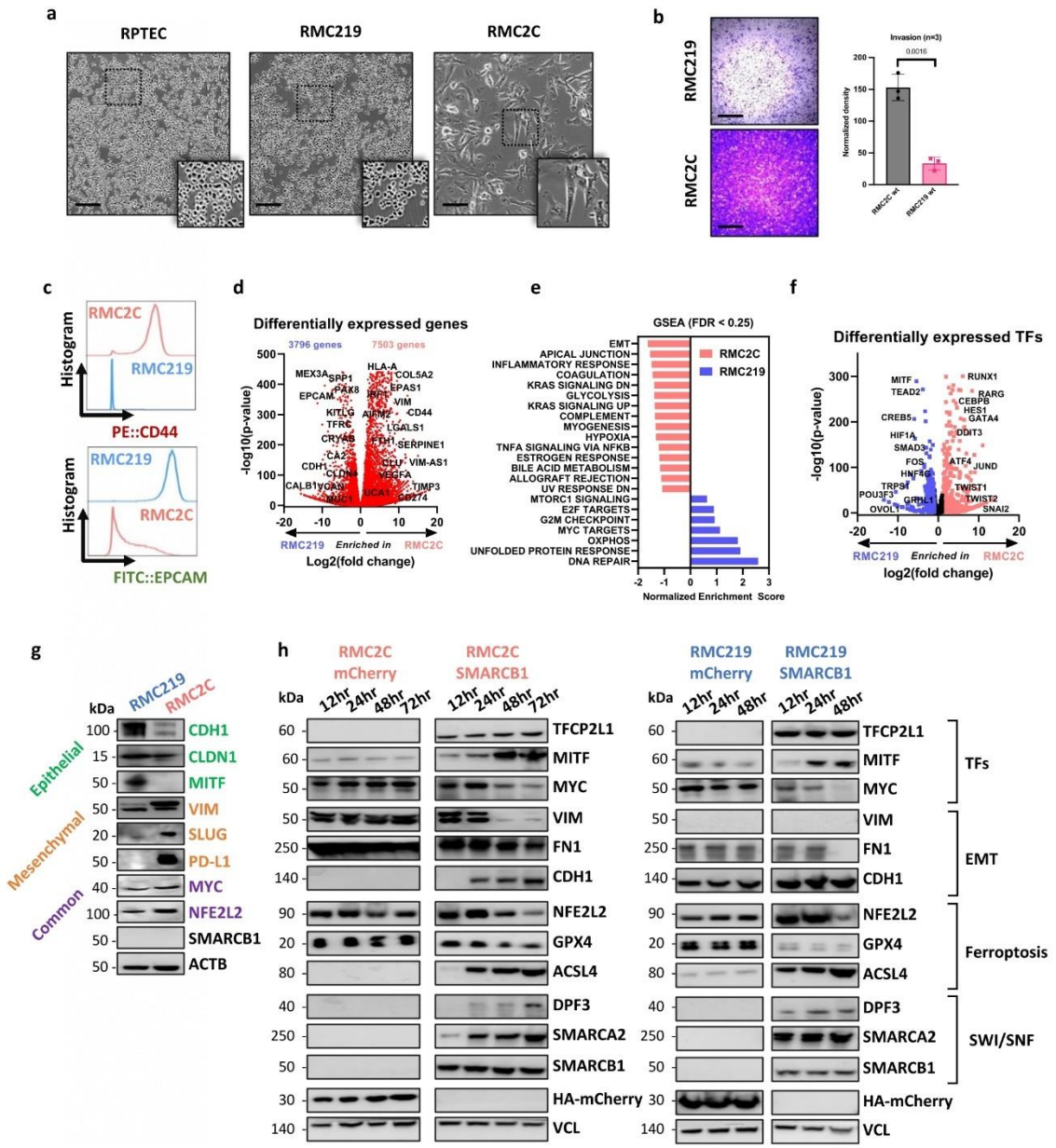


Fig. 3 | Cultured RMC cells recapitulate EMT cell states. a Phase-contrast microscopy at 20X magnification of normal kidney (RPTEC) and tumour cells (RMC219 and RMC2C) showing distinct morphologies of RMC lines. Scale bar: 250 μm. $n = 3$ independent biological replicates. **b.** Brightfield microscopy at ×4 magnification of Boyden chamber matrigel assays using RMC lines (left) and absolute quantification using absorbance of resuspended crystal violet (right). Scale bar: 1000 μm. Biological triplicates are plotted as means ± SD and one-sided unpaired *t* test analyses were calculated by Prism 5; *p* value is indicated. **c** Flow cytometry of membrane protein expression of EPCAM and CD44 in RMC lines. **d** Volcano plot depicting differentially expressed genes using normalized bulk RNA-seq of RMC lines. *P* values were derived using the Wald test and adjusted using Benjamini-Hochberg FDR correction. **e** GSEA using the Hallmarks genesets showing pathways enrichment in respective RMC lines. Note that only pathways with FDR < 0.25 are shown. **f** Volcano plot of differentially expressed 1681 FANTOM5-defined TFs using normalized bulk RNA-seq of RMC lines. *P* values were derived using the Wald test and adjusted using Benjamini-Hochberg FDR correction. **g** Immunoblots detecting the indicated proteins. $n = 3$ independent biological replicates. **h** Immunoblots showing expression of selected proteins upon re-expression of SMARCB1 in RMC2C (left) and RMC219 (right). $n = 3$ independent biological replicates. Source data are provided in the Source Data files.

oncogenic program, we analysed RMC2C and RMC219 cells[6,29]. RMC219 cells displayed a regular rounded morphology similar to primary RPTEC renal epithelial cells (Fig. 3a). RMC2C cells were larger with a more mesenchymal morphology and were much more invasive than the RMC219 cells (Fig. 3a, b). Similarly, flow cytometry indicated that RMC219 cells were EPCAM high, whereas RMC2C cells were CD44 high (Fig. 3c), a marker of RCC aggressiveness[30]. A similar analysis of the UOK360 and UOK353[31] lines by flow cytometry revealed intermediate phenotypes. UOK360 displayed higher EPCAM and lower CD44 than UOK353 and more resembled RMC219 cells (Fig. S4a, b).

Article 1: SMARCB1 regulates a TFCP2L1-MYC transcriptional switch promoting renal medullary carcinoma transformation and ferroptosis resistance

Note however that UOK360 expressed both CD44 and EPCAM discriminating them for the most epithelial RMC219 cells. UOK353 on the other hand, had lower EPCAM, but CD44 levels closer to the RMC2C cells. Moreover, we observed a progressive increase in invasive capacity along the EMT gradient from RMC219-UOK360-UOK353-RMC2C (Fig. S4c). Cultured RMC cells therefore formed an EMT gradient as observed in the scRNA-seq data on the human tumours.

We used RNA-seq to characterize the most epithelial and mesenchymal RMC219 and RMC2C lines identifying an extensive set of differentially expressed genes with preferential expression of epithelial markers in RMC219 cells and mesenchymal markers in RMC2C cells (Fig. 3d). GSEA revealed enrichment of EMT, inflammatory response and hypoxia in RMC2C cells, as seen in the RMC2 tumour population, and enrichment of cell cycle and DNA repair in RMC219 cells (Fig. 3e). Moreover, while OXPHOS was enriched in RMC219, glycolysis was enriched in RMC2C suggesting a metabolic switch upon EMT. *MITF* and *POU3F3*, previously reported determinants of nephron morphogenesis and TAL cell differentiation[32,33], were preferentially expressed in RMC219 cells, whereas EMT-transcription factors like *TWIST1/2* and *SNAI2* were preferentially expressed in RMC2C cells (Fig. 3f). Immunoblot analyses confirmed higher expression of VIM, and SNAI2 in RMC2C and higher expression of CDH1 and MITF in RMC219 cells (Fig. 3g). Both cell lines however showed expression of NFE2L2 and MYC and lacked SMARCB1. These cell lines therefore reproduced epithelial-like and mesenchymal-like phenotypes analogous to those observed in human tumours.

SMARCB1 re-expression in RMC cells represses the oncogenic program

We analysed expression of SWI/SNF subunits in RMC2C cells compared to other SMARCB1-deficient cell lines and HEK293T kidney cells. As expected SMARCB1 was absent from all tumour lines (Fig. S5a, b). The catalytic ATPase subunit SMARCA2 (BRM) was absent in all lines except VA-ES-BJ (epithelioid sarcoma), while SMARCA4 (BRG1) was detected in all lines except CHLA-06-ATRT (rhabdoid tumour). RMC2C cells showed the most important changes in SWI/SNF composition with absence of SMARCD3, ARID2 and lowest expression of DPF3, PBRM1, BRD7 and ARID1A. Although the bulk patient RNA-seq data also comprised signal from CAF and TAM cells, RMC-specific reductions in SMARCA2, and DPF3 expression could still be observed (Fig. S5c).

We engineered RMC2C and RMC219 cells to re-express SMARCB1, or mCherry as control, in a doxycycline (Dox)-dependent manner. SMARCB1 was maximally expressed in both RMC cell lines already 12 h after Dox addition (Fig. 3h). SMARCB1 expression in RMC2C cells was comparable to that seen in HEKT cells, seen in almost all cells of the population and was integrated into SWI/SNF and co-precipitated with BRG1 (Fig. S6a–c). The renewed presence of SMARCB1 induced rapid re-expression of SMARCA2, but slower re-expression of DPF3 (Fig. 3h). Similarly, the TAL-associated TFCP2L1, MITF and CDH1 were also induced, whereas MYC, NFE2L2 and EMT markers VIM and FN1 were down-regulated. Each line showed a similar response, but with faster kinetics in the epithelial RMC219 cells where the oncogenic program was more rapidly repressed and the epithelial program faster induced than in RM2C2 cells. SMARCB1 re-expression therefore reversed key transcriptional changes observed during TAL to RMC transformation.

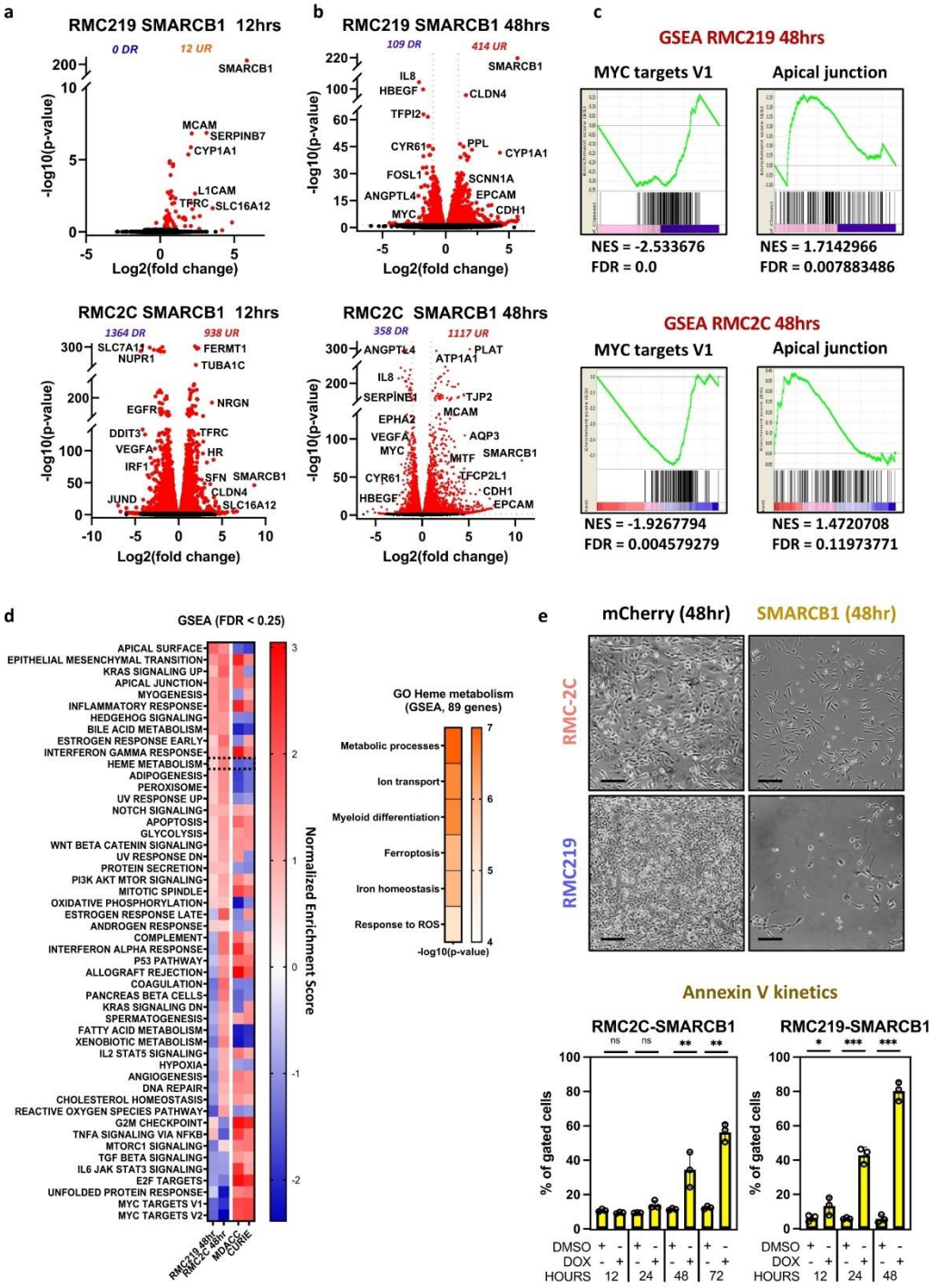
To globally assess gene expression upon SMARCB1 re-expression, we performed RNA-seq in each cell line 12 and 48 h after Dox-treatment. In RMC2C cells, a rapid transcriptional response was seen with 938 down-regulated and 1364 up-regulated genes after 12 h compared to RMC219 cells where only 12 genes were up-regulated over the same period (Fig. 4a, Supplementary Data 3). After 48 h, a larger number of up and down-regulated genes were observed in both cell lines (Fig. 4b and Fig. S6d). Despite the differences in kinetics and numbers of affected genes, GSEA analyses revealed that in both lines, genes down-regulated were involved in oncogenic functions such as cell cycle and proliferation, designated by the GSEA terms MYC or E2F-targets in agreement with accumulation of G1/S phase RMC2C cells 12 and 48 h after Dox treatment (Fig. S6e). Up-regulated genes were designated by epithelial-like program terms such as cell adhesion, apical junction and apical surface (Fig. 4c). Comparison with bulk-RNA-seq from RMC patients relative to their NAT from both MDACC and Institut Curie cohorts showed the opposite profile with genes up-regulated in the SMARCB1-deficient tumours enriched in proliferation, cell cycle and JAK-STAT3 pathway, whereas those down-regulated associated with apical surface (Fig. 4d). Similarly, while OXPHOS was increased upon SMARCB1 expression in cell lines, it was reduced in RMC tumours. RMC cell lines hence reproduce phenotypes and transcriptional signatures seen in RMC tumours whose key features were reversed by SMARCB1 re-expression.

SMARCB1 re-expression in RMC cells induces ferroptotic cell death

SMARCB1 re-expression induced cell death with a 10–20-fold increase in the number of Annexin V-expressing cells (Fig. 4e). RMC219 cells responded rapidly with many dead cells detected by 24 h after Dox addition, whereas death of RMC2C cells was evident at 48 h, but required a longer time to reach higher levels (Fig. 4e). To understand the mechanism of cell death, we examined the gene expression changes and noted that Heme metabolism was amongst the pathways strongly up-regulated upon SMARCB1 re-expression and down-regulated in RMC patients (Fig. 4d). The Heme metabolism GSEA term covers iron homoeostasis, response to reactive oxygen species and ferroptosis (Fig. 4d, right panel). Following SMARCB1 re-expression, key anti-ferroptosis genes such as *NFE2L2*, *NUPR1* and their target *GPX4*, a well-characterized inhibitor of lipid peroxidation[34] were down-regulated in both lines (Figs. 5a and 3h). On the other hand, Transferrin (*TF*) and transferrin receptor (*TFRC*) regulating iron uptake were both rapidly induced in RMC219 and RMC2C cells (Figs. 4a, b, 5a and Fig. S6c). Following these acute events, at 48 h we observed increased expression of a subset of genes involved in lipid peroxidation namely *DPP4*, *LOX*, *LPCAT* paralogs and *ACSL4* (Figs. 5a and 3h). These data suggested that SMARCB1 re-expression induced an acute increase in iron uptake followed by increased lipid peroxidation and ferroptosis.

Complementary observations were made from our scRNA-seq dataset where SCENIC showed that RMC tumour cells were characterized by the activation of the NFE2L2/3 regulon a major regulator of ferroptosis (Figs. 2d, i)[35,36]. Consequently, expression of *NFE2L2*, *GPX4* and other anti-ferroptosis genes was upregulated in RMC cells from the MVAC-treated tumour compared to TAL cells, whereas many pro-ferroptosis genes were higher expressed in TAL cells (Fig. 5b). Similarly, in the naive tumour, *GPX4* and anti-ferroptosis genes were upregulated in RMC compared to TAL cells (Fig. 5b). RMC tumours further showed staining with 4-Hydroxynonenal (4-HNE) antibodies compared to the surrounding stromal cells confirming their propensity to undergo lipid peroxidation (Fig. S7a). Moreover, in agreement with their pre-tumoural phenotype, the RMC-associated TAL3 cells showed up-regulated expression of anti-ferroptosis genes and down-regulated expression of the pro-ferroptosis genes compared to the TAL2 cells. Activation of the *MYC* and *NFE2L2/3* regulons in these cells was therefore accompanied by activation of the ferroptosis resistance program.

We next assessed if SMARCB1 re-expression and increased expression of the lipid peroxidation genes translated into an elevation of lipid ROS assessed using BODIPY-C11-based flow cytometry (Fig. 5c and Fig. S8). SMARCB1 re-expression induced a strong increase of lipid ROS in both lines not seen in mCherry control lines. High lipid ROS was associated with increased AnnexinV-positive cells. Importantly, the increase in lipid ROS and in Annexin-V positive cells were both impaired by ferrostatin-1, a selective ferroptosis inhibitor (Fig. 5c). In contrast, SMARCB1 expression did not induce the activated Caspase 3



apoptosis marker unlike Campothecin treatment. To further confirm ferroptotic cell death, we treated RMC cells with the GPX4 inhibitor RSL3. The RMC cells had IC50 values 2–4 times lower than other RT cell lines and more than 100-fold lower than control RPTEC or HEKT cells (Fig. 5d). We additionally assessed the ability of the pan-caspase inhibitor zVAD-fmk or the necroptosis inhibitor necrostatin-1 (nec1) to inhibit SMARCB1 or RSL3-induced cell death. Reduced cell viability after Dox-induced SMARCB1 expression was rescued by ferrostatin-1 and by nec1, consistent with the previously reported ability of higher concentrations of nec1 to rescue ferroptosis in other tumour cell lines[37,38], but not by zVAD-fmk (Fig. S7b). Similarly, cell viability in presence of RSL3 was also rescued by ferrostatin-1 and Nec1, but not zVAD-fmk (Fig. S7b). Flow cytometry confirmed that RSL3-induced cell death was rescued by high but not low concentrations of Nec1

Article 1: SMARCB1 regulates a TFCP2L1-MYC transcriptional switch promoting renal medullary carcinoma transformation and ferroptosis resistance

Fig. 4 | Tumour-suppressor function of SMARCB1. a Volcano plot revealing up- and down-regulated genes at 12 h after SMARCB1 re-expression in RMC lines. *P* values were derived using the Wald test and adjusted using Benjamini-Hochberg FDR correction. **b** Volcano plot revealing up- and down-regulated genes at 48 h after SMARCB1 re-expression in RMC lines. *P* values were derived using the Wald test and adjusted using Benjamini-Hochberg FDR correction. **c** GSEA showing top up- and down-regulated pathways upon SMARCB1 re-expression (48 h) with similar ontologies observed in both lines. **d** Integrative heatmap showing GSEA Hallmarks enrichments (left panel) in SMARCB1 re-expressing RMC lines and 2 cohorts of RMC primary tumours (MDACC: $n = 11$; Curie: $n = 5$) and Metascape ontology analysis of genes constituting the GSEA "Heme metabolism" term (right panel). FDR values were derived by GSEA using permutation and Benjamini-Hochberg correction. **e** Phase-contrast microscopy at ×10 magnification of RMC lines 48 h after re-expression of either SMARCB1 or mCHERRY control. Scale bar: 500 μm. (upper panel) Quantification of cell death in RMC lines at selected time-points upon SMARCB1 re-expression, as assessed by flow cytometry (lower panel). Note that the % of cells staining positive for either ANXA5 or propidium iodide were tagged as "dead". The remaining unstained cells were tagged "viable". Biological triplicates are plotted as means ± SD and one-sided unpaired *t* test analyses were performed by Prism 5 by comparing matched time-points: *p* values: ns= $p > 0.05$; *=$p < 0.05$; **=$p < 0.01$; ***=$p < 0.001$, RMC2C: *p* values 0.074, 0.082, 0.008 0.00006. RMC219: *p* values 0.046, 0.00008, 0.00001. Source data are provided as a Source Data files.

(Fig. S7c). As ferristatin does not rescue other forms of death[34], these data support the observation that SMARCB1 expression induced ferroptotic cell death. Moreover, further evidence for ferroptosis came from immunofluorescence (Fig. S6c) showing not only that TFRC was rapidly induced by SMARCB1 re-expression, but also that while it was located in the cytoplasm in most RMC2C2 cells at 24 h, there were already some small rounded dying cells where TFRC was relocated to the plasma membrane, hallmarks of ferroptosis[39]. TFRC located to the plasma membrane in almost all RMC219 cells at the same stage consistent with the observation that these cells undergo very rapid ferroptosis. These results confirmed that RMC cells were highly sensitive to GPX4 inhibition and that cell death was due to ferroptosis.

IFNg, secreted by the immune microenvironment in tumours in situ, induces tumour cell dedifferentiation and ferroptotic cell death in melanoma[40,41]. IFNg treatment of RMC219 and RMC2C resulted in durable expression of PDL1, STAT1 and IRF1 and of mesenchymal markers JUN and ZEB1, induced in RMC219 cells and up-regulated in RMC2C cells (Fig. 5e, f). In contrast, NFE2L2 expression was reduced. IFNg treatment induced death of RMC2C cells between 48 and 72 h, whereas death of RMC219 cells required 72 h (Fig. 5g). Importantly, treatment with ferrostatin 1 diminished the IFNg-induced cell death showing that it involved ferroptosis (Fig. 5h), while as control no induced cell death was seen with HEK293T.

These results revealed that TAL cells were characterized by a ferroptosis sensitivity program that was progressively replaced in pre-tumoural TAL3 cells, in the RMC tumour populations and in RMC cell lines by a NFE2L2 and GPX4-high ferroptosis resistance program. This process was reversed by SMARCB1 re-expression that down-regulated NFE2L2 and GPX4 or by IFNg treatment leading to cell death by ferroptosis.

SMARCB1 re-expression promotes genomic SWI/SNF re-localization to enhancers with TFCP2L1 motifs

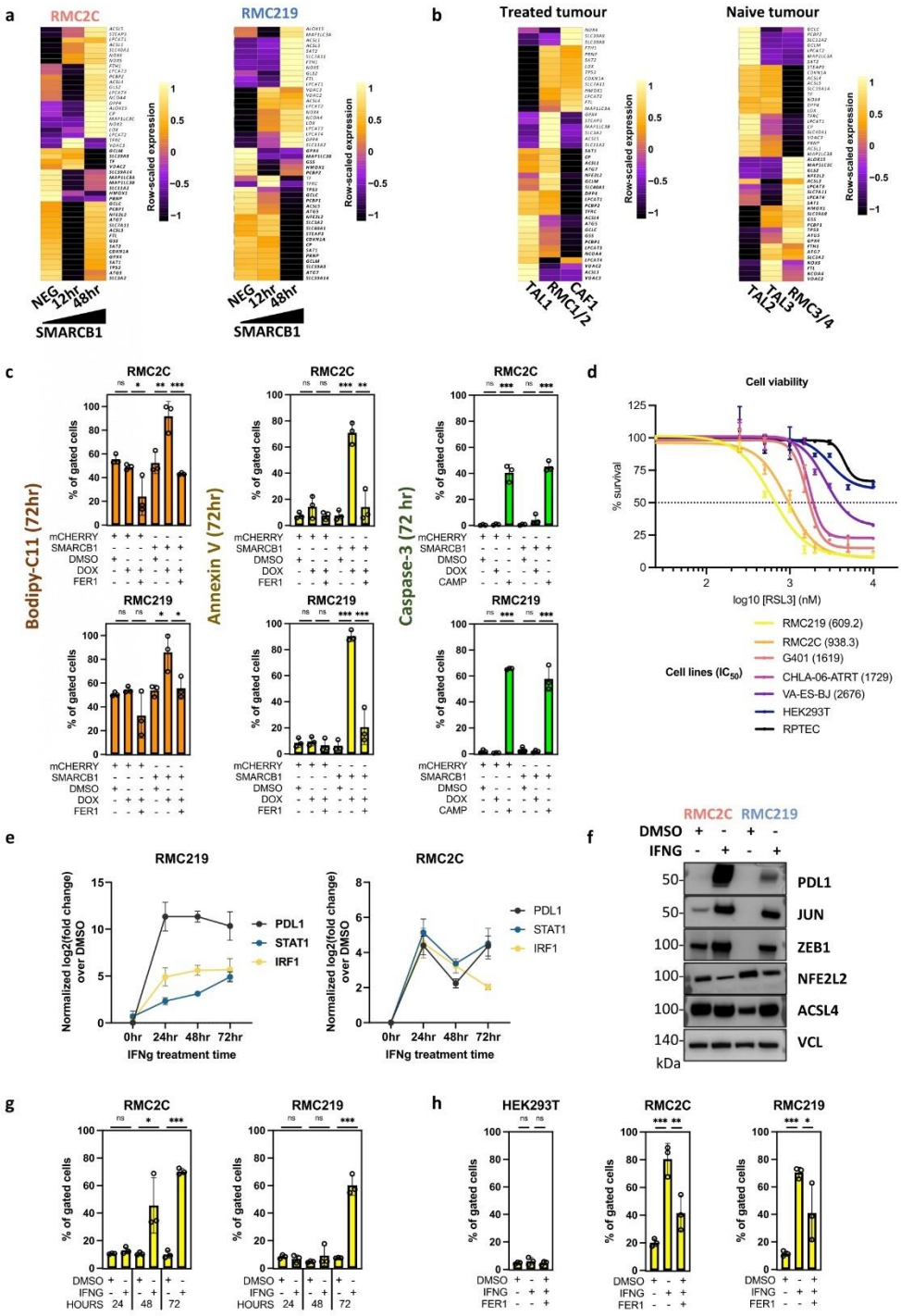
To investigate the consequences of SMARCB1 re-expression on SWI/SNF localization and the epigenome of RMC2C cells, we performed BRG1 and H3K27ac ChIP-seq 48 h after Dox treatment of SMARCB1 or control mCherry expressing cells. RMC219 cells could not be used due to the rapid cell death upon SMARCB1 expression. SMARCB1 re-expression increased the overall number of H3K27ac peaks, but had little impact on their relative genomic distribution with similar fractions of sites at transcription start sites (TSS) and other genomic regions (Fig. S9a, b). However, comparison of read density at more than 46000 non-redundant H3K27ac sites revealed a gain of sites located distal to the TSS following SMARCB1 re-expression (cluster G2, Fig. S9c), whereas only a minor change was seen at the TSS. A fraction of gained peaks were extended regions reminiscent of super-enhancers (SE) known to regulate genes involved in critical aspects of lineage identity or oncogenic transformation[42,43]. While a large number of H3K27ac-marked SEs and their associated genes were shared between the mCherry and SMARCB1 expressing cells, 240 SE-associated genes were specific to the mCherry line and associated with a variety of functions notably DNA repair and cell cycle (Fig. S9f). More strikingly, 330 SE-associated genes specific to SMARCB1 expressing cells were associated with kidney epithelium development and differentiation as well as cell polarity and junction (Fig. S9g).

SMARCB1 re-expression also modified BRG1 genomic occupancy with a loss mainly at the TSS (H4, Fig. S9d), but a gain at distal sites (H8, Fig. S9d). Integration of BRG1 and H3K27ac read density profiles at more than 40,000 non-redundant co-occupied sites identified those with concomitant gain of H3K27ac and BRG1 following SMARCB1 re-expression (A3, Fig. 6a) predominantly located distal to the TSS (C2). In contrast, cluster A2 defined sites with reduced BRG1 predominantly located at the TSS (A2/B1) with a smaller set at distal sites (C1). Correlation with RNA-seq data indicated that genes associated with cluster A3/C2 showed increased expression following SMARCB1 re-expression (Fig. 6a). RSAT analyses revealed a strong enrichment for TFCP2L1, HOXB9, and MITF binding motifs at the distal gained A2/C3 sites (Fig. 6b). Moreover, ontology analyses of the nearest genes to the A3/C3 sites showed enrichment in differentiation, cell adhesion and kidney epithelium development (Fig. 6c). SMARCB1-dependent BRG1 recruitment and H3K27ac modification was exemplified at intragenic regulatory elements at the MITF locus, at a putative regulatory element at the 3′ end of the TFCP2L1 gene, or at intragenic regulatory regions of the SYT13 and DOCK1 genes, that were all re-activated following SMARCB1 re-expression (Fig. S9h). SMARCB1 re-expression therefore led not only to re-expression of TFCP2L1 and MITF, but also re-localization of BRG1 to putative H3K27ac marked distal enhancers and super-enhancers associated with the epithelial gene expression program.

We additionally performed Cut&Tag experiments to profile BRG1 and SMARCB1 genomic localization 24 h after Dox treatment. While no SMARCB1 signal was seen in control mCherry cells, strong SMARCB1 occupancy was seen following its Dox-induced expression (Fig. S10a). At a subset of sites, low BRG1 binding and H3K27ac was seen in absence of SMARCB1 (cluster A1), whereas at the remainder BRG1 and H3K27ac were seen only in presence of SMARCB1 (cluster A2). When SMARCB1 occupancy was examined at the 10983 distal sites observed at 48 h (Fig. 6a), de novo recruitment of SMARCB1, BRG1 and marking by H3K27ac was observed at 24 h (Fig. S10b, clusters B2 and B3). Rapid de novo SMARCB1 recruitment along with BRG1 and H3K27ac was exemplified at the MITF, TFCP2L1, SYT13 and DOCK 1 loci as shown above (Fig. S9h). Moreover, in accordance with the strong enrichment for TFCP2L1 binding motifs at these sites (Fig. 6b), TFCP2L1 co-precipitated with BRG1 from extracts of Dox-treated cells (Fig. S10c). Together these results showed that upon its re-expression, SMARCB1 integrated the SWI/SNF complex that interacted with TFCP2L1 and was rapidly recruited to the H3K27ac-marked regulatory elements associated with epithelial genes.

As mentioned above, TFRC was rapidly induced 12 h after SMARCB1 re-expression. The *TFRC* promoter was strongly marked by H2K27ac in both the mCherry control and 24 h after SMARCB1 re-expression (Fig. S10d). Moreover, BRG1 and SMARCB1 were recruited at 24 h. TFRC therefore behaved as an 'immediate-early' gene whose promoter was pre-marked with H3K27ac, but whose activation was associated with rapid BRG1 and SMARCB1 recruitment. This contrasts with epithelial program genes whose activation was slower and where

Article https://doi.org/10.1038/s41467-023-38472-y



both BRG1/SMARCB1 recruitment and H3K27ac modification occurred de novo.

SMARCB1 re-expression remodels MYC genomic binding

It has been reported that SMARCB1 interacts directly with MYC to antagonize its DNA binding and genomic occupancy in RT cells[44,45]. To address this in RMC cells, we performed MYC ChIP-seq in SMARCB1-expressing and mCherry control cells 48 h after Dox addition. We identified 54,786 non-redundant MYC sites, a much larger number than previously observed[44]. All MYC-bound sites in G401 RT cells, that were predominantly located close to the TSS, were occupied also in RMC2C cells (Fig. S11a, b). For example, MYC sites commonly bound in G401, RMC2C2 and in the Hela ENCODE data sets were observed at the *NCL* and *CDK4* loci (Fig. S11b).

Article 1: SMARCB1 regulates a TFCP2L1-MYC transcriptional switch promoting renal medullary carcinoma transformation and ferroptosis resistance

Fig. 5 | SMARCB1 regulates ferroptosis. a Heatmap showing the KEGG ferroptosis gene signature in SMARCB1 re-expressing RMC2C (left) and RMC219 (right) cells. **b** Heatmap showing expression of the ferroptosis gene signature in RMC and TAL clusters. **c** Flow cytometry quantification of Bodipy-C11, ANXA5 and cleaved CASP3 at 72 h in SMARCB1 or mCHERRY expressing cells and using either Ferrostatin-1 (Fer1) or camptothecin (CAMP) as controls. Biological triplicates are plotted as means ± SD and one-sided unpaired *t* test analyses were performed by Prism 5, ns=$p > 0.05$; *=$p < 0.05$; **=$p < 0.01$; ***=$p < 0.001$ and ****=$p < 0.0001$. *P* values: upper left panel: 0.076, 0.027 0.005, 0.001; lower left: 0.05 0.06, 0.01, 0.02; upper centre panel: 0.12, 0.13, 0.0001 0.001; lower centre panel: 0.37, 0.21 0.000007 0.0004; upper right panel: 0.16, 0.0002, 0.09, 0.0001; lower right panel: 0.09, 2.27 E−09, 0.18 0. 0002. **d.** Cell viability (IC50) upon increasing concentrations of RSL3, a class II ferroptosis inducer. Biological triplicates are plotted as means ± SEM. **e** Gene expression changes of known IFNg downstream targets upon treatment of RMC lines with 10 ng/mL recombinant human IFNg. Biological triplicates are plotted as means ± SEM. **f** Immunoblots showing expression of selected EMT and ferroptosis markers in RMC lines treated either with IFNg or DMSO vehicle control. $n = 3$ independent biological replicates. Molecular mass markers in kDa are indicated. **g** Cell death quantified by flow cytometry using annexin-V in RMC lines. Biological triplicates are plotted as means ± SD and one-sided unpaired *t* test analyses were performed by Prism 5, ns=$p > 0.05$; *=$p < 0.05$; **=$p < 0.01$; ***=$p < 0.001$. *P* values: left panel: 0,09 0,01 3,72E-06; right panel: 0.22, 0.17, 0.0001. **h** Flow cytometry-based quantification of cell death at 72 h upon treatment with IFNg alone, IFNg with Fer1 or DMSO in RMC lines and normal kidney cells as control. Represented values are the mean of 3 biological replicates as means ± SD and unpaired *t* test analyses were performed with Prism5 by comparing conditions to matched DMSO. *P* values: ns=$p > 0.05$; *=$p < 0.05$; **=$p < 0.01$; ***=$p < 0.001$ and ****=$p < 0.0001$. *P* values: left panel: 0.21 0.23; centre panel: 0.0004 0.008; right panel 9.98 E−06, 0.037. Source data are provided as a Source Data files.

In keeping with reduced MYC expression, around 50% fewer peaks were observed in SMARCB1 expressing cells where its occupancy was remodelled with a relative re-localization to the TSS that increased from 24 to 41% of the detected peaks (Fig. S9a, b). Read density profiles at the non-redundant MYC sites identified those with gained (I2/I8, Fig. S9e) or diminished (I3/I9) occupancy located at both TSS proximal and distal regions. Notably, integration with BRG1 and H3K27ac datasets revealed that MYC occupancy was increased at TSS proximal sites marked by H3K27ac, but characterized by diminished BRG1 occupancy (D1/E1, Fig. 6d). In contrast, a large set of distal located sites were lost upon SMARCB1 re-expression (D4/F3) with a smaller number showing increased occupancy (D3/F2). Global analyses confirmed that BRG1 flanking a subset of MYC bound sites in the mCherry control cells was diminished following SMARCB1 re-expression, whereas H3K27ac was unchanged (Fig. S9c). RSAT analysis of the top 1000 MYC peaks confirmed a strong enrichment of the cognate E-box motif (Fig. 6e). MYC-binding motifs were also strongly enriched at the D1-D4 sub-clusters, together with MAZ at D1 sites, whose activity was associated with TAL transformation (Fig. 2g–i and S11d).

As shown above, the term 'MYC targets' was a prominent hallmark of genes down-regulated by SMARCB1 re-expression. We determined the % of genes in the GSEA hallmark gene sets overlapping with those associated with each MYC sub-cluster. Genes associated with D1 sites were strongly enriched in MYC targets, mitotic spindle, mTOR, E2F. DNA repair and G2M hallmark signatures (Fig. S12a). Genes associated with D4 also displayed a similar, yet lower, enrichment in many of these pathways. Correspondingly, genes associated with D1 and D4 showed global down-regulation (Fig. S12b), whereas those associated with D2 and D3 showed up-regulated expression. Thus, many genes associated with oncogenic transformation and down-regulated by SMARCB1 re-expression were associated with a gain of promoter-proximal MYC, but strongly reduced BRG1 binding.

A similar analysis of BRG1 sub-clusters, showed genes associated with A2 were strongly enriched in the above oncogenic-associated hallmarks (Fig. S12c). In contrast, A3 sites with strongly gained BRG1 binding were enriched in genes associated with apical junction/surface and kidney morphogenesis hallmarks, consistent with re-activation of an epithelium program. We used ROSE to identify MYC-H3K27ac-marked or BRG1-H3K27ac-marked SEs in control and SMARCB1-expressing cells (Fig. S12d, e). The ontology of the SE-associated genes was consistent with a switch from MYC/BRG1 driving proliferation and oncogenesis in absence of SMARCB1 to TFCP2L1/BRG1 driving an epithelium program in presence of SMARCB1.

To better understand the paradoxical observation that MYC binding increases at down-regulated oncogenic genes, we looked more closely at the large set of diminished D4 sites associated with similar ontology terms to D1. Re-clustering of D4 identified a small number (J1, Fig. S12f) of promoter-proximal sites associated with H3K27ac and a large majority of distal sites (J2, Fig. S12f). Strikingly, a large number of genes were commonly associated with both clusters (Fig. S12g). Hence many genes of the oncogenic program had both promoter-proximal and distal MYC sites showing increased and decreased occupancy, respectively. Importantly, the D4 sites were enriched in binding motifs for HIF1A and SNAI1 in agreement with coordinate activation of MYC, HIF1A and EMT programs in RMC. Loss of MYC at the D4 sites upon SMARCB1 expression was therefore consistent with their role in driving transformation.

Overall, these results showed that SMARCB1 re-expression did not repress MYC genomic occupancy, but rather remodelled its binding profile in a manner suggesting that altered enhancer-promoter communications and loss of promoter-proximal BRG1 binding underlie reduced expression of the proliferation/oncogenic program.

Discussion

Oncogenic transformation of TAL cells into RMC

Here we integrate transcriptomic data from RMC patients with gain and loss of SMARCB1 function in cell-based models to decipher the mechanism of a transcriptional switch driving oncogenic transformation and ferroptosis resistance of TAL epithelial cells.

ScRNA-seq analyses of RMC cells compared to NAT identified TAL cells as RMC cell-of-origin. TAL cells were marked by strong activity of *TFCP2L1*, *HOXB9* and *MITF* transcription factors associated with the epithelial expression program. TAL transformation was characterized by loss of expression and activity of these factors, but gain of *MYC* and *NFE2L2* that drive proliferation and ferroptosis resistance. Further evidence for this series of oncogenic events came from the fortuitous capture of tumour-associated TAL2/3 cells that displayed a pre-transformed state retaining *TFCP2L1* activity, while at the same time showing *MYC* and *YY1* activity accompanied by a hypoxia and stress signature.

TAL transformation generated an epithelial-mesenchymal gradient of RMC tumour cells that was reproduced by RMC219, UOK360, UOK353 and RMC2C cell lines. Mesenchymal-like cells were observed in the treated tumour and the mesenchymal transcriptional signature was present in primary tumours from naive patients and was predominant in the lymph nodes. Thus, de-differentiation into this mesenchymal state is not specific to drug-treated tumours, but appears to be an intrinsic feature of RMC tumours that likely contributes to their metastatic spread.

SMARCB1 re-expression in RMC2C cells provided experimental mechanistic support for the above model of TAL-RMC transformation. SMARCB1 expression reactivated *TFCP2L1*, *HOXB9* and *MITF* expression and promoted BRG1 re-localization to enhancers and super-enhancers driving expression of an epithelial expression program that were de novo marked by H3K27ac and enriched in binding motifs for these factors (Fig. 6f). The lack of ChIP-grade TFCP2L1 and MITF antibodies did not allow us to directly confirm their presence at these enhancers. However, we previously showed that MITF interacts with

Article 1: SMARCB1 regulates a TFCP2L1-MYC transcriptional switch promoting renal medullary carcinoma transformation and ferroptosis resistance

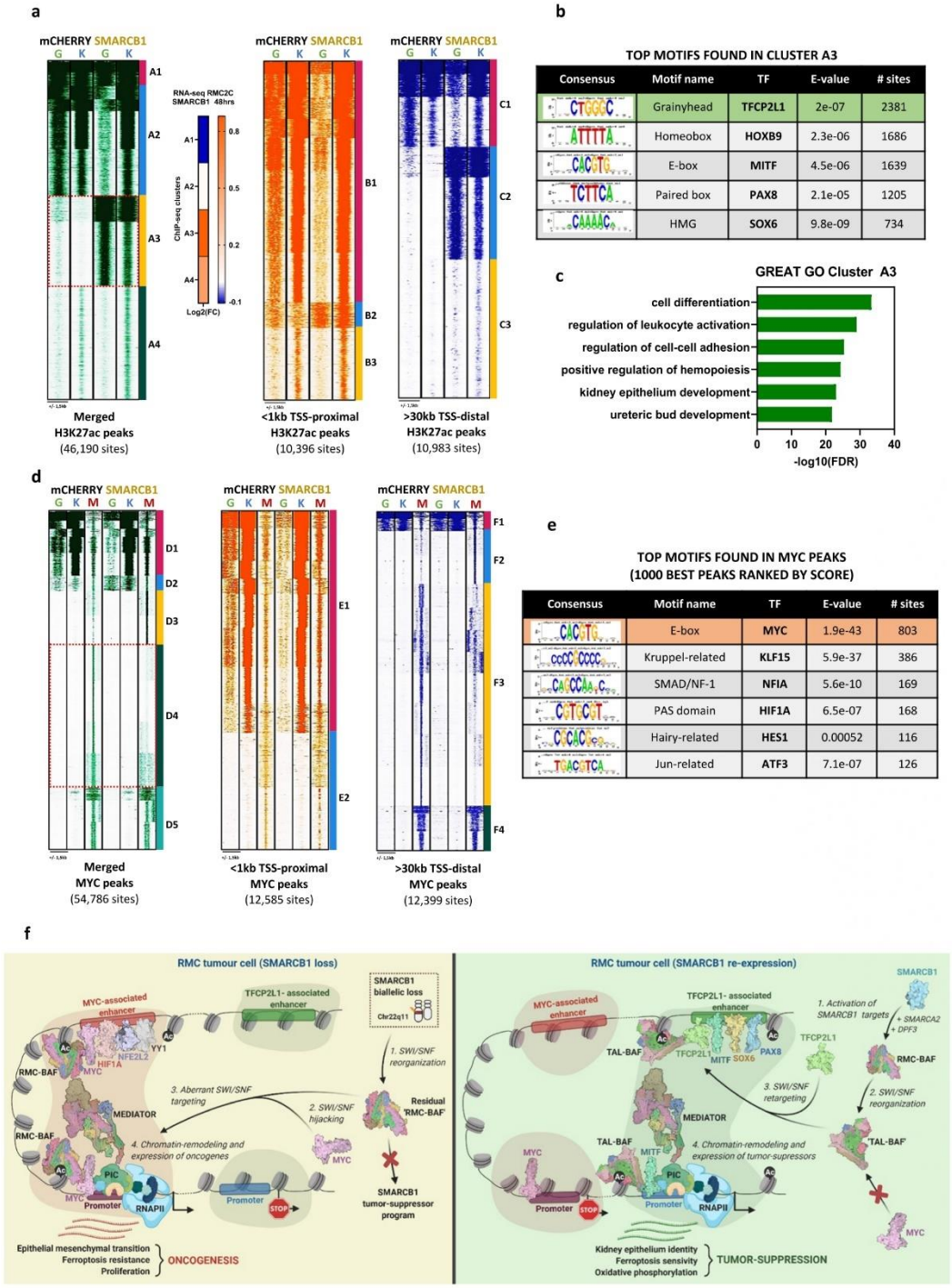


Fig. 6 | SMARCB1 retargets SWI/SNF complexes to enhancers bearing TFCP2L1 motifs. a Read density maps showing genome localization BRG1 (G) and H3K27ac (K) in RMC2C cells expressing either SMARCB1 or mCHERRY using as a reference all merged H3K27ac sites (1st panel), all TSS-proximal H3K27ac sites (3rd panel) and all TSS-distal H3K27ac sites (4th panel). Expression changes for genes associated with BRG1/H3K27ac- clusters following SMARCB1 re-expression are shown in the 2nd panel. **b** RSAT-based motif enrichment analysis using A3 sites ranked by number of sites. **c** Ontology analysis of genes associated with A3 as annotated by GREAT. **d** Read density maps showing genome localization of BRG1 (G), H3K27ac (K) and MYC (M) in RMC2C cells expressing either SMARCB1 or mCHERRY using as a reference all merged MYC sites (1st panel), all TSS-proximal MYC sites (2nd panel) and all TSS-distal MYC sites (3rd panel). **e** RSAT-based motif enrichment analysis using one thousand best MYC peaks ranked (by peak score). **f** Working model of the oncogenic and SMARCB1 tumour-suppressor events in RMC. Created with BioRender.com.

Article 1: SMARCB1 regulates a TFCP2L1-MYC transcriptional switch promoting renal medullary carcinoma transformation and ferroptosis resistance

SWI/SNF and actively recruits BRG1 to melanoma-cell promoters and enhancers[46] and here we showed that TFCP2L1 also co-precipitated with SWI/SNF. In contrast, SMARCB1 re-expression led to reduced levels of MYC and NFE2L2. Genomic profiling revealed a remodelling of MYC genomic binding with sites showing both gained or reduced occupancy. Paradoxically, while MYC binding increased at the proximal promoters of genes involved in oncogenesis, it was lost at sites distal to these genes. Although there are clear limitations in assigning distal binding sites to regulation of a given gene, a large set of genes showed increased MYC binding at the promoter and diminished binding at distal sites suggesting the importance of enhancer-promoter communication in their activation. More importantly however, BRG1 occupancy was strongly reduced at these promoters showing that MYC cooperated with SWI/SNF lacking SMARCB1 to activate the oncogenic program and that BRG1 eviction and not MYC loss reversed the oncogenic process.

Integrating patient and in cellulo-derived data converged to show that pre-tumoral TAL2/3 cells displayed a hypoxia/stress state activating *MYC* and *NFE2L2* to drive ferroptosis resistance allowing survival under conditions favourable to SMARCB1 loss (Fig. 6f). Subsequently, SMARCB1 loss led to BRG1 recruitment at promoters of MYC occupied oncogenic genes and inhibition of the *TFCP2L1/HOXB9/MITF*-driven TAL epithelial program. In RMC cells, SWI/SNF lacking SMARCB1 cooperates with MYC to drive the oncogenic program, whereas SMARCB1-containing SWI/SNF is evicted from MYC-driven oncogenic promoters and re-located to enhancers driving the TAL epithelial program.

Distinct cells-of-origin and oncogenic mechanisms in RMC and RT

The above observations highlight major differences with previous studies on RT cells. In G401 RT cells, SMARCB1 antagonized MYC DNA binding and chromatin occupancy[44]. In contrast, in RMC cells, antagonism translated not as a loss of MYC binding, but eviction of SMARCB1-containing SWI/SNF from MYC occupied promoters and reduced oncogenic gene expression. It has been reported that BRD9-containing non-canonical (nc)BAF plays a critical role in driving the oncogenic state in SMARCB1-deficient RT[47,48]. NcBAF strongly colocalized with CTCF although other enriched transcription factor motifs were also identified. However, MYC was not amongst the strongly enriched motifs in either study. This leads to the paradoxical observation that despite the essential role of BRD9/ncBAF in driving the transformed state, it is not enriched at MYC-bound sites, whereas depletion studies in RT cells[44] and our current data revealed MYC as the essential oncogenic driver. Given the strong association of ncBAF with CTCF and not MYC, it is unlikely that the BRG1 seen at the MYC promoters in RMC cells corresponds to ncBAF. Moreover, CTCF motifs were not enriched at the promoter sites where BRG1 was evicted, but were present at distal sites, where no BRG1 was detected. Thus, the role of ncBAF in RMC remains to be determined.

In RT cells, SMARCB1 re-expression led to SWI/SNF re-localization to what have been described as lineage-specific enhancers[49,50]. However, the transcription factor motifs at SMARCB1-bound enhancers were not always informative as to the nature of the cell of origin. The lack of a clearly defined cell(s) of origin, and their intrinsic biology has hampered a detailed understanding of the transformation process. Mechanistic studies were often limited to SMARCB1 re-expression in RT cell lines with little supporting patient data. An exception is atypical teratoid RT (AT/RT) where the epigenetic profiles of the AT/RT tumours were compared with other types of brain tumours or normal brain[51]. Nevertheless, the validity of these comparisons is limited since more recent data provide strong evidence that RT arise following arrest of neural crest cell differentiation into mesenchyme, in particular Schwann cells[52]. Many of the above limitations have been overcome in our study, where patient-derived and functional in cellulo data converged to define the transcriptional program of the TAL cell of origin and to decipher the mechanistic details of a reversible transcriptional switch driving their transformation into epithelial- and mesenchymal-type RMC states. We thus highlight the fundamental difference between RMC arising from mesoderm-derived differentiated epithelial TAL cells and RT derived from differentiating neural crest cells.

A link between RMC ferroptosis and sickle cell trait

A key finding of our study is activation of a ferroptosis resistance pathway in RMC cells. Analyses of gene expression signatures in scRNA-seq, patient cohort RNA-seq and the RMC cell lines defined how the ferroptosis sensitivity signature in TAL cells is replaced by a ferroptosis resistance signature in RMC cells. This process is reversed in RMC cells upon SMARCB1 re-expression leading to their ferroptotic cell death unlike other RT cells where SMARCB1 re-expression leads to cell cycle arrest or apoptosis[49,50,53]. Indeed, RMC cells are more sensitive to GPX4 inhibition than RT lines. Ferroptosis is therefore a specific vulnerability of RMC tumours.

The above observations link the RMC oncogenic process with sickle cell trait. The kidney medulla is amongst the most hypoxic micro-environments in the organism[54]. Due to its central role in urine concentration, the loop of Henle is characterized by increasing osmolarity and hypoxia that are highest in the TAL region. Msaouel et al. proposed a model where the high interstitial NaCl concentration induces DNA double strand breaks (DSB), whereas microcirculatory ischaemia induced by red blood cell (RBC) sickling reduces this osmolarity reactivating DSB repair in a chronic hypoxic environment by NEHJ favoring translocations and deletions, particularly in fragile regions such as chromosome 22q where the *SMARCB1* locus is located[55].

Our observations enrich this model with iron release by RBC sickling favouring ferroptosis of TAL cells and their renewal to maintain the homoeostasis of the epithelium[56,57]. Early initiation of ferroptosis resistance observed in the pre-tumoural TAL cells would thus promote their survival under the high NaCl and hypoxic conditions driving error-prone DSB repair. The increased extracellular iron concentration due to the fragility of the sickled RBCs acts as a selective pressure for survival of ferroptosis resistant TAL cells in an environment propitious to the mutagenic events associated with RMC development. This unique set of circumstances may explain why RMC is the only SMARCB1-deficient tumour arising from epithelial cells, compared to RTs arising from a developmental block of neural crest differentiation.

Methods

Ethical approval

Tumour sample collection for further research analysis was approved by ethical Committees of Strasbourg University Hospital and Curie Institute and all patients provided an informed written consent for the use of material for further research. Animal care and use for this study were performed in accordance with the recommendations of the European Community (2010/63/UE) for the care and use of laboratory animals and carried out in accordance with the principles of the Declaration of Helsinki and with GDPR regulations. The experiments were approved by the Curie Institute animal ethical committee CEEA-IC #118 (Authorization APAFIS#11206-2017090816044613-v2 given by National Authority) and performed in accordance with the internal, national and European guidelines of Animal Care and Use. The establishment of PDX received approval by the Institut Curie institutional review board OBS170323 CPP ref 3272; n de dossier (2015- A00464-45). Written institutional informed consent was obtained from the patient.

Article 1: SMARCB1 regulates a TFCP2L1-MYC transcriptional switch promoting renal medullary carcinoma transformation and ferroptosis resistance

Tumour samples

The two RMC samples subjected to scRNAseq were collected from Strasbourg University Hospital and Curie Institute, according to institutional guidelines. One tumour came from a 21-year-old female patient on a post-treatment primary nephrectomy from an RMC patient with lung metastases at diagnosis, whereas the second came from a 16-year-old male patient with regional lymph node and adrenal gland metastases (pT4N1M1). Bulk RNAseq came from 11 patients recently reported[6] and we generated an additional dataset of multi-region RNAseq of a cohort of 4 of the RMC patients, including multiple sections and lymph nodes metastasis (Supplementary Data 2).

Human single-cell sample preparation and RNA-seq

Following the treated tumour resection, samples from the tumour and adjacent non-malignant normal adjacent tissue were each conserved at 4 °C in 1 mL of MACS Tissue Storage Solution (Miltenyi Biotech). Single cell suspensions were prepared using gentleMACS™ dissociator and human tumour dissociation kit (Miltenyi Biotech) following manufacturer's instructions. Samples were applied to a MACS SmartStrainer 70 μm (Miltenyi Biotech) placed on a 15 mL Falcon tube and 10 mL DMEM were used to wash C tube and SmartStrainer. Following centrifugation at 300 g and 4 °C for 10 min, cells were sorted using CD45 (TIL) Microbeads (Miltenyi Biotech). CD45+ and CD45- fractions were centrifuged (300 g, 10 min, 4 °C) and dead cells were removed using Dead cell removal kit (Miltenyi Biotech). CD45- and CD45+ were mixed in 1 to 4 ratios. Cell viability and concentration were assessed before 3′-mRNA single-cell libraries were prepared using the Chromium (10x Genomics) following the manufacturer's instructions. Libraries were sequenced 2x100bp on HiSeq4000 sequencer.

Folowing resection of the naive tumour, the sample was cut in small pieces then dissociated 30 min at 37 °C in CO2-independent medium (Gibco) + 0,4 g/l of human albumin (Vialebex) with Liberase TL (Roche) 150 ug/ml and DNase 1 (Sigma) 150 µg/ml. Dissociated cells were then filtered with a 40 mm cell strainer, then washed and resuspended with CO2-independent medium + 0,4 g/l of human albumin. A fraction of the cell suspension was used to enrich tumour cells using Tumour isolation kit (Miltenyi Biotech, cat#130-108-339). Cells were then resuspended at 800 cells/ul in PBS + BSA 0,04%. Tissues were processed within 1 h after tumour resection, and sorted cells were loaded in a 10x Chromium instrument within 6 h.

Patient-derived xenograft sample preparation

Renal medullary carcinoma (RMC) patient derived xenograft (IC-pPDX-132) was established from a resected RMC tumour treated with 6 cycles of cisplatin, gemcitabine and bevacizumab. The undissociated tumour was engrafted in the subscapular fat pad of NSG (NOD.Cg-Prkdc^{scid} IL2^{rgtm1Wjl}/SzJ) mice. A PDX tumour fragment was then serially transplanted using the same procedure into Swiss Nude (Crl:NU(Ico)-Foxn1^{nu}) mice until passage 4 which was used for the single cell RNA-seq experiments. No PDX tumour was allowed to grow beyond the 1000 mm3 size limitation. Mice were maintained in IVC cages in a semi pathogen-free facility under standard housing conditions with continuous access to food and water. Curie Institute animal facilities comply with all appropriate standards (cages, space per animal, temperature (22 °C), light, 12-hour light/dark cycle, 50% humidity, continuous access to food and water), and all cages are enriched with nesting materials.

scRNA-seq analysis of human primary RMC tumours

After sequencing, raw reads were processed using CellRanger (v 3.1) to align on the hg19 human genome, remove unexpressed genes and quantify barcodes and UMIs. Data were then analysed in R (v4.0.2). For the treated tumour, tumour and NAT samples were aggregated with the cellranger 'aggr' command. The resulting aggregation was analysed with Seurat v3.2.0 following the recommended workflow. Cells were filtered for feature count ranging from 120 to 2000 and percentage of mitochondrial reads <15%. Counts were normalized with the "LogNormalize" method and data scaled to remove unwanted sources of variation (UMI count and mitochondrial reads). The number of principal components was determined from the Jackstraw plots. Clustering was performed on variable features using the 25 most significant principal components and a resolution of 1.15. For the naive tumour, the same Seurat pipeline was performed using feature counts from 200 to 6000, mitochondrial read fraction <20% and a resolution of 1.0 using the 20 most significant principal component for the clustering. Aggregate analyses of tumours 1 and 2 was performed by merging the two R objects and using the Seurat sctransform with batch correction function to normalize and scale data reducing the impact of technical factors.

scRNA-seq analysis of patient-derived RMC xenograft

For the IC-pPDX-132 sample raw reads were aligned on an hg19-mm10 hybrid genome. Cells were filtered based on feature counts ranging from 200 to 7000 and global clustering performed with a resolution of 0.3 using the 20 most significant components. Human and Mouse cells were re-clustered separately by first filtering cells with mitochondrial read fraction >20% and then using a resolution of 0.4 with 25 principal components.

Functional analysis using scRNA-seq data

Regulome analyses of active transcription factors were performed using the SCENIC v1.1.2.2 package[17]. Transcription factor activities were visualized on the UMAP using AUCell v1.8.0 or as heatmaps using the R-package 'pheatmap'. RMC correlations with the different renal tubule clusters were computed by Clustifyr v1.0.0[58] using cluster marker signatures for RMC (*TIMP1, FN1, CTHRC1, DCBLD2, COL1A2, COL1A1, ARL4C, COL6A2, LGALS1, CD44, VIM, CLU, MMP7, SERPINA1, WFDC2, SFRP2, MUC1, KRT18, KRT7, EPCAM, CDH1, CLDN4, CLDN10, DEFB1*), RMC1 (*WFDC2, FXYD2, SLPI, CLDN4, KRT7, KLF6, GSTP1, EEF1A1, CLDN3, TM4SF1*) and RMC2 (*FN1, COL1A2, COL1A1, TIMP1, CD44, CTHRC1, RARRES3, BGN, TFPI2, COL6A2*). Trajectory analyses were plotted and visualized using Similarity Weighted Nonnegative Embedding (SWNE)[59]. Gene set variation analysis were performed using the r-package GSVA[60].

For the "bulk RMC signature", the upregulated genes from the differential analysis of the MDACC RMC cohort (11 tumours versus 6 NAT) were selected using log2FC > 2 and FDR < 0.01[6]. For all signatures, gene sets were retrieved from either Hallmarks MSigDB or KEGG pathways. Gene signatures were computed and visualized on UMAPs using the R package VISION v2.1.0 (https://github.com/YosefLab/VISION).

Anti-4 Hydroxynonenal staining of RMC tumours

Sections from 2 independent RMC tumours and as control a colorectal cancer were fixed in 10% neutral-buffered formalin, paraffin embedded, sectioned, and stained with hematoxylin and eosin. 4-µm tissue sections were processed on VENTANA-Benchmark-XT, with incubation at room temperature in an antigen retrieval process (EDTA citrate buffer, pH 8,3, CC1 buffer, 8 mn), then incubated with 4HNJ-2 (Anti-4 Hydroxynonenal antibody, mouse monoclonal, clone HNEJ-2, Abcam; dilution: 1/5000 during 32 mn), revealed with 'Ultra View' Universal DAB Detection kit and counterstained with Hematoxylin solution (Ventana Roche Systems).

Cell culture, establishment of RMC lines stably expressing SMARCB1

RMC219 cells were grown in HAM-F12/D-MEM (1:1) medium supplemented with 10% foetal calf serum (FCS), Glutamine 2 mM, AANE and PS. UOK360 and UOK353 cells were grown in D-MEM medium supplemented with 10% foetal calf serum (FCS) and Glutamine 2 mM.

Article 1: SMARCB1 regulates a TFCP2L1-MYC transcriptional switch promoting renal medullary carcinoma transformation and ferroptosis resistance

RMC2C cells were grown in MEM medium with 10% FCS, AANE, 50 ng/mL EGF and PS. Cell lines were provided by colleagues and are not commercially available. Authentication performed by immunoblot showing absence of SMARCB1 expression and by RNA-seq. HEK 293 T cells were obtained from ATCC. All cell lines were regularly tested as negative for Mycoplasma infection using the Venor™ GeM Mycoplasma Detection Kit, and used at less than 10 passages. RMC cells infected with lentiviral constructs were grown in respective media replacing normal FCS with tetracyclin-free FCS (Dutscher) and supplemented with G418 (300ug/mL). SMARCB1 expression was induced by treatment with either DMSO or 2 μM of doxycycline.

Lentiviral pInducer20 vector was obtained from Addgene and the cDNA of either SMARCB1 or mCherry was cloned into the vector by Gateway. We then used pInducer20-mCherry or -SMARCB1 containing lentiviruses to infect 1×10^6 RMC2C or RMC219 cells. Cells were selected using 500ug/mL G418 for a week and then maintained under these conditions.

In vitro treatments

For ferroptosis, cells were either treated with DMSO or 2uM doxycycline alone or co-treated with 2uM doxycycline and 1uM ferrostatin-1 (SelleckChem, #S7243), zVAD-fmk (MedChemExpress, #HY-16658B) or necrostatin-1 (MedChemExpress, #HY-15760) for the indicated times. For the Caspase-3 assays, cells were either treated with 5uM camptothecin (SelleckChem, #S1288) for 4 h, DMSO or 2uM doxycycline for the indicated times. For the IFNg experiments, cells were either treated with DMSO or 10 ng/mL of IFNg (Peprotech, 300-02).

Cell death, caspase-3 and lipid peroxidation analyses by flow cytometry

Cells were harvested at the indicated times and co-stained with Annexin-V-FITC and propidium iodide following manufacturer instructions (BioLegend, #640914). To assess active Caspase-3, cells were fixed and permeabilized before incubation with the FITC-conjugated caspase-3 antibody (dilution 1/6) following manufacturer's instructions (BD Biosciences # 550914) for subsequent flow cytometry analysis. To assess membrane lipid perodixation, cells were stained using 10uM of Bodipy 581/591 C11 (ThermoFisher, #D3861) following manufacturer's instructions. To assess senescence, cells were treated with 100 nM bafilomycin A1 (Sigma, #19-148) for 1 h followed by 2 mM C12FDG (Invitrogen, #D2893) for 2 h before being washed and harvested for flow cytometry analyses. All assays were analysed on a LSRII Fortessa (BD Biosciences) and data were analysed using Flowjo v6.8.

Immunofluorescence

Cells grown on glass slides in 24-well plates, were fixed with 4% paraformaldheyde for 15 min. After two washes with PBS buffer, they were permeabilized in PBS+triton X-100 0,1% for 5 min and blocked with PBS + 10% FCS inactivated for 20 min. Primary antibodies were incubated overnight at 4 °C and after three washes with PBS+Triton 0,1%, cells were stained for 1 h at room temperature with AlexaFluor-488 conjugated secondary antibodies (Invitrogen goat anti mouse # A11001 and goat anti-rabbit # A32731) diluted 1/500 in PBS + 10% FCS. After three washes with PBS+Triton 0,1%, cells were stained with DAPI (final concentration 1 ug/ml) and mounted on microscopy slides. Images were captured with a confocal (Leica DMI6000) microscope.

Cell viability assay by fluorescence screening

In total, 5×10^3 of indicated cell types were seeded on 96-well plates in four technical replicates on day 1. The next day, cells were treated either with DMSO control or with an increasing concentration of RSL3 (SelleckChem, 8155) ranging from 0 to 10 μM. At day 3, cells were washed with PBS and stained using PrestoBlue (Invitrogen, A13261) according to manufacturer instructions before fluorescence was quantified on a multi-modal spectrometer (Berthold Mithras, LB940). IC50 values were calculated using the fraction of DMSO control.

Immunostaining quantification by flow cytometry

Wildtype RMC219 and RMC2C cells were harvested and 1×10^6 cells were resuspended in buffer A (PBS 1X, EDTA 2 mM, inactivated FCS 1%) and 5uL of Human TruStain FcX (Biolegend, 422301) was added for 10 min at room temperature. Following blocking, cells were stained for 1 h with 5 μL of conjugated EPCAM-FITC (Biolegend, 324203) and conjugated CD44-PE (Biolegend, 103023). Following two PBS washes, cells were resuspended in buffer A before flow cytometry on a LSRII Fortessa (BD Biosciences) and analysis using Flowjo v6.8.

Boyden Chamber Invasion assays

Before seeding, 100ul of diluted Matrigel (1:20, 356234, Corning) was added in each insert (24-well 8um inserts, Corning) and left to dry for 2 h at 37 °C before being washed twice with PBS. Subsequently, RMC cells were harvested and 2×10^5 cells and seeded in the Boyden chambers in corresponding media without serum. 24 h later, migrated cells were fixed using PFA 4% for 10 min before being stained using Crystal violet for 10 min. Excess stain was washed 3 times in PBS before images were captured on phase contrast microscope. Quantification of migrated cells was done by resuspension of staining using 100 mM acetic acid for 15 min before luminescence was measured on a BioTek Luminescence microplate reader (using Gen5 software).

RNA preparation and quantitative PCR

RNA isolation was performed according to standard procedures (Macherey Nagel RNA Plus kit). RT-qPCR was carried out with SYBR Green I (Roche) and SuperScript IV Reverse Transcriptase (Invitrogen) and monitored using a LightCycler 480 (Roche). The mean of ACTB, TBP, RPL13A and GAPDH gene expressions was used to normalize the results. Primer sequences for each cDNA were designed using Primer3 Software and are available in Supplementary Table 1.

Public data correlation analysis using TGCA and CCLE database

Spearman correlation for all selected genes were retrieved from co-expression studies using the Cancer Cell Line Encyclopaedia (Broad, 2019) and the TCGA chromophobe renal cell carcinoma (KICH) databases. All transcription factors were extracted using the "Full Human TFs" list from[61]. Scatter plots were made using Prism5. For the correlation with TFCP2L1, the epithelial and mesenchymal genes were retrieved from Watanabe et al.[23].

Bulk RNA sequencing

RMC cell lines were analysed by RNA-seq under the different indicated conditions. After sequencing raw reads were pre-processed in order to remove adapter and low-quality sequences (Phred quality score below 20) using cutadapt version 1.10. and reads shorter than 40 bases were discarded. Reads were maping to rRNA sequences using bowtie version 2.2.8, were also removed. Reads were mapped onto the hg19 assembly of Homo sapiens genome using STAR version 2.5.3a. Gene expression quantification was performed from uniquely aligned reads using htseq-count version 0.6.1p1, with annotations from Ensembl version 75 and "union" mode. Only non-ambiguously assigned reads were retained for further analyses. Read counts were normalized across samples with the median-of-ratios method. Comparisons of interest were performed using the Wald test for differential expression and implemented in the Bioconductor package DESeq2 version 1.16.1. Genes with high Cook's distance were filtered out and independent filtering based on the mean of normalized counts was performed. P-values were adjusted for multiple testing using the Benjamini and Hochberg method. Deregulated genes were defined as genes with log2(foldchange) >1 or <−1 and adjusted p value < 0.05.

Article 1: SMARCB1 regulates a TFCP2L1-MYC transcriptional switch promoting renal medullary carcinoma transformation and ferroptosis resistance

Analysis of bulk RNA-seq of patient samples

For RMC cohorts, raw counts were retrieved in excel format and normalized first by sequencing depth using DESeq2 sizefactors and then divided by median of gene length. Samples were clustered using the hclust function with "ward.D2" linkage function and visualized as heatmaps using pheatmap package v1.0.12. The deconvolution of immune and stromal cells was done using MCP-counter v1.2.0[62]. Sample compositions were also estimated by deconvolution from our single-cell data using the CIBERSORTx algorithm[63]. Volcano plots were generated with ggplot2 v3.3.2. Gene set enrichment analyses were done with the GSEA software v3.0 using the hallmark gene sets of Molecular Signature Database v6.2. Gene Ontology analysis was done using DAVID (http://david.abcc.ncifcrf.gov/). Gene list intersections and Venn diagrams were performed by Venny.

Protein extraction and Western blotting

Whole cell extracts were prepared by the standard freeze-thaw technique using LSDB 500 buffer (500 mM KCl, 25 mM Tris at pH 7.9, 10% glycerol (v/v), 0.05% NP-40 (v/v), 16 mM DTT, and protease inhibitor cocktail). Cell lysates were subjected to SDS–polyacrylamide gel electrophoresis (SDS-PAGE) and proteins were transferred onto a nitrocellulose membrane. Membranes were incubated with primary antibodies in 5% dry fat milk and 0.01% Tween-20 overnight at 4 °C. The membrane was then incubated with HRP-conjugated secondary antibody (Jackson ImmunoResearch; Goat against Mouse: 115-036-71; Goat against Rabbit: 111-035-144 dilution 1.2000) for 1 h at room temperature, and visualized using the ECL detection system (GE Healthcare). The references of all antibodies are available in Supplementary Table 2.

Chromatin immunoprecipitation and sequencing (ChIP-seq)

BRG1 ChIP experiments were performed on native MNase-digested chromatin. Between 10 to 20 × 10^8 freshly harvested RMC2C cells bearing either SMARCB1 or mCHERRY and treated 2uM doxycycline for 48 h were resuspended in 1.5 ml ice-cold hypotonic buffer (0.3 M Sucrose, 60 mM KCl, 15 mM NaCl, 5 mM MgCl2, 0.1 mM EDTA, 15 mM Tris–HCl pH 7.5, 0.5 mM DTT, 0.1 mM PMSF, PIC) and cytoplasmic fraction was released by incubation with 1.5 ml of lysis-buffer (0.3 M sucrose, 60 mM KCl, 15 mM NaCl, 5 mM MgCl2, 0.1 mM EDTA, 15 mM Tris–HCl pH 7.5, 0.5 mM DTT, 0.1 mM PMSF, PIC, 0.5% (vol/vol) IGEPAL CA-630) for 10 min on ice. The suspension was layered onto a sucrose cushion (1.2 M sucrose, 60 mM KCl, 15 mM NaCl, 5 mM MgCl2, 0.1 mM EDTA, 15 mM Tris–HCl [pH 7.5], 0.5 mM DTT, 0.1 mM PMSF, PIC) and centrifuged for 30 min 4 °C at 5000 g in a swing rotor. The nuclear pellet was resuspended in digestion buffer (0.32Msucrose, 50 mM Tris–HCl [pH 7.5], 4 mM MgCl2, 1 mM CaCl2, 0.1 mM PMSF) and incubated with 10ul of Micrococcal Nuclease (NEB) for 7 min at 37 °C. The reaction was stopped by addition of 20ul of EDTA 0,5 M and suspension chilled on ice for 10 min. The suspension was cleared by centrifugation at 8000 g (4 °C) for 10 min and supernatant (chromatin) was used for further purposes. Chromatin was digested to around 80% of mono-nucleosomes as judged by extraction of the DNA and agarose gel electrophoresis. H3K27ac and MYC ChIP experiments were performed on 0.4% PFA-fixed chromatin isolated from RMC2C cells bearing either SMARCB1 or mCHERRY and treated 2uM doxycycline for 48 h according to standard protocols[64]. ChIP-seq libraries were prepared using MicroPlex Library Preparation kit v2 and sequenced on the Illumina Hi-seq 4000 as single-end 50-base reads[65]. Sequenced reads were mapped to the Homo sapiens genome assembly hg19 using Bowtie with the following arguments: -m 1 --strata --best -y -S -l 40 -p 2. Cut&Tag was performed using the Active Motif CUT&Tag-IT kit following the manufacturer's instructions.

ChIP-seq analysis

After sequencing, peak detection was performed using the MACS software (Zhang et al., 2008). Peaks were annotated with Homer using the GTF from ENSEMBL v75. Global clustering analysis and quantitative comparisons were performed using seqMINER[66]. Super-enhancers were called with the python package Ranking Of Super Enhancers (ROSE) https://github.com/stjude/ROSE.

De novo motif discovery on FASTA sequences corresponding to windowed peaks was performed using MEME suite (meme-suite.org). Motif correlation matrix was calculated with in-house algorithms using JASPAR database as described in[67]. Motif discovery from ChIP-seq peaks was performed using the RSAT peak-motifs algorithm (http://rsat.sb-roscoff.fr/peak-motifs_form.cgi).

Motif analysis Searching of known TF motifs from the Jaspar 2014 motif database at BRG1-bound sites was made using FIMO[68] within regions of 200 bp around peak summits, FIMO results were then processed by a custom Perl script which computed the frequency of occurrence of each motif. To assess the enrichment of motifs within the regions of interest, the same analysis was done 100 times on randomly selected regions of the same length as the BRG1 bound regions and the results used to compute an expected distribution of motif occurrence. The significance of the motif occurrence at the BRG1-occupied regions was estimated through the computation of a Z-score (z) with $z = (x - \mu)/\sigma$, where: − x is the observed value (number of motif occurrence), − μ is the mean of the number of occurrences (computed on randomly selected data), − σ is the standard deviation of the number of occurrences of motifs (computed on randomly selected data). The source code is accessible at https://github.com/slegras/motif-search-significance.git.

Statistics

All experiments were performed in biological triplicates, unless stated otherwise in the figure legends. All tests used for statistical significance were calculated using Prism5 and indicated in the figure legends along with p values (****$p < 0.0001$, ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, ns: $p > 0.05$).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data used in this study are available in the main article or as supplementary information to the manuscript. Source data are provided as a Source Data file. The sequencing data used in this study are publicly available in the GEO database under accession number GSE181001. Source data are provided with this paper.

References

1. Davis, C. J., Mostofi, F. K. & Sesterhenn, I. A. Renal medullary carcinoma. The seventh sickle cell nephropathy. *Am. J. Surg. Pathol.* **19**, 1–11 (1995).
2. Cajaiba, M. M. et al. The classification of pediatric and young adult renal cell carcinomas registered on the children's oncology group (COG) protocol AREN03B2 after focused genetic testing. *Cancer* **124**, 3381–3389 (2018).
3. Alvarez, O., Rodriguez, M. M., Jordan, L. & Sarnaik, S. Renal medullary carcinoma and sickle cell trait: A systematic review. *Pediatr. Blood Cancer* **62**, 1694–1699 (2015).
4. Msaouel, P., Carugo, A. & Genovese, G. Targeting proteostasis and autophagy in SMARCB1-deficient malignancies: where next? *Oncotarget* **10**, 3979–3981 (2019).
5. Carugo, A. et al. p53 Is a Master Regulator of Proteostasis in SMARCB1-Deficient Malignant Rhabdoid Tumors. *Cancer Cell* **35**, 204–220.e9 (2019).
6. Msaouel, P. et al. Comprehensive molecular characterization identifies distinct genomic and immune hallmarks of renal medullary carcinoma. *Cancer Cell* **37**, 720–734.e13 (2020).

Article 1: SMARCB1 regulates a TFCP2L1-MYC transcriptional switch promoting renal medullary carcinoma transformation and ferroptosis resistance

7. Msaouel, P., Walker, C. L., Genovese, G. & Tannir, N. M. Molecular hallmarks of renal medullary carcinoma: more to c-MYC than meets the eye. *Mol. Cell. Oncol.* **7**, 1777060 (2020).
8. Swartz, M. A. et al. Renal medullary carcinoma: clinical, pathologic, immunohistochemical, and genetic analysis with pathogenetic implications. *Urology* **60**, 1083–1089 (2002).
9. Gupta, R. et al. Carcinoma of the collecting ducts of Bellini and renal medullary carcinoma: clinicopathologic analysis of 52 cases of rare aggressive subtypes of renal cell carcinoma with a focus on their interrelationship. *Am. J. Surg. Pathol.* **36**, 1265–1278 (2012).
10. Beckermann, K. E. et al. Renal medullary carcinoma: establishing standards in practice. *J. Oncol. Pract.* **13**, 414–421 (2017).
11. Dimashkieh, H., Choe, J. & Mutema, G. Renal medullary carcinoma: a report of 2 cases and review of the literature. *Arch. Pathol. Lab. Med.* **127**, e135–e138 (2003).
12. Elliott, A. & Bruner, E. Renal medullary carcinoma. *Arch. Pathol. Lab. Med.* **143**, 1556–1561 (2019).
13. Cheval, L., Pierrat, F., Rajerison, R., Piquemal, D. & Doucet, A. Of mice and men: divergence of gene expression patterns in kidney. *PloS One* **7**, e46876 (2012).
14. Lake, B. B. et al. A single-nucleus RNA-sequencing pipeline to decipher the molecular anatomy and pathophysiology of human kidneys. *Nat. Commun.* **10**, 2832 (2019).
15. Muto, Y. et al. Single cell transcriptional and chromatin accessibility profiling redefine cellular heterogeneity in the adult human kidney. *Nat. Commun.* **12**, 2190 (2021).
16. Young, M. D. et al. Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science* **361**, 594–599 (2018).
17. Van de Sande, B. et al. A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat. Protoc.* **15**, 2247–2276 (2020).
18. Werth, M. et al. Transcription factor TFCP2L1 patterns cells in the mouse kidney collecting ducts. *eLife* **6**, e24265 (2017).
19. LeBleu, V. S. et al. PGC-1α mediates mitochondrial biogenesis and oxidative phosphorylation in cancer cells to promote metastasis. *Nat. Cell Biol.* **16**, 1–15 (2014). 992–1003.
20. Dimova, D. K. & Dyson, N. J. The E2F transcriptional network: old acquaintances with new faces. *Oncogene* **24**, 2810–2826 (2005).
21. Mullan, P. B., Quinn, J. E. & Harkin, D. P. The role of BRCA1 in transcriptional regulation and cell cycle control. *Oncogene* **25**, 5854–5863 (2006).
22. Sengez, B. et al. The Transcription Factor Elf3 Is Essential for a Successful Mesenchymal to Epithelial Transition. *Cells* **8**, E858 (2019).
23. Watanabe, K. et al. OVOL2 induces mesenchymal-to-epithelial transition in fibroblasts and enhances cell-state reprogramming towards epithelial lineages. *Sci. Rep.* **9**, 6490 (2019).
24. Liu, Z.-H., Dai, X.-M. & Du, B. Hes1: a key role in stemness, metastasis and multidrug resistance. *Cancer Biol. Ther.* **16**, 353–359 (2015).
25. Yin, J. et al. HGF/MET regulated epithelial-mesenchymal transitions and metastasis By FOSL2 in non-small cell lung cancer. *OncoTargets Ther.* **12**, 9227–9237 (2019).
26. Senft, D. & Ronai, Z. A. UPR, autophagy, and mitochondria crosstalk underlies the ER stress response. *Trends Biochem. Sci.* **40**, 141–148 (2015).
27. LeBleu, V. S. & Kalluri, R. A peek into cancer-associated fibroblasts: origins, functions and translational impact. *Dis. Model. Mech.* **11**, dmm029447 (2018).
28. Conejo-Garcia, J. R. & Rodriguez, P. C. c-Maf: a bad influence in the education of macrophages. *J. Clin. Invest.* **130**, 1629–1631 (2020).
29. Dong, Y. et al. Tumor xenografts of human clear cell renal cell carcinoma but not corresponding cell lines recapitulate clinical response to sunitinib: feasibility of using biopsy samples. *Eur. Urol. Focus* **3**, 590–598 (2017).
30. Zanjani, L. S. et al. Increased expression of CD44 is associated with more aggressive behavior in clear cell renal cell carcinoma. *Biomark. Med.* **12**, 45–61 (2018).
31. Wei, D. et al. Novel renal medullary carcinoma cell lines, UOK353 and UOK360, provide preclinical tools to identify new therapeutic treatments. *Genes. Chromosomes Cancer* **59**, 472–483 (2020).
32. Nakai, S. et al. Crucial roles of Brn1 in distal tubule formation and function in mouse kidney. *Dev. Camb. Engl.* **130**, 4751–4759 (2003).
33. Phelep, A. et al. MITF—A controls branching morphogenesis and nephron endowment. *PLoS Genet.* **13**, e1007093 (2017).
34. Dixon, S. J. et al. Ferroptosis: an iron-dependent form of non-apoptotic cell death. *Cell* **149**, 1060–1072 (2012).
35. Chen, X., Kang, R., Kroemer, G. & Tang, D. Broadening horizons: the role of ferroptosis in cancer. *Nat. Rev. Clin. Oncol.* **18**, 280–296 (2021).
36. Dai, C. et al. Transcription factors in ferroptotic cell death. *Cancer Gene Ther.* **27**, 645–656 (2020).
37. Yuk, H., Abdullah, M., Kim, D.-H., Lee, H. & Lee, S.-J. Necrostatin-1 Prevents Ferroptosis in a RIPK1- and IDO-Independent Manner in Hepatocellular Carcinoma. *Antioxidants* **10**, 1347 (2021).
38. Stockwell, B. R. Ferroptosis turns 10: Emerging mechanisms, physiological functions, and therapeutic applications. *Cell* **185**, 2401–2421 (2022).
39. Feng, H. et al. Transferrin receptor is a specific ferroptosis marker. *Cell Rep.* **30**, 3411–3423.e7 (2020).
40. Tsoi, J. et al. Multi-stage differentiation defines melanoma subtypes with differential vulnerability to drug-induced iron-dependent oxidative stress. *Cancer Cell* **33**, 890–904.e5 (2018).
41. Wang, W. et al. CD8+ T cells regulate tumour ferroptosis during cancer immunotherapy. *Nature* **569**, 270–274 (2019).
42. Hnisz, D. et al. Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
43. Whyte, W. A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
44. Weissmiller, A. M. et al. Inhibition of MYC by the SMARCB1 tumor suppressor. *Nat. Commun.* **10**, 2014 (2019).
45. Woodley, C. M. et al. Multiple interactions of the oncoprotein transcription factor MYC with the SWI/SNF chromatin remodeler. *Oncogene* **40**, 3593–3609 (2021).
46. Laurette, P. et al. Transcription factor MITF and remodeller BRG1 define chromatin organisation at regulatory elements in melanoma cells. *eLife* **4**, e06857 (2015).
47. Michel, B. C. et al. A non-canonical SWI/SNF complex is a synthetic lethal target in cancers driven by BAF complex perturbation. *Nat. Cell Biol.* **20**, 1410–1420 (2018).
48. Wang, X. et al. BRD9 defines a SWI/SNF sub-complex and constitutes a specific vulnerability in malignant rhabdoid tumors. *Nat. Commun.* **10**, 1881 (2019).
49. Nakayama, R. T. et al. SMARCB1 is required for widespread BAF complex-mediated activation of enhancers and bivalent promoters. *Nat. Genet.* **49**, 1613–1623 (2017).
50. Wang, X. et al. SMARCB1-mediated SWI/SNF complex function is essential for enhancer regulation. *Nat. Genet.* **49**, 289–295 (2017).
51. Erkek, S. et al. Comprehensive analysis of chromatin states in atypical teratoid/rhabdoid tumor identifies diverging roles for SWI/SNF and polycomb in gene regulation. *Cancer Cell* **35**, 95–110.e8 (2019).
52. Custers, L. et al. Somatic mutations and single-cell transcriptomes reveal the root of malignant rhabdoid tumours. *Nat. Commun.* **12**, 1407 (2021).
53. Betz, B. L., Strobeck, M. W., Reisman, D. N., Knudsen, E. S. & Weissman, B. E. Re-expression of hSNF5/INI1/BAF47 in pediatric tumor cells leads to G1 arrest associated with induction of p16ink4a and activation of RB. *Oncogene* **21**, 5193–5203 (2002).

Article 1: SMARCB1 regulates a TFCP2L1-MYC transcriptional switch promoting renal medullary carcinoma transformation and ferroptosis resistance

54. Evans, R. G., Smith, D. W., Lee, C.-J., Ngo, J. P. & Gardiner, B. S. What makes the kidney susceptible to hypoxia? *Anat. Rec. Hoboken NJ 2007* **303**, 2544–2552 (2020).
55. Msaouel, P., Tannir, N. M. & Walker, C. L. A model linking sickle cell hemoglobinopathies and SMARCB1 loss in renal medullary carcinoma. *Clin. Cancer Res. J. Am. Assoc. Cancer Res.* **24**, 2044–2049 (2018).
56. Humphreys, B. D. et al. Intrinsic epithelial cells repair the kidney after injury. *Cell Stem Cell* **2**, 284–291 (2008).
57. Scindia, Y., Leeds, J. & Swaminathan, S. Iron homeostasis in healthy kidney and its role in acute kidney injury. *Semin. Nephrol.* **39**, 76–84 (2019).
58. Fu, R. et al. clustifyr: an R package for automated single-cell RNA sequencing cluster classification. *F1000Research* **9**, 223 (2020).
59. Wu, Y., Tamayo, P. & Zhang, K. Visualizing and interpreting single-cell gene expression datasets with similarity weighted nonnegative embedding. *Cell Syst.* **7**, 656–666.e4 (2018).
60. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinforma.* **14**, 7 (2013).
61. Lambert, S. A. et al. The human transcription factors. *Cell* **172**, 650–665 (2018).
62. Becht, E. et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* **17**, 218 (2016).
63. Newman, A. M. et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **37**, 773–782 (2019).
64. Strub, T. et al. Essential role of microphthalmia transcription factor for DNA replication, mitosis and genomic stability in melanoma. *Oncogene* **30**, 2319–2332 (2011).
65. Laurette, P. et al. Chromatin remodellers Brg1 and Bptf are required for normal gene expression and progression of oncogenic Braf-driven mouse melanoma. *Cell Death Differ.* **27**, 29–43 (2020).
66. Ye, T. et al. seqMINER: an integrated ChIP-seq data interpretation platform. *Nucleic Acids Res.* **39**, e35 (2011).
67. Joshi, S. et al. TEAD transcription factors are required for normal primary myoblast differentiation in vitro and muscle regeneration in vivo. *PLoS Genet.* **13**, e1006600 (2017).
68. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinforma. Oxf. Engl.* **27**, 1017–1018 (2011).

Acknowledgements

We thank all the staff of the IGBMC common facilities in particular Betty Heller and Patricia Wagner from Cell Culture, Claudine Ebel and Muriel Philipps from Flow Cytometry and Dr. Paola Rossolillo and Karim Essabri of the molecular biology facility. We also would like to thank the Institut Curie facilities and in particular Dr Pascale Philippe-Chomette, Pr Michel Peuchmaur, Dr Yves Allory and Dr Pascale Maille for proving the primary specimen for the PDX experiments. This work was supported by institutional grants from the Centre National de la Recherche Scientifique, the Institut National de la Santé et de la Recherche Médicale, the Université de Strasbourg, the Association pour la Recherche contre le Cancer (CR, contract number PJA 20181208268), the Ligue Nationale contre le Cancer, the Institut National du Cancer (PLBIO-2021-079), the ANR-10-LABX-0030-INRT French state fund through the Agence Nationale de la Recherche under the frame programme Investissements d'Avenir labelled ANR-10-IDEX-0002-02 and MSD AVENIR research grant. The IGBMC high throughput sequencing facility is a member of the "France Génomique" consortium (ANR10-INBS-09-08). FB was supported by grants from the INCA for the scRNAseq program (INCA PRTK19-2020-036) and the Saint Bladrick Foundation. ID is an 'équipe labellisée' of the Ligue Nationale contre le Cancer. BHV was supported by fellowships from the ANR, the Ligue Nationale contre le Cancer.

Author contributions

Conceptualization, B.H.V., I.D., G.G.M.; Methodology, B.H.V., G.D., N.T.P.S., A.H., M.R., A.R.H., J.G., J.T,. P.B., P.M., X.S., D.S., R.B., J.E.K., F.B., N.M.T., I.D., G.G.M.; Formal Analysis, B.H.V., G.D., I.D., G.G.M.. Investigation, B.H.V., G.D., N.T.P.S, V.M., S.Z., P.R,. A.H., M.R., A.R.H., J.G., J.T., P.B., P.M., X.S., H.L., T.T., V.L., D.S., R.B., J.E.K., F.B., N.M.T., I.D., G.G.M.. Bioinformatics: G.D.,. B.H.V., X.S.; Writing—Original Draft, B.H.V., I.D., G.G.M.; Writing—Review and Editing, B.H.V., G.D., I.D., G.G.M.. Investigation, B.H.V., G.D., N.T.P.S., V.M., S.Z., P.R., A.H., M.R., A.R.H., J.G., J.T., P.B., P.M., X.S., H.L., T.T., V.L., D.S., R.B., J.E.K., F.B., N.M.T., I.D., G.G.M.; Resources, P.M., X.S., H.L., V.M., S.Z., P.R., T.T., V.L., D.S., R.B., F.B., N.M.T., I.D. and G.G.M.; Supervision: I.D., G.G.M.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-023-38472-y.

Correspondence and requests for materials should be addressed to Irwin Davidson or Gabriel G. Malouf.

Peer review information *Nature Communications* thanks A Ari Hakimi, Thomas Mitchell and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

[1]Department of Cancer and Functional Genomics, Institute of Genetics and Molecular and Cellular Biology, CNRS/INSERM/UNISTRA, 67400 Illkirch, France. [2]Department of Medical Oncology, Institut de Cancérologie Strasbourg Europe, 67200 Strasbourg, France. [3]INSERM U830, Équipe Labellisée LNCC, Diversity and Plasticity of Childhood Tumors Lab, Institut Curie Research Centre, 75005 Paris, France. [4]Department of Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. [5]Department of Genitourinary Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston,

Article 2: An atlas of the human liver circadian transcriptome and its carcinogenic perturbation by hepatitis C virus infection

This next article describes the novel influence of circadian rhythm on liver physiology and disease. It is currently under minor revisions in Nature Communications. Circadian rhythm corresponds to the changes in the organism over a 24 h time period. Notably, hepatitis C virus infection of chimeric mice with human hepatocytes showed that both the rhythmicity of the transcriptome and epigenome was disturbed compared to non-infected mice.

I participated in this study by performing chromatin immunoprecipitation on fixed chimeric mice livers infected or not by HCV and at different time points after the infection. I precipitated open chromatin with an antibody targeting H3K27ac and H3K9ac marks, followed by sequencing. This showed that almost 1500 genes expression were impacted by HCV infection, often showing loss of rhythmic expression and leading to chronic liver disease.

An atlas of the human liver circadian transcriptome and its carcinogenic perturbation by hepatitis C virus infection

Atish Mukherji[1,*,#], Frank Jühling[1,*], Yogy Simanjuntak[1], Emilie Crouchet[1], Fabio Del Zompo[1], Yuji Teraoka[2], Alexandre Haller[3], Philippe Baltzinger[3], Soumith Paritala[4], Fahmida Rasha[4], Naoto Fujiwara[4], Cloé Gadenne[1], Nevena Slovic[1], Marine A. Oudot[1], Sarah C. Durand[1], Clara Ponsolles[1], Catherine Schuster[1], Xiaodong Zhuang[5], Jacinta Holmes[6], Ming-Lun Yeh[7], Hiromi Abe-Chayama[8], Mathias Heikenwälder[9,10], Angelo Sangiovanni[11], Massimo Iavarone[11], Massimo Colombo[12], Steven K.H. Foung[13], Jane A McKeating[5], Irwin Davidson[3], Ming-Lung Yu[7,14], Raymond T. Chung[15], Yujin Hoshida[4], Kazuaki Chayama[16,17,18], Joachim Lupberger[1,§], and Thomas F. Baumert[1,19,20, 21,§,#]

Affiliations

[1]University of Strasbourg, Institute of Translational Medicine and Liver Diseases (ITM), Inserm UMR_S1110, Strasbourg, France.

[2]Department of Gastroenterology, National Hospital Organization Kure Medical Center, Hiroshima, Japan.

[3]Department of Functional Genomics and Cancer, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS/INSERM/University of Strasbourg, Illkirch, France.

[4]Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, USA.

[5]Nuffield Department of Medicine, University of Oxford, Oxford OX3 7FZ, UK.

[6]University of Melbourne, St Vincent's Hospital, Melbourne, Australia.

[7]Hepatobiliary Division, Department of Internal Medicine, School of Medicine and Hepatitis Research Center, College of Medicine, and Center for Liquid Biopsy and Cohort Research, Kaohsiung Medical University Hospital, Kaohsiung Medical University, Kaohsiung 80708, Taiwan.

[8]Center for Medical Specialist Graduate Education and Research, Hiroshima, Japan.

[9]Division of Chronic Inflammation and Cancer, German Cancer Research Center (DKFZ), Heidelberg, Germany.

[10]M3 Research Center, Tübingen, Germany.

[11]Division of Gastroenterology and Hepatology, Fondazione IRCCS Cà Granda Ospedale Maggiore Policlinico, Milan, Italy.

[12]EASL International Liver Foundation, Geneva, Switzerland.

[13]Department of Pathology, Stanford University School of Medicine, California 94305, USA.

[14]School of Medicine and Doctoral Program of Clinical and Experimental Medicine, College of Medicine and Center of Excellence for Metabolic Associated Fatty Liver Disease, National Sun Yat-sen University, Kaohsiung, Taiwan.

[15]Gastrointestinal Division, Hepatology and Liver Center, Massachusetts General Hospital, Boston, MA 02114, USA.

[16]Collaborative Research Laboratory of Medical Innovation, Graduate School of Biomedical and Health Sciences, Hiroshima University, Hiroshima, Japan.

[17]RIKEN Center for Integrative Medical Sciences, Yokohama, Japan.

[18]Hiroshima Institute of Life Sciences, Hiroshima, Japan.

[19]Gastroenterology and Hepatology Service, Center for Liver and Digestive Disease, Strasbourg University Hospitals, France.

[20]Institut Universitaire de France, Paris, France.

[21]IHU Strasbourg, France

[#]Correspondence to: Prof. Thomas F. Baumert, MD, and Dr. Atish Mukherji, Ph.D., Inserm U1110, University of Strasbourg, 3 Rue Koeberlé, F-67000 Strasbourg, France. Phone: +33368853703, email: thomas.baumert@unistra.fr , mukherji@unistra.fr

[*]A.M. and F.J. contributed equally to this work. §J.L. and T.F.B. having jointly supervised this work.

Article 2: An atlas of the human liver circadian transcriptome and its carcinogenic perturbation by hepatitis C virus infection

Summary:

Chronic liver disease and cancer are global health challenges. The role of the circadian clock (CC) as a regulator of liver physiology and disease is well established in rodents, however, the identity and epigenetic regulation of rhythmically expressed genes in human disease is unknown. Here we unraveled the rhythmic transcriptome and epigenome of human hepatocytes using human liver chimeric mice. We identified a large number of rhythmically expressed protein coding genes in human hepatocytes of chimeric mice, which includes key transcription factors, chromatin modifiers, and critical enzymes. We show that hepatitis C virus (HCV) infection, a major cause of liver disease and cancer, perturbs the transcriptome by altering the rhythmicity of the expression of more than 1500 genes, and affects epigenome, leading to an activation of critical pathways mediating metabolic alterations, fibrosis, and cancer. HCV-perturbed circadian pathways remain dysregulated in patients with advanced liver disease. Collectively, these data support a role for virus-induced perturbation of the liver clock in cancer development and may provide opportunities for cancer prevention and biomarkers to predict HCC risk.

Introduction

Mammals are endowed with an endogenous timing system known as the circadian clock (CC), a well-known regulator of physiology and behavior[1-4]. The CC is self-sustained and present in virtually all cell and tissue types. The CC-oscillator represents the core of this molecular 'clock' and operates as a transcriptional-translational feedback system[1-6]. Notably, components of the CC-oscillator drive daily rhythmicity of their own expression and the temporal expression of target genes (CC-controlled genes; CCGs), thereby ensuring coordinated function of cells, tissues, and organs[1-4]. Studies with animal model systems reveal an essential role for the CC to regulate metabolism, immune and endocrine functions[1-4,7]. Furthermore, perturbation of CC function in murine models associates with diverse pathologies including metabolic diseases and cancer[1-4,7-11]. However, the role of disrupted liver CC in the development of human chronic liver disease is largely unknown due to our limited knowledge on diurnal gene expression in the human liver.

Hepatitis C virus (HCV) is a major cause of chronic liver disease and hepatocellular carcinoma (HCC). Direct-acting antivirals (DAA) can eliminate HCV and cure infection[12,13], however, a significant HCC risk persists, especially in patients with advanced fibrosis even after viral elimination[12-17]. The lack of reliable biomarkers to predict HCC risk after viral cure poses a clinical challenge[14-16]. Despite significant strides, the molecular basis of HCV-induced liver disease and HCC development is still only partially understood[16,17]. Several studies suggest a role for HCV-induced transcriptional reprogramming in the development of HCC[18-21], however, the molecular drivers and mechanisms remain unknown. Here, we aimed to investigate the role of the hepatic clock in liver disease biology by identifying the circadian transcriptome and epigenome of human hepatocytes and its perturbation in chronic HCV infection.

Results

Comparative analysis of the rhythmic transcriptome in human and mouse hepatocytes.

Assessing the rhythmicity of the liver transcriptome in humans is challenging as it would require multiple biopsy samples over a 24 h period which poses a non-acceptable risk for the patient. We therefore used a human liver chimeric mouse (HLCM) model as a surrogate. These mice are immunodeficient and engrafted with primary human hepatocytes (PHH) and it is well established to recapitulate critical aspects of patient liver disease biology including chronic viral and metabolic liver disease[20,22-24]. To identify the genes displaying a rhythmic expression pattern in human hepatocytes, we investigated temporal changes in transcript abundance in liver tissue samples from HLCM (Fig. 1a). Hence, 3 HLCM were sacrificed every 4 hours throughout the 24-hour period starting at *Zeitgeber* 0 (ZT0; time of light ON, and ZT12- time of light OFF). We performed 2 independent animal experiments (Series 1 and 2; Supplementary Tables 1 and 2) to map the human hepatic diurnal transcriptome. Measurement of human serum albumin levels indicated the comparable levels of hepatocyte engraftment (degree of humanization) in both experiments (Series 1: ~14203 µg/mL, and Series 2: ~14973 µg/mL; Supplementary Table 1). This was further confirmed by comparing human vs. mouse RNA-sequencing (RNA-seq) reads (Supplementary Table 2) and human hepatocyte-specific CK8-18 immunostaining (Extended Data Fig. 1) suggesting ~65% to ~70% humanization of the chimeric livers. Importantly, hematoxylin-eosin (HE) staining revealed a conserved lobular hepatic architecture in HLCM with no sign of engraftment-induced stress or steatosis (Extended Data Fig. 1). To assess the temporal changes in the liver transcriptome of HLCM RNA-sequencing (RNA-seq) was performed, which produced an average of 35 million reads per sample (Supplementary Table 2). To distinguish transcripts originating from human and mouse tissue, sequence reads were mapped and annotated

to both the human and mouse genome (Methods, and Supplementary Table 2). Post-annotation unsupervised clustering was performed to determine the biological heterogeneity amongst HLCM samples with respect to the expression phase of all the core CC-oscillator genes and CC-output regulatory transcription factors in human hepatocytes (Extended Data Fig. 2a, b). This unbiased approach allowed us to choose 2 samples from each timepoint from Series 1 and all 3 samples for each timepoint from Series 2 for subsequent evaluations (Extended Data Fig. 2a, b). One sample of Series 1 had to be excluded because of insufficient reads (Extended Data Fig. 2a; indicated in red). Taken together, we analyzed 5 HLCM per timepoint to ensure a robust data set for further analysis of diurnal transcriptome in human hepatocytes.

The rhythmicity of CC-gene expression in the chimeric livers were analyzed by MetaCycle[25]. Notably, core CC-genes in engrafted human hepatocytes from both experiments showed comparable oscillation profiles to residual murine hepatocytes and showed the characteristic rhythmicity of these genes (Series 1 vs. Series 2 in Fig. 1b, and merged Series 1 and 2 in Extended Data Fig. 3a)[1-4]. These comprise the RORE-containing genes *CRY1*, *BMAL1*, *NPAS2*, *CLOCK*, *NFIL3*, which reach maximal expression during active phase (ZT12-ZT0) and minimal levels during rest phase. In contrast, BMAL1/CLOCK-transcriptional activity dependent E-Box containing genes *DBP*, *TEF, PER1,* and *PER2* are mostly expressed during rest phase (Fig. 1b and Extended Data Fig. 3a). Next, we compared the expression of CC-genes between HLCM liver and in the liver of wild-type (WT) mice[6,26] which suggested their phase advancement in the transplanted liver (Extended Data Fig. 3b). Importantly, our analyses of the expression pattern of the CC-genes in engrafted human hepatocytes, residual murine cells, and their comparison with the WT mice liver clock (Fig. 1b and Extended Data Fig. 3a, b) is consistent with recent observations[27]. Next, we applied the recently developed algorithm *dryR*[26] to analyze

differential rhythmicity of gene expression in HLCM liver. This algorithm determines variations of amplitude (fold changes), phase (peak expression time) and mean expression levels of circadian orthologous gene expression, allowing the comparison of datasets with multiple conditions and between species. The *dryR* algorithm distributes groups of genes to different models of rhythmicity. Through *dryR* we identified five distinct categories of transcriptional profile in HLCM livers that include genes with: cycling in human hepatocytes (model 2; colored green in the cartogram), cycling in mouse hepatocytes (model 3; pink in the cartogram), unaltered rhythm [model 4, grey: merger of pink (mouse) and green (human)], and altered rhythm (model 5; diverged expression between two species either in phase or amplitude while displaying some rhythmicity) (Fig. 1c and Supplementary Table 3). Additionally, a fifth model includes non-rhythmic genes in both species (model 1; Supplementary Table 3). *dryR* revealed ~1,700 rhythmic protein coding orthologous genes in HLCM liver of which 824 were uniquely rhythmic for human hepatocytes (model 2; Fig. 1c and Supplementary Table 3). We also found 103 genes (model 5) whose expression was altered between human and mouse hepatocytes, while 749 genes (model 4) showed unaltered rhythmic expression in two species in HLCM liver (Fig. 1c and Supplementary Table 3). Comparing cycling genes in human hepatocytes of HLCM with that of WT mice livers[6] and post-mortem human livers[28] revealed rhythmic genes which are shared (Extended Data Fig. 4a, b and Supplementary Table 3). However, the 'number' of shared genes were not statistically (hypergeometric test) significant probably due to (i) whole liver tissue contains cells other than hepatocytes, and (ii) using *dryR* we analyzed only protein coding genes (with known murine orthologues) while different categories of non-coding genes are also known to have diurnal rhythmicity in whole liver tissue.

Next, a classification of rhythmic genes in human hepatocytes of HLCM for their biological function using the Molecular Signature database (MSigDB)[29,30] allowed a comparison with known 'clock'-controlled physiological processes reported in the mouse liver[6,26] (Fig. 1d and Extended Data Fig. 5a, b). Temporally resolved enrichment analysis revealed an overwhelming role for the hepatic clock to regulate pathways involved in leptin, bile acid, cholesterol, fatty acid, and heme metabolism. In addition, we found that human rhythmic pathways involved in stress response (DNA repair, reactive oxygen species, p53 signaling), signaling pathways (NOTCH, KRAS, TGF-β), and cell cycle clustered in comparable temporal windows ('peak' expression ± 1 ZT) as seen in mice (Fig. 1d and Extended Data Fig. 5a, b). Human hepatocytes showed a rhythmic expression of cellular processes (apoptosis, hypoxia response), cytokine signaling (IFNα, IFNγ, IL6, TNFα, complement system), VEGF (angiogenesis) and unfolded protein response which operate at different temporal windows compared to murine hepatocytes. Consistent with a recent study[27], we also observed that some of the key genes involved in growth hormone (GH)-STAT5[31] and mTOR signaling pathways, e.g., *STAT1*, *STAT3*, *STAT5B*, *MTOR*, *LAMTOR1-5* were non-rhythmic (model 1) in transplanted human hepatocytes (Supplementary Table 3). Taken together, our study identifies the diurnal transcriptome and rhythmic physiological processes in human hepatocytes.

Diurnal regulation of transcription factors and enhancer regions impact pathways relevant for liver homeostasis.

Rhythmic expression of transcription factors (TFs) can generate temporal patterns of target gene expression. To identify rhythmically expressed TFs in human hepatocytes of HLCM we analyzed a dataset containing 1600 human TFs[32]. Analyses of *dryR*-classified rhythmic genes led to the

identification of ~140 rhythmically expressed TFs (~8% of all rhythmic genes) in human hepatocytes of HLCM livers (Fig. 1e and Extended Data Fig. 6, 7a, b) that are well-known regulators of physiopathology. Evaluating the expression of TFs in human and mouse cells of HLCM revealed that key TFs like *IRF2*, *NCOR2*, *JUNB*, *RELB*, and *IRF1* were uniquely rhythmic in human hepatocytes (model 2; Fig. 1f, Extended Data Fig. 6 and Supplementary Table 3), correlating with the pathway analysis indicating that several of the cytokine and stress response pathways are temporally distinct in human and mouse livers (Fig. 1d). We found shifted expression phase of TFs regulating xenobiotic metabolism (*NR1I2*, and *HLF;* Extended Data Fig. 6, and 7a) and altered expression (model 5) for *CEBPG* (Fig. 1e). Importantly, amongst rhythmically expressed TFs in both species in HLCM liver contained CC-regulatory TFs (*BMAL1/ARNTL*, *CLOCK*, *NR1D1*), and orchestrators of metabolism (*PPARD*, *CREB3L3*). In HLCM livers, we identified several human orthologues of zinc finger TFs (ZNFs) being expressed in different *dryR* categories, e.g., *ZNFs* (*217*, *248*, and *318*; model 2 - *only cycling in human hepatocytes*), while *ZNFs* (*330*, *277*, *362* and *367*, model 4 - *unaltered in two species*) (Extended Data Fig. 6 and Supplementary Table 3).

Gene expression requires epigenetic changes at promoter and enhancer regions. To determine whether these rhythmicity in transcript levels are associated with epigenetic remodeling of the chromatin, we performed ChIP-sequencing (ChIP-seq) of samples from the HLCM liver (Fig. 1a). We initially focused on the temporal variation in histone 3 lysine 27 acetylation (H3K27ac) levels an established marker of promoter and enhancer activation[33,34], and whose deregulation is related to chronic liver disease[19,35,36]. Comparative analysis revealed the diurnal variation of H3K27ac levels in enhancer and promoter regions in human hepatocytes and in the liver of WT mice[6]. Analyzing the enhancer activation pattern showed that the changes for

CC output regulatory TF D-box binding protein (*DBP)* follows a similar pattern in human and murine hepatocytes, while variations in H3K27ac levels surrounding *IRF2* promoter-enhancer were only rhythmic in the human hepatocytes (Fig. 1g and Extended Data Fig. 8).

Taken together, our results provide an atlas of human hepatocellular diurnal transcriptome, unveiling its epigenomic variations *in vivo.* Importantly, altered expression of several of the identified rhythmic genes and processes are reported to associate with chronic liver disease[19,20,35].

HCV infection perturbs the rhythmicity of human liver transcriptome in vivo.

To understand whether perturbation of the human hepatocellular rhythmic transcriptome associates with liver disease and cancer, we investigated the impact of HCV infection on diurnal gene expression in two independent animal experiments (Series 1 and 2; Supplementary Tables 1, 2). HCV exclusively infects human hepatocytes[12-14] and does neither infect nor replicates in murine cells[22,24,37]. HLCM were infected with patient-derived HCV for 10 weeks (Fig. 2a and Supplementary Table 1) and assessed for hepatic diurnal gene expression. RNA-seq from the HCV infected HLCM livers (3 HLCM / experiment / circadian timepoint) yielded specific transcriptomic profiles that clustered in an unsupervised manner (Supplementary Table 2 and Extended Data Fig. 9a, b). While humanization of the mouse livers in control and HCV animals were similar at the time of infection according to human albumin levels in the sera, we observed a slightly reduced humanization in the HCV animals after 4th week of infection (Extended Data Fig. 10c) as described previously in a similar model[24] (Extended Data Fig. 10a-c and Supplementary Table 1). Loss of human hepatocytes is likely due to an induction of cellular stress

in HCV-infected livers[19,20,23,38], which was also reflected histologically by CK8-18 staining. Importantly, we next employed a species-specific 'reads' mapping strategy which allows for an independent analysis of human and mouse transcriptome (see methods). Additionally, this method of analyzing human (and mouse) reads independent from each other also takes into account artifacts arising from different humanization levels in different groups of HLCM.

Due to the restricted tropism of HCV to human hepatocytes, we studied the impact of HCV-infection on diurnal transcriptome only in human hepatocytes (control and infected) of HLCM. Notably, analyzing two independent animal infection experiments showed overall perturbation in the expression of key CC genes (Fig. 2b and Extended Data Fig. 11a). The 'peak' expression for *BMAL1* (ZT20) was consistently but modestly reduced (FDR<0.067), which correlated with a reduced maximal level (ZT4) of its target gene, transcription factor *DBP* (Fig. 2b, Extended Data Fig. 11a, b). However, other key CC genes (*CLOCK*, *PER1* and *REV-ERBα/NR1D1*) displayed only overall tendencies of deregulation in infected hepatocytes (Fig. 2b, Extended Data Fig. 11a). Also, between ZT4-ZT8 the expression level of *REV-ERBα* target gene *NFIL3* was increased in HCV-infected livers (Fig. 2b, Extended Data Fig. 11a, b). Importantly, the *dryR* analyses of HCV infection (see below) confirmed that two key CC TFs (*BMAL1/ARNTL* and *NFIL3*) belong to the 'altered' group of genes (model 5; Supplementary Table 4). Taken together, these analyses indicate that by perturbing the expression of circadian TFs, HCV infection likely impacts the global gene expression of their downstream targets, e.g., BMAL1-activated (E-Box-containing), and NFIL3-repressed (D-Box-containing) genes. Interestingly, we found that serum HCV RNA levels in HLCM displayed a modest diurnal rhythmicity with a peak in the active phase (ZT16-ZT20). However, the clinical translation of this finding remains to be explored (Extended Data Fig. 12).

Next, in the human hepatocytes of HLCM we investigated HCV-induced perturbation of the entire rhythmic transcriptome and pathways using *dryR*[26]. Similar to Fig. 1c, we identified five distinct categories of transcriptional profile in HCV-infected HLCM livers (loss, gain, unaltered, and altered rhythm (Fig. 2c and Supplementary Table 4), plus one additional category of non-rhythmic genes in comparison to non-infected controls. The global impact of HCV-induced perturbation of the hepatocellular diurnal transcriptome (~22% of all the rhythmic genes) was reflected by the deregulation of several physiological processes (Fig. 2d and Extended Data Fig. 13), including pathways of HCV-induced proteogenomic changes[20,39,40]. Importantly, pathways (enriched for cycling genes; FDR <0.05) being significantly dysregulated comprise key drivers of chronic liver disease, including metabolic alterations (fatty acids, lipids, peroxisome organization), fibrosis (TGFβ-signaling, SMAD activity, fibroblast proliferation, and EMT response), and oncogenic pathways linked to HCC development (liver cancer signatures, MYC, H-RAS, and EGFR signaling) (Fig. 2d, Extended Data Fig. 13). All these pathways were previously determined to be cycling by dryR (Fig. 1d) within model 2 (loss of rhythmicity) (Fig. 2d). Although HLCM lack T- and B-cells (adaptive immunity deficient), engrafted hepatocytes can activate cell intrinsic innate immunity-related pathways[19,20], which may act to defend against stress and infection. Consistently, HCV-infected HLCM liver showed an upregulation of innate immune and inflammatory pathways (NF-κB, TNFα, IL6/STAT3, and Type I interferons) (Fig. 2d, Extended Data Fig. 13). Next, we performed immunostaining of MYC a well-known oncogene driving various aspects of chronic liver disease and HCC. While MYC was barely detectable in the uninfected liver, it was significantly (p=0.015) upregulated upon HCV infection in HLCM liver supporting activation of downstream target pathways at the transcriptomic level (Fig. 2d and Extended Data Fig. 14a-c). Several studies have noted that the

lack of adaptive immune cells prevents the development of steatosis and that only marginal fibrosis develops in HLCM liver[38,41,42]. Consistently, histology of HLCM liver did not detect any steatosis (Extended Data Fig. 15a). Sirius red staining and quantification of collagen positive areas (CPA) showed modest but significant liver fibrosis in the infected mice (Extended Data Fig. 15b, c). In summary, these data shows that HCV infection dysregulates the diurnal liver transcriptome and rhythmic processes, which likely predisposes to the advancement of chronic liver disease.

HCV infection disrupts rhythmic epigenetic variations that drive gene expression in vivo.

The diurnal variation of epigenomic changes driving promoter-enhancer activation/repression in health and chronic liver disease in human hepatocytes is unexplored. Hence, we profiled 24-hour changes in two epigenetic markers of gene activation: histone 3 lysine 27 acetylation (H3K27ac) and histone 3 lysine 9 acetylation (H3K9ac) in liver obtained from infected and non-infected HLCM. We identified global rhythmicity in both H3K27ac and H3K9ac levels in promoter-enhancer regions of cycling genes in uninfected livers (Fig. 3a, d and Supplementary Table 3). Amongst cycling genes, we identified transcription start site (TSS)-enriched peaks for both H3K27ac and H3K9ac. Moreover, we also found peaks of H3K27ac and H3K9ac within gene bodies and intergenic regions that suggests possible diurnal epigenetic regulation of yet to be characterized distal enhancer regions (Fig. 3a-f and Supplementary Table 5). Importantly, overall circadian variation in H3K27ac was completely lost in the infected liver. The loss of H3K27ac variation emerged as a generalized plateau of TSS-associated H3K27ac peaks (Fig. 3a-c). Also, the observed circadian intergenic enhancer regions were persistently hyper-H3K27 acetylated in the infected liver (Fig. 3a). In contrast to H3K27ac, HCV-induced higher H3K9ac levels mostly from the end of the rest phase and during the active phase that caused a saturation of peak numbers

(ZT8-ZT20) when compared to the non-infected animals (Fig. 3d-f). Thus, HCV alters circadian epigenome at both promoter-enhancer (H3K27ac and H3K9ac), and gene body (H3K9ac) levels (Fig. 3a-f). In addition, analyses of gene-specific circadian variations in levels of H3K27ac and H3K9ac of liver disease driving genes (*XBP1* and *RAF1*) and CC-component *CRY1* confirmed the HCV-induced epigenetic perturbations (Fig. 3g, Extended Data Fig. 16a, b). To unravel the TF motifs most enriched in H3K27ac peaks in HCV-infected HLCM livers, we intersected TSS and enhancer-associated peaks with TF binding site predictions from JASPAR[43,44] (Fig. 3h). This revealed an overall suppression of H3K27ac levels in the regulatory regions of genes controlled by CC-output regulatory TFs (e.g., DBP, HLF, TEF, NFIL3; regulators of D-Box genes) correlating with an overall dysregulation of their target pathways (bile acid metabolism, xenobiotic metabolism; Extended Data Fig. 13). Importantly, we also observed an enrichment of H3K27ac peaks in enhancer regions for known transcriptional drivers of chronic liver disease progression in the HCV-infected HLCM, e.g., NRF1/NFE2L1 (regulator of lipid metabolism, oxidative/hypoxic stress), E2Fs and EGRs (regulator of cell cycle, stress response) (Fig. 2d and 3h, Extended Data Fig. 13).

Histone acetylases and deacetylases are known enzymatic regulators of histone acetylation. In HLCM, no significant transcriptional change of key histone acetylases (*EP300,* and *CREBBP)* or deacetylases of the *HDAC* or *SIRTUIN* families were observed upon HCV-infection (Fig. 3j, Extended Data Fig. 17a, b). Notably, *dryR* identified *SIRT1* to be uniquely rhythmic in human hepatocytes (model 2, Supplementary Table 3), while the expression of all other *SIRTUINs* (*SIRTs 2-7*) showed no rhythmicity in control HLCM livers (Fig. 3i, Extended Data Fig. 17c and Supplementary Table 3). Our results demonstrate that HCV infection impairs

the hepatic transcriptome and epigenome to create an environment that is conducive to the development of liver disease.

The HCV non-structural (NS) proteins 4 and 5A suppress SIRT1 and BMAL1 protein expression in an infectious cell culture model.

SIRT1 is the only epigenetic regulator which was identified by *dryR* to display diurnal rhythmicity in human hepatocytes of chimeric mice and is a known regulator of CC-oscillator in mice[45]. Given the technical challenges associated of performing specific loss-of-function studies in human hepatocytes of the chimeric mice, we utilized the state-of-the art cell culture system of HCV infection[46,47] (previously used to discover the essential biological process involved in HCV replication and viral pathogenesis as well as the clinical antiviral therapy)[13,19,20,35] to investigate potential mechanistic links between *SIRT1* and key CC gene *BMAL1*.

For loss-of-function studies, we established CRISPR/Cas9 guided control (*sgCTRL*) and *SIRT1* (*sgSIRT1*) knock-out (KO) in human hepatocyte-derived Huh7.5.1 cells (Fig. 4a). In *sgSIRT1* cells, BMAL1 protein was significantly reduced compared to the controls (Fig. 4a and Extended Data Fig. 20). SIRT1 regulates *BMAL1* transcription as in *sgSIRT1* cells the "basal" level of *BMAL1* promoter driven luciferase (*BMAL1-Luc*) activity was significantly reduced (Fig. 4b). Consistently, SIRT1 agonist (SRT2183)[46] increased the "basal" level of *BMAL1-Luc* activity in Huh7.5.1 cells (Fig. 4c) and this effect was abolished in *SIRT1* KO cells (Fig. 4c). These results suggest a potential functional role for SIRT1 to regulate BMAL1 expression in the HCV cell culture model.

Next, we explored the role of individual HCV proteins in the expression of SIRT1 and BMAL1 in Huh7.5.1 cells, by ectopically expressing viral-encoded structural (core: C and core-envelope: CE2) and non-structural (NS3, NS4, NS5A and NS5B) proteins. We confirmed that

HCV infection significantly reduced SIRT1 and BMAL1 expression in Huh7.5.1 cells (Fig. 4d and Extended Data Fig. 21). HCV NS4 and NS5A reduced the levels of both SIRT1 and BMAL1 (Fig. 4d and Extended Data Fig. 21). We investigated which HCV protein affects the expression of both *SIRT1* and *BMAL1* transcriptionally by determining their promoter-luciferase activity. We found that of all the HCV proteins only NS4 and NS5A significantly reduce the promoter activities of both *SIRT1-Luc* and *BMAL1-Luc* (Fig. 4e, f), suggesting that decreased *SIRT1* (deacetylase) levels not only leads to a reduction in *BMAL1* levels, but also likely contributes (along with other chromatin regulators) to the HCV-induced epigenomic alterations in the cell culture model.

HCV perturbation of the diurnal gene expression is associated with HCC risk in patients with chronic liver disease.

To investigate the clinical impact of these findings, we investigated the virus-induced perturbation of the rhythmic pathways in patients with chronic hepatitis C. Our analysis of transcriptomic changes in these patients broadly supports our observations in HLCM by revealing that HCV perturbs the expression liver disease-relevant rhythmic pathways (Fig. 5a), as shown in a side-by-side comparison in this well characterized cohort of pooled transcriptomic data from chronically infected patients (F1-F4 stage liver disease) and control subjects without liver disease[14] (Fig. 5a). Most of the liver specimens in these patient cohorts were obtained in the morning, however, we cannot exclude a moderate temporal heterogeneity due to a variation in clinical schedules.

Next, we studied whether DAA cure results in reversal of transcriptomic changes of the evaluated pathways in the HLCM mouse model (HCV-infected vs DAA-treated; Supplementary Table 7)

as well as in HCV-infected and viral-cured patients (Pat 1,2, and 3; Supplementary Table 6), which revealed that the virus-induced changes in the human transcriptome are only partially reversed following viral clearance (Extended Data Fig. 18a). Importantly, an independent analysis of these disease relevant pathways either in viral-cured patients with HCC (Supplementary Table 6) or with MASH showed their perturbation (Extended Data Fig. 18b, c).

Finally, we aimed to study whether the perturbation of rhythmic gene expression was associated with HCC risk in patients. Notably, using we have previously shown that HCV-induced epigenetic changes are associated with a persistent HCC risk post viral cure[35]. The effect on carcinogenic pathways is reflected by the perturbation of the well characterized prognostic liver signature (PLS)[21,39,40], which robustly predicts survival, liver disease progression and HCC risk in patients with different etiologies including chronic hepatitis C. The PLS was perturbed to a HCC high risk status during chronic HCV infection. A systematic dissection of the 186 PLS genes revealed that genes associated with a low HCC risk and good prognosis are cycling into a much higher proportion compared to the expression of HCC high risk/poor prognosis genes (Fig. 5b). Thus HCV-induced CC-perturbation seems to affect the expression good prognosis genes much more than the poor prognosis genes (Extended Data Fig. 19).

Next, we aimed to identify a gene signature that reflects HCV-induced perturbation of CC genes and at the same time predict patient outcome. Therefore, we assessed whether diseased human livers express the various *dryR* gene sets we identified in the HCV-infected HLCM liver (Fig. 2c). We compared the perturbed rhythm associated with disease phenotype by using our previously described genome-wide transcriptome profiles of liver tissues from 216 early-stage (Child-Pugh class A) cirrhosis patients[18]. We defined gene sets of HCV-infected and control hepatocytes for early (ZT0-ZT8) and late (ZT12-ZT20) time points, as up-regulated genes with

fold changes >4 in each condition. We observed correlation of some of the gene sets from early time points of HCV-infected hepatocytes with several disease severity-related features, e.g., severer (larger) gastroesophageal varices and high-risk pattern of the PLS[18,21,40,48-51] (Fig. 5b, Supplementary Tables 8, 9). In contrast, the gene sets derived from the control and late time points in HCV-infected hepatocytes were associated with low-risk pattern of the PLS. Using these gene sets as a common signature termed HCVclock, we were able to classify the HCC patients into high-risk (poor outcome) and low-risk (good outcome) groups. Importantly, we observed changes in gene set enrichment index (GSEI) which indicates "magnitude of correlation" with overall survival according to HCVclock classification (Fig. 5c, d), suggesting possible clinical relevance. Finally, our analyses of HCV-disrupted rhythmic pathways in HLCM and patient cohorts revealed a firm link with all major 'hallmarks' of cancer (Fig. 5e). Collectively, our analyses demonstrate that the HCV-induced perturbation of the human hepatocellular diurnal transcriptome correlates with clinical risk of developing HCC.

Discussion

The liver is the central metabolic organ in mammals and performs evolutionary conserved metabolic processes. However, given an evolutionary distance of ~80 million years with nocturnal rodents in which the CC has been extensively studied, it is important to identify genes and corresponding pathways that are rhythmic in human tissues, and hepatocytes, and which are deregulated in liver disease. Our study provides a comprehensive temporal gene expression atlas and identifies epigenomic changes in human hepatocytes that facilitates bioinformatic analyses to study their potential role in liver disease. The abundance and complexity of transcripts in a multifunctional tissue like the liver does not explain all dimensions of its physiology, but the

study of temporal gene expression patterns as provided here offers a new and crucial dimension of understanding human liver in health and disease. It has generally been assumed, given the conservation of biological processes across mammals, that diurnal transcriptomes are comparable across species while expressing mostly orthologous rhythmic genes. However, this investigation identifies crucial genes (TFs, epigenetic remodelers, enzymes etc.) and pathways which show rhythmic expression only in human hepatocytes (Fig. 1 and Extended Data Fig. 5-7). Critically, many of these pathways were recently identified as human hepatocyte specific[27]. We found that the CC genes in engrafted human hepatocytes and resident murine hepatocytes display comparable patterns of expression (Fig. 1b). This effect was recently confirmed and suggested to be due to a dominant cellular 'clock' synchronizing ability of engrafted human hepatocytes[27]. Together, our investigation and that of Delbés and co-workers[27] confirm that the engraftment of human hepatocytes into a new environment (neuronal and systemic signals, food, light cycle, temperature) did not prevent the re-establishment of a functioning rhythmic hepatic CC-oscillator. The HLCM model is well established to investigate viral and metabolic chronic liver disease. However, the CC-oscillator in engrafted human hepatocytes was phase-advanced when compared to the WT mice liver[27]. This is consistent with the literature in the field suggesting that this being due to the differential sensitivity of human hepatocytes to metabolic and systemic signaling[27,52,53]. It should be noted that HLCM are immunodeficient, and transcripts of liver resident immune cells cannot be studied extensively using such animal models. However, the engrafted hepatocytes have an intrinsic capacity to induce innate immune responses to stress and infection, as evidenced by their expression of several cytokine (TNF, IL6, and type I interferons) signaling pathways (Fig. 2d). Physiologically, the CC governs the expression of numerous critical genes controlled directly by the CC-components (e.g., *BMAL1*, *DBP*, *TEF*), or through CC-regulated rhythmic TFs

(Fig. 1). Even though HCV-infection perturbed the expression of only a few CC genes (Fig. 2b), this perturbation was associated with a dysregulation of the transcriptomic oscillation of 22% of rhythmic genes (loss, gain, and altered rhythmicity) (Fig. 2 and Extended Data Fig. 13). Importantly, previous studies demonstrated that viral infections and chronic liver disease is driven through similar perturbed pro-oncogenic pathways in various disease etiologies[19,21,40,54,55].

Using chronic hepatitis C as a model disease, our data indicate a critical role of the dysregulated hepatocellular CC for virus-induced cancer as shown by perturbation of the oscillation of the major cancer hallmark pathways in both HLCM and patients (Fig. 2, 3, 5). Mechanistic studies identified a previously unrecognized epigenetic imprinting (dysregulated histone acetylation) in enhancer regions of circadian genes, that correlated with impairments of cancer hallmark pathways upon viral infection. A strength of our study is the use of two state-of-the-art model systems for HCV infection with data integration of liver tissues of HCV-infected patients, a technical limitation of the study is that these model systems still only partially mimic the cell circuits in patients and a diurnal analyses of the patient liver transcriptome and proteome would require serial biopsies over a 24 hour time period which are not available due to ethical constraints. Our analyses obtained in a state-of-the-art cell culture model for HCV infection suggest a role of the NAD^+-dependent deacetylase SIRT1[56] for HCV-induced CC perturbation. In the cell culture model, HCV reduced SIRT1 protein and BMAL1 levels (Fig. 4) suggesting virus-induced perturbed SIRT1 and BMAL1 expression as candidate mechanism contributing to the virus-induced changes of the rhythmicity of part of the transcriptome. Of note, the regulatory role of SIRT1 on BMAL1 has been previously established in mice[45,57,58]. Supporting this hypothesis, we observed a perturbed expression of BMAL1 targets in HCV-infected HLCM livers including TFs like DBP and several other genes (Extended Data Fig. 11b and Supplementary

Table 4). While these data may suggest a virus-induced BMAL1-associated suppression in infected HLCM, additional *in vivo* validation experiments would be required for arresting conclusions. Interestingly, we also observed an HCV-induced perturbation of the promoter-enhancer associated histone mark (H3K27ac) in vivo while gene-body associated H3K9ac was only disrupted during ZT8-ZT20 (Fig. 3a-h), indicating a dynamic act at the epigenomic level between different chromatin remodelers regulating gene expression. Taken together, these alterations in different TFs, chromatin modifiers and their downstream targets suggest intricate mechanisms that could largely explain HCV-induced deregulation of the cellular transcriptome that collectively have the potential to activate pathways known to drive liver disease (Fig. 2-4 and Extended Data Fig. 20).

Moreover, our findings of gene expression in patients suggest a profound effect of HCV infection on HCC risk biomarkers (i.e., PLS) (Fig. 5b-d). As, CC-targeting compounds[2] have been shown to attenuate liver disease progression in rodent models[59], our results may provide opportunities to discover CC-based biomarkers to predict HCC risk and novel therapeutic targets for cancer prevention.

References

1 Asher, G. & Sassone-Corsi, P. Time for food: the intimate interplay between nutrition, metabolism, and the circadian clock. *Cell* **161**, 84-92 (2015). https://doi.org:10.1016/j.cell.2015.03.015

2 Mukherji, A., Bailey, S. M., Staels, B. & Baumert, T. F. The circadian clock and liver function in health and disease. *J Hepatol* **71**, 200-211 (2019). https://doi.org:10.1016/j.jhep.2019.03.020

3 Panda, S. Circadian physiology of metabolism. *Science* **354**, 1008-1015 (2016). https://doi.org:10.1126/science.aah4967

4 Takahashi, J. S. Transcriptional architecture of the mammalian circadian clock. *Nat Rev Genet* **18**, 164-179 (2017). https://doi.org:10.1038/nrg.2016.150

5 Fang, B. *et al.* Circadian enhancers coordinate multiple phases of rhythmic gene transcription in vivo. *Cell* **159**, 1140-1152 (2014). https://doi.org:10.1016/j.cell.2014.10.022

6 Koike, N. *et al.* Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. *Science* **338**, 349-354 (2012). https://doi.org:10.1126/science.1226339

7 Scheiermann, C., Gibbs, J., Ince, L. & Loudon, A. Clocking in to immunity. *Nat Rev Immunol* **18**, 423-437 (2018). https://doi.org:10.1038/s41577-018-0008-4

8 Masri, S. & Sassone-Corsi, P. The emerging link between cancer, metabolism, and circadian rhythms. *Nature medicine* **24**, 1795-1803 (2018). https://doi.org:10.1038/s41591-018-0271-8

9 Sun, M. *et al.* Night shift work exposure profile and obesity: Baseline results from a Chinese night shift worker cohort. *PLoS One* **13**, e0196989 (2018). https://doi.org:10.1371/journal.pone.0196989

10 Verlande, A. & Masri, S. Circadian Clocks and Cancer: Timekeeping Governs Cellular Metabolism. *Trends Endocrinol Metab* **30**, 445-458 (2019). https://doi.org:10.1016/j.tem.2019.05.001

11 Wang, X. S., Armstrong, M. E., Cairns, B. J., Key, T. J. & Travis, R. C. Shift work and chronic disease: the epidemiological evidence. *Occup Med (Lond)* **61**, 78-89 (2011). https://doi.org:10.1093/occmed/kqr001

12 Baumert, T. F., Berg, T., Lim, J. K. & Nelson, D. R. Status of Direct-Acting Antiviral Therapy for Hepatitis C Virus Infection and Remaining Challenges. *Gastroenterology* **156**, 431-445 (2019). https://doi.org:10.1053/j.gastro.2018.10.024

13 Chung, R. T. & Baumert, T. F. Curing chronic hepatitis C--the arc of a medical triumph. *N Engl J Med* **370**, 1576-1578 (2014). https://doi.org:10.1056/NEJMp1400986

14 Boldanova, T., Suslov, A., Heim, M. H. & Necsulea, A. Transcriptional response to hepatitis C virus infection and interferon-alpha treatment in the human liver. *EMBO Mol Med* **9**, 816-834 (2017). https://doi.org:10.15252/emmm.201607006

15 El-Serag, H. B., Kanwal, F., Richardson, P. & Kramer, J. Risk of hepatocellular carcinoma after sustained virological response in Veterans with hepatitis C virus infection. *Hepatology* **64**, 130-137 (2016). https://doi.org:10.1002/hep.28535

16 Kanwal, F. *et al.* Risk of Hepatocellular Cancer in HCV Patients Treated With Direct-Acting Antiviral Agents. *Gastroenterology* **153**, 996-1005 e1001 (2017). https://doi.org:10.1053/j.gastro.2017.06.012

17 van der Meer, A. J. *et al.* Risk of cirrhosis-related complications in patients with advanced fibrosis following hepatitis C virus eradication. *J Hepatol* **66**, 485-493 (2017). https://doi.org:10.1016/j.jhep.2016.10.017

18 Hoshida, Y. *et al.* Prognostic gene expression signature for patients with hepatitis C-related early-stage cirrhosis. *Gastroenterology* **144**, 1024-1030 (2013). https://doi.org:10.1053/j.gastro.2013.01.021

19 Juhling, F. *et al.* Targeting clinical epigenetic reprogramming for chemoprevention of metabolic and viral hepatocellular carcinoma. *Gut* **70**, 157-169 (2021). https://doi.org:10.1136/gutjnl-2019-318918

20 Lupberger, J. *et al.* Combined Analysis of Metabolomes, Proteomes, and Transcriptomes of Hepatitis C Virus-Infected Cells and Liver to Identify Pathways Associated With Disease Development. *Gastroenterology* **157**, 537-551 e539 (2019). https://doi.org:10.1053/j.gastro.2019.04.003

21 Nakagawa, S. *et al.* Molecular Liver Cancer Prevention in Cirrhosis by Organ Transcriptome Analysis and Lysophosphatidic Acid Pathway Inhibition. *Cancer Cell* **30**, 879-890 (2016). https://doi.org:10.1016/j.ccell.2016.11.004

22 Grompe, M. & Strom, S. Mice with human livers. *Gastroenterology* **145**, 1209-1214 (2013). https://doi.org:10.1053/j.gastro.2013.09.009

23 Mailly, L. *et al.* In vivo combination of human anti-envelope glycoprotein E2 and -Claudin-1 monoclonal antibodies for prevention of hepatitis C virus infection. *Antiviral Res* **162**, 136-141 (2019). https://doi.org:10.1016/j.antiviral.2018.12.018

24 Mailly, L. *et al.* Clearance of persistent hepatitis C virus infection in humanized mice using a claudin-1-targeting monoclonal antibody. *Nat Biotechnol* **33**, 549-554 (2015). https://doi.org:10.1038/nbt.3179

25 Wu, G., Anafi, R. C., Hughes, M. E., Kornacker, K. & Hogenesch, J. B. MetaCycle: an integrated R package to evaluate periodicity in large scale data. *Bioinformatics* **32**, 3351-3353 (2016). https://doi.org:10.1093/bioinformatics/btw405

26 Weger, B. D. *et al.* Systematic analysis of differential rhythmic liver gene expression mediated by the circadian clock and feeding rhythms. *Proc Natl Acad Sci U S A* **118** (2021). https://doi.org:10.1073/pnas.2015803118

27 Delbes, A. S. *et al.* Mice with humanized livers reveal the role of hepatocyte clocks in rhythmic behavior. *Sci Adv* **9**, eadf2982 (2023). https://doi.org:10.1126/sciadv.adf2982

28 Talamanca, L., Gobet, C. & Naef, F. Sex-dimorphic and age-dependent organization of 24-hour gene expression rhythms in humans. *Science* **379**, 478-483 (2023). https://doi.org:10.1126/science.add0846

29 Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425 (2015). https://doi.org:10.1016/j.cels.2015.12.004

30 Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739-1740 (2011). https://doi.org:10.1093/bioinformatics/btr260

31 Lu, M., Flanagan, J. U., Langley, R. J., Hay, M. P. & Perry, J. K. Targeting growth hormone function: strategies and therapeutic applications. *Signal Transduct Target Ther* **4**, 3 (2019). https://doi.org:10.1038/s41392-019-0036-y

32 Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* **172**, 650-665 (2018). https://doi.org:10.1016/j.cell.2018.01.029

33 Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat Rev Genet* **17**, 487-500 (2016). https://doi.org:10.1038/nrg.2016.59

34 Calo, E. & Wysocka, J. Modification of enhancer chromatin: what, how, and why? *Mol Cell* **49**, 825-837 (2013). https://doi.org:10.1016/j.molcel.2013.01.038

35 Hamdane, N. *et al.* HCV-Induced Epigenetic Changes Associated With Liver Cancer Risk Persist After Sustained Virologic Response. *Gastroenterology* **156**, 2313-2329 e2317 (2019). https://doi.org:10.1053/j.gastro.2019.02.038

36 Perez, S. *et al.* Hepatitis C virus leaves an epigenetic signature post cure of infection by direct-acting antivirals. *PLoS Genet* **15**, e1008181 (2019). https://doi.org:10.1371/journal.pgen.1008181

37 Ploss, A. & Kapoor, A. Animal Models of Hepatitis C Virus Infection. *Cold Spring Harb Perspect Med* **10** (2020). https://doi.org:10.1101/cshperspect.a036970

38 Bissig, K. D. *et al.* Human liver chimeric mice provide a model for hepatitis B and C virus infection and treatment. *J Clin Invest* **120**, 924-930 (2010). https://doi.org:10.1172/JCI40094

39 Crouchet, E. *et al.* A human liver cell-based system modeling a clinical prognostic liver signature for therapeutic discovery. *Nat Commun* **12**, 5525 (2021). https://doi.org:10.1038/s41467-021-25468-9

40 Hoshida, Y. *et al.* Gene expression in fixed tissues and outcome in hepatocellular carcinoma. *N Engl J Med* **359**, 1995-2004 (2008). https://doi.org:10.1056/NEJMoa0804525

41 Mercer, D. F. *et al.* Hepatitis C virus replication in mice with chimeric human livers. *Nature medicine* **7**, 927-933 (2001). https://doi.org:10.1038/90968

42 Washburn, M. L. *et al.* A humanized mouse model to study hepatitis C virus infection, immune response, and liver disease. *Gastroenterology* **140**, 1334-1344 (2011). https://doi.org:10.1053/j.gastro.2011.01.001

43 Stormo, G. D. Modeling the specificity of protein-DNA interactions. *Quant Biol* **1**, 115-130 (2013). https://doi.org:10.1007/s40484-013-0012-4

44 Wasserman, W. W. & Sandelin, A. Applied bioinformatics for the identification of regulatory elements. *Nat Rev Genet* **5**, 276-287 (2004). https://doi.org:10.1038/nrg1315

45 Chang, H. C. & Guarente, L. SIRT1 mediates central circadian control in the SCN by a mechanism that decays with aging. *Cell* **153**, 1448-1460 (2013). https://doi.org:10.1016/j.cell.2013.05.027

46 Lindenbach, B. D. *et al.* Complete replication of hepatitis C virus in cell culture. *Science* **309**, 623-626 (2005). https://doi.org:10.1126/science.1114016

47 Wakita, T. *et al.* Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nature medicine* **11**, 791-796 (2005). https://doi.org:10.1038/nm1268

48 Goossens, N. *et al.* Nonalcoholic Steatohepatitis Is Associated With Increased Mortality in Obese Patients Undergoing Bariatric Surgery. *Clin Gastroenterol Hepatol* **14**, 1619-1628 (2016). https://doi.org:10.1016/j.cgh.2015.10.010

49 King, L. Y. *et al.* A genomic and clinical prognostic index for hepatitis C-related early-stage cirrhosis that predicts clinical deterioration. *Gut* **64**, 1296-1302 (2015). https://doi.org:10.1136/gutjnl-2014-307862

50 Ono, A. *et al.* Persisting risk of hepatocellular carcinoma after hepatitis C virus cure monitored by a liver transcriptome signature. *Hepatology* **66**, 1344-1346 (2017). https://doi.org:10.1002/hep.29203

51 Qian, T. *et al.* Molecular Signature Predictive of Long-Term Liver Fibrosis Progression to Inform Antifibrotic Drug Development. *Gastroenterology* **162**, 1210-1225 (2022). https://doi.org:10.1053/j.gastro.2021.12.250

52 Cao, R. mTOR Signaling, Translational Control, and the Circadian Clock. *Front Genet* **9**, 367 (2018). https://doi.org:10.3389/fgene.2018.00367

53 Jouffe, C. *et al.* The circadian clock coordinates ribosome biogenesis. *PLoS Biol* **11**, e1001455 (2013). https://doi.org:10.1371/journal.pbio.1001455

54 Zhuang, X., Edgar, R. S. & McKeating, J. A. The role of circadian clock pathways in viral replication. *Semin Immunopathol* **44**, 175-182 (2022). https://doi.org:10.1007/s00281-021-00908-2

55 Zhuang, X. *et al.* Circadian control of hepatitis B virus replication. *Nat Commun* **12**, 1658 (2021). https://doi.org:10.1038/s41467-021-21821-0

56 Feng, S. *et al.* Regulation of HepG2 cell apoptosis by hepatitis C virus (HCV) core protein via the sirt1-p53-bax pathway. *Virus Genes* **51**, 338-346 (2015). https://doi.org:10.1007/s11262-015-1253-2

57 Asher, G. *et al.* SIRT1 regulates circadian clock gene expression through PER2 deacetylation. *Cell* **134**, 317-328 (2008). https://doi.org:10.1016/j.cell.2008.06.050

58 Bellet, M. M. *et al.* Pharmacological modulation of circadian rhythms by synthetic activators of the deacetylase SIRT1. *Proc Natl Acad Sci U S A* **110**, 3333-3338 (2013). https://doi.org:10.1073/pnas.1214266110

59 Ni, Y. *et al.* Pharmacological activation of REV-ERBalpha improves nonalcoholic steatohepatitis by regulating intestinal permeability. *Metabolism* **114**, 154409 (2021). https://doi.org:10.1016/j.metabol.2020.154409

60 Tateno, C. *et al.* Generation of Novel Chimeric Mice with Humanized Livers by Using Hemizygous cDNA-uPA/SCID Mice. *PLoS One* **10**, e0142145 (2015). https://doi.org:10.1371/journal.pone.0142145

61 Katsuda, T. *et al.* Generation of human hepatic progenitor cells with regenerative and metabolic capacities from primary hepatocytes. *Elife* **8** (2019). https://doi.org:10.7554/eLife.47313

62 Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**, 907-915 (2019). https://doi.org:10.1038/s41587-019-0201-4

63 Putri, G. H., Anders, S., Pyl, P. T., Pimanda, J. E. & Zanini, F. Analysing high-throughput sequencing data in Python with HTSeq 2.0. *Bioinformatics* (2022). https://doi.org:10.1093/bioinformatics/btac166

64 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). https://doi.org:10.1186/s13059-014-0550-8

65 Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847-2849 (2016). https://doi.org:10.1093/bioinformatics/btw313

66 Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25 (2009). https://doi.org:10.1186/gb-2009-10-3-r25

67 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589 (2010). https://doi.org:10.1016/j.molcel.2010.05.004

68 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008). https://doi.org:10.1186/gb-2008-9-9-r137

69 Liu, T. Use model-based Analysis of ChIP-Seq (MACS) to analyze short reads generated by sequencing protein-DNA interactions in embryonic stem cells. *Methods Mol Biol* **1150**, 81-95 (2014). https://doi.org:10.1007/978-1-4939-0512-6_4

70 Castro-Mondragon, J. A. *et al.* JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **50**, D165-D173 (2022). https://doi.org:10.1093/nar/gkab1113

71 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005). https://doi.org:10.1073/pnas.0506580102

72 Hanzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013). https://doi.org:10.1186/1471-2105-14-7

73 Gel, B. & Serra, E. karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics* **33**, 3088-3090 (2017). https://doi.org:10.1093/bioinformatics/btx346

74 Reich, M. *et al.* GenePattern 2.0. *Nat Genet* **38**, 500-501 (2006). https://doi.org:10.1038/ng0506-500

75 Hoshida, Y. Nearest template prediction: a single-sample-based flexible class prediction with confidence assessment. *PLoS One* **5**, e15543 (2010). https://doi.org:10.1371/journal.pone.0015543

Acknowledgements

We acknowledge the Centre de Ressources Biologiques (Biological Resource Center), Strasbourg, France for the management of patient-derived liver tissues. We also thank the GenomEast platform (IGBMC, Illkirch, France), a member of the 'France Genomique' consortium (ANR-10-INBS-0009) for library preparation and sequencing of ChIP samples. The authors thank Prof. R. Bartenschlager (University of Heidelberg, Germany) and Prof. C. Rice (Rockefeller University, New York, USA) for providing plasmids for HCVcc production and, F. Chisari (The Scripps Research Institute, La Jolla, California, USA) for the gift of Huh7.5.1 cells. Plasmids expressing different HCV proteins were a gift from Dr. T. Wakita (National Institute of Infectious Disease, Tokyo, Japan). Bmal1-Luciferase plasmid was provided by Dr. D. Ray (Oxford, UK) and SIRT1-Luc was a gift from Dr. J. Auwerx (Laussane, Switzerland). We thank D. Heide (DKFZ, Heidelberg) for assistance with MYC immunostaining. We thank Dr. C. Eller and Dr. K. Herzog for their assistance in the initial phase of the study.

Funding

This work was funded by the European Union (ERC-AdG-2014-HEPCIR #671231 and ERC-AdG-2020-FIBCAN # 101021417 to T.F.B. and Y.H, EU H2020-HEPCAR #667273 to T.F.B, J.A.M. and J.L, HORIZON-HLTH-2021-DISEASE-04-07 D-SOLVE #101057917 to T.F.B and J.L., Fondation ARC- www.fondation-arc.org, Paris (TheraHCC2.0 IHUC201901299 to T.F.B.), the University of Strasbourg Institute for Advanced Study (USIAS-2020-029 to A.M.), the ANRS Maladies infectieuses émergentes (ANRS-MIE; ECTZ103701 to T.F.B., ECTZ160436, ECTZ131760, ECTZ171594 to J.L.), the French National Research Agency RHU DELIVER (ANR-21-RHUS-0001) and LABEX ANR-10-LABX-0028_HEPSYS (T.F.B.), the US National Institute of Health (U19-AI123862) to T.F.B and S.K.H.F, and the US National Institute of Health (R01CA233794) to Y.H and T.F.B., the UK Medical Research Council (MRC; MR/R022011/1 to J.A.M.), the Wellcome Trust (IA 200838/Z/16/Z to J.A.M.) and the Japan Agency for Medical Research and Development (AMED; 19fk0210020h0003 to K.C.), and the Inserm Plan Cancer 2019–2023 to T.F.B. Y.H. acknowledges the funding from US National Institute of Health (CA233794, CA255621), and Cancer Prevention and Research Institute of Texas, USA (RR180016, RP200554). The work in the lab of M-L. Y was partly supported by the "Center of Excellence for Metabolic Associated Fatty Liver Disease, National Sun Yat-sen University, Kaohsiung" from the featured areas research center program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE), Taiwan. This work of the Interdisciplinary Thematic Institute IMCBio, as part of the ITI 2021–2028 program of the University of Strasbourg, CNRS, and Inserm, was supported by IdEx Unistra (ANR-10-IDEX-0002) and by the SFRI-STRATUS project (ANR 20-SFRI-0012) and EUR IMCBio (ANR-17-EURE-0023) under the framework of the French Investments for the Future Program.

Author Contributions

T.F.B. initiated and coordinated the study. J.L. and T.F.B. jointly supervised the study. A.M., J.L., T.F.B. designed experiments. F.J. performed the computational analyses and representation of NGS data and figures. A.M., Y.S., E.C, F.D.Z, Y.T., A.H., P.B., C.G., N.S., M.A.O., S.C.D, C.P., X.Z., H.A.C. performed experiments. A.S., M.C., Y.H.J.H., R.T.C, M.L.Y, K.C. provided samples from patient cohorts. K.C. provided liver tissue samples from HLCMs. A.S., M.C., Y.H. generated the HCVclock gene signature. S.P., F.R., N.F., Y.H. performed the bioinformatic analyses of cancer gene and HCVclock signatures. A.M., F.J., J.L., T.F.B. wrote the manuscript. M.H., I.D., S.K.H.F., C.S., J.A.M. edited and revised the manuscript. All the authors read and approved the manuscript contents and the author list and its order.

Competing interests

We declare that none of the authors have competing financial or non-financial interests.

Materials and Methods

Generation of uPA/SCID humanized liver chimeric mice (HLCM) and HCV infection. Humanized liver chimeric mice were produced by splenic injection of cryopreserved PHH (BD Biosciences, San Jose, CA) into uPA/SCID mice as previously described[60]. The PHH donor (HF284) was a deceased 2-year-old female patient with no recorded history of liver disease and viral infection. The isolated PHH showed positive enzymatic activities for multiple human Cytochromes (CYP). PHH were stored and thawed as instructed to maintain their high viability (~75%). Both control and HCV-infected mice were engrafted with the same batch of PHH into

the spleen. All animal protocols described in this study were performed in accordance with the Guide for the Care and Use of Laboratory. Animals (https://grants.nih.gov/grants/olaw/guide-for-the-care-and-use-of-laboratory-animals.pdf) and the experimental protocol was approved by the Ethics Review Committee for Animal Experimentation of the Graduate School of Biomedical Sciences, Hiroshima University with approval number A14-195. All mice (male) were kept in isolated cages and fed ad libitum on CFR1 diet (Oriental Yeast Co., Ltd, Tokyo, Japan). Twelve weeks after hepatocyte transplantation, mice were injected intravenously with or without 10^5 copies of genotype 1b HCV-infected serum[61]. After serum inoculation, mouse blood samples were obtained serially, and human albumin concentrations and serum HCV RNA levels were measured. Human serum samples were obtained from a patient who had provided written informed consent for participation in the study. The serum sample was divided into small aliquots and stored -80 °C until use. Both control and HCV-infected mice were provided ad libitum chow diet and water. Both groups of mice were maintained in 12 h light-dark cycle for 10 weeks and were sacrificed through carbon dioxide every four hours.

Measurement of human albumin levels. Human hepatocyte repopulation rates were estimated by blood human albumin levels. Blood samples (5 μL) were collected periodically from the tail vein, and the levels of human albumin levels were determined with the Human Albumin ELISA Quantitation kit (Bethyl Laboratories Inc., Montgomery, TX, USA).

Measurement of HCV load from serum samples. RNA was extracted from mouse serum samples using Sepa Gene RV-R (EIDIA Co., LTD., Tokyo, Japan). Extracted RNA was reverse transcribed using random primer (TakaRa Bio Inc., Shiga, Japan) and M-MLV reverse transcriptase (ReverTra Ace; TOYOBO Co., LTD., Osaka, Japan) according to the instructions provided by the manufacturer. HCV RNA levels in mice were measured using the COBAS

TaqMan HCV test (Roche Diagnostics, Tokyo, Japan). The lower detection limit for the assay in mice was 3.45 log IU/mL.

RNA-seq NGS library preparation. The amount of total RNA was quantified using the Qubit 2.0 Fluorometric Quantitation system (Thermo Fisher Scientific, Waltham, MA, USA) and the RNA integrity number (RIN) was determined using the Experion Automated Electrophoresis System (Bio-Rad, Hercules, CA, USA). RNA-seq libraries were prepared with the TruSeq Stranded mRNA LT sample preparation kit (Illumina, San Diego, CA, USA) using Sciclone and Zephyr liquid handling workstations (PerkinElmer, Waltham, MA, USA) for pre- and post-PCR steps, respectively. Library concentrations were quantified with the Qubit 2.0 Fluorometric Quantitation system (Life Technologies, Carlsbad, CA, USA) and the size distribution was assessed using the Experion Automated Electrophoresis System (Bio-Rad, Hercules, CA, USA). For sequencing, samples were diluted and pooled into NGS libraries in equimolar amounts. Expression profiling libraries were sequenced on HiSeq 3000/4000 instruments (Illumina, San Diego, CA, USA) following a 50 bp, single-end recipe. Raw data acquisition (HiSeq Control Software, HCS, HD 3.4.0.38) and base calling (Real-Time Analysis Software, RTA, 2.7.7) was performed on-instrument, while the subsequent raw data processing off the instruments involved two custom programs based on Picard tools (2.19.2). In a first step, base calls were converted into lane-specific, multiplexed, unaligned BAM files suitable for long-term archival (IlluminaBasecallsToMultiplexSam, 2.19.2-CeMM). In a second step, archive BAM files were demultiplexed into sample-specific, unaligned BAM files (IlluminaSamDemux, 2.19.2-CeMM).

H3K27ac and H3K9ac chromatin immunoprecipitation (ChIP). Chromatin immunoprecipitation (ChIP) of H3K27ac and H3K9ac were performed as described previously[35], with minor modifications. In brief, for ChIP from humanized mice we used ~100 mg liver per

time point. Liver tissues were dounce homogenized in cold PBS, on ice. Subsequently, fresh formaldehyde was added (1% final concentration (v/v) to the PBS containing cells and incubated on a flip-flop rocker for 10 min at room temperature, followed by the addition of 2 M glycine (0.125 M final concentration) and incubation for 5 min (room temperature). Cross-linked cells were pelleted (400 x g, 5 min) at 4 °C, washed twice in ice-cold PBS and resuspended in ice cold lysis buffer [10 mM EDTA, 50 mM Tris-HCl (pH 8), 1% SDS, 0.5 x PMSF and protease inhibitor cocktail Complete, EDTA-free (Roche)]. Cross-linked lysates were sonicated [Covaris; (Power 200 Watts, Duty factor 28%, Cycles/burst 200 and time/sample 20 min)] at 4 °C to generate 200-500 bp chromatin fragments. Cellular debris were removed by centrifugation (10000 x g, 10 min) at 4 °C and the supernatant (chromatin extract) was quantified and verified for sonication efficiency. Next, the chromatin extracts were pre-cleared with 30 µL of protein G Dynabeads (per sample) for 1 h at 4 °C. Next, beads were discarded, and equal amounts of lysates were subjected to immunoprecipitation. 10% of the lysate was saved and served as input DNA. Next, individual samples were diluted 1:8 (v/v) in ChIP dilution buffer (16.7 mM Tris, 0.01% SDS, 1% Triton X-100, 1 mM EDTA, 16 mM NaCl, 1X Complete, EDTA-free (pH 8.1) and incubated overnight at 4 °C on a rotating wheel with control IgG (Diagenode, #C15410206) and either H3K27ac (Active Motif, #39133) or H3K9ac antibody (Active Motif, #39137). Next, 50 µL of pre-cleaned Dynabeads G were added for 60 min at 4 °C on a rotating wheel. Protein G-bound immunocomplexes were then washed twice with low-salt buffer (20 mM Tris, 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 150 mM NaCl, pH 8.1), twice in high-salt buffer (20 mM Tris, 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 500 mM NaCl, pH 8.1), twice in LiCl buffer (10 mM Tris, 250 mM LiCl, 1% NP-40, 1% sodium deoxycholate, 1 mM EDTA, pH 8.1), and finally washed twice in 1 mL TE buffer (10 mM Tris, 1 mM EDTA, pH 8). Chromatin was released from the

beads by incubation with 250 µL of elution buffer (1% SDS, 100 mM $NaHCO_3$, pH 8) for 10 min at room temperature on a shaker at 1250 rpm in a tabletop centrifuge. The elution step was repeated and both eluates were pooled. 1 µL of RNaseA (10 mg/mL; Fermentas) and NaCl (200 mM final concentration) was added to the eluate and incubated at 65 °C overnight. The reaction was stopped by 3 µL proteinase K (10 mg/mL; Fermentas) and incubated at 50 °C for 1 h. DNA was subsequently purified from the eluate using QIAGEN PCR purification kit in a final volume of 50 µL. This DNA was used for library preparation.

ChIP-seq NGS library preparation. ChIP samples were purified using Agencourt AMPure XP beads (Beckman Coulter) and quantified with the Qubit (Invitrogen). ChIP-seq libraries were prepared from 3 to 10 ng of double-stranded purified DNA using the MicroPlex Library Preparation kit v2 (C05010014, Diagenode s.a., Seraing, Belgium), according to manufacturer's instructions. In the first step, the DNA was repaired and yielded molecules with blunt ends. In the next step, stem-loop adaptors with blocked 5 prime ends were ligated to the 5-prime end of the genomic DNA, leaving a nick at the 3-prime end. The adaptors cannot ligate to each other and do not have single-strand tails, avoiding non-specific background. In the final step, the 3 prime ends of the genomic DNA were extended to complete library synthesis and Illumina compatible indexes were added through a PCR amplification (7 + 4 cycles). Amplified libraries were purified and size-selected using Agencourt AMPure XP beads (Beckman Coulter) to remove unincorporated primers and other reagents. Libraries were sequenced on an Illumina HiSeq 4000 sequencer as single read 50 base reads. Image analysis and base calling were performed using RTA version 2.7.7 and bcl2fastq version 2.20.0.422.

Patient cohort: The cohort c included chronically infected patients (F1-F4) and control subjects without liver disease[14]. The patient characteristics have been described[35]. Most of the liver

specimens in these cohorts were obtained in the morning, however, we cannot exclude a moderate temporal heterogeneity due to variation in clinical schedule. The humanized samples in cohort b were all sacrificed during the circadian rest phase.

Transcriptome analyses. NGS reads from humanized mouse livers and human patient samples were mapped to an artificial genome containing the human and mouse Genome Reference Consortium GRCh38 and GRCm38 assemblies (humanized mouse) or the common GRCh38 assembly using HISAT2[62]. The study protocol for human patient samples was approved by the Hiroshima University ethical committee (approval number HI-98-21) in accordance with the Helsinki declaration. All patients provided written informed consent. Mouse and human reads were counted with htseq-count[63], and normalization as well as differentially expression analysis performed with Bioconductor DESeq2 (1.28.1)[64] package based on a model using the negative binomial distribution. Clustering for selecting closest samples at each ZT was performed using ward.D2 algorithms as implemented in the Bioconductor ComplexHeatmap package[65].

ChIP-seq analyses. Raw reads were demultiplexed using bcl2fastq v2.20.0.422 mapped to the human genome hg38 using the Bowtie aligner v1.2.2[66]. Quality control checks were performed using FastQC. Bigwig files were created using make UCSCfile from HOMER[67] with the parameters- norm 14028944-fraglength 200. All original alignment files were scaled down to the same number of reads set as the lowest number of the mapped reads (14M for H3K27ac) or to the maximum of 20M (in the case of H3K9ac) using the tool MACS2 randsample[68]. with the following parameters-n14028944-f BAM. All original Peaks were then called in uniquely mapped reads using MACS3[69] v3.0.0b3 with parameters -f BED-g hs. Target genes were defined using the predicted transcription factor binding sites listed in the JASPAR[70] database with a score higher than 500.

The *dryR* analyses of cycling genes and pathways. Differences in cycling gene pattern and assignments to expression modules comparing mouse vs. human hepatocytes, as well as between control (uninfected human hepatocytes) and HCV-infected human hepatocytes were analyzed using *dryR*[26]. In the case of comparing rhythmicity between human hepatocytes and mouse hepatocytes, only genes with orthologues in both species and with a mean of more than 10 read counts were used as input for the *dryR* analyses, as described[27]. Intersections with external data were performed based on human gene names and hypergeometric tests were calculated for overlaps. Gene set enrichments for MsigDB v7.2[71] sets comparing conditions were calculated using local javaGSEA, and sample-wise enrichments were calculated using GSVA[72]. Genome tracks were generated using the Bioconductor karyoploteR[73] package based on down-samples BAM alignment files, heatmaps were generated using the Bioconductor ComplexHeatmap[65] package, and other customized plots were generated using ggplot2.

Assessment of HCV$_{CLOCK}$ signature. To assess the presence of the 4 CC-oscillator transcriptomic model (i.e., loss, gain, altered, and unaltered) in liver disease patient cohort for association with clinical characteristics and prognosis, CC-oscillator gene sets were first defined as differentially expressed genes (at 4-fold or more) between early (ZT0-ZT8) or late (ZT12-ZT20) time points in each of HCV-infected and control conditions in each model. Subsequently, enrichment of the gene sets were assessed in our previously generated genome-wide transcriptome profiles of 216 early-stage (Child-Pugh class A) HCV-related cirrhosis patients[18] (GSE 15654) by using the eseach algorithm utilizing Kolmogorov-Smirnov statistic-based enrichment score[21,51] implemented in the eseach module of GenePattern genomic analysis toolkit[74] (github.com/genepattern/gparc-module-docs). Gene set enrichment index (GSEI) was calculated as -log10 (nominal p-value for the enrichment score) with sign of enrichment score for

each signature and each patient as previously reported[21,51]. GSEI was also calculated for our previously reported Prognostic Liver Signature (PLS) that predicts poor prognosis of liver cirrhosis patients[18,21,40,48-51]. Correlation of GSEIs with clinical characteristics and PLS was evaluated by Spearman rank correlation test. Using the CC-oscillator signatures correlated with clinical features reflecting liver disease severity, the 216 cirrhosis patients were classified into high- or low-risk groups using the Nearest Template Prediction (NTP) algorithm[75] implemented in the Gene Pattern Nearest Template Prediction module. Association of the classification with time to overall death was assessed by Kaplan-Meier method and Cox regression. All analyses were performed using R statistical language (www.R-project.org).

Histology and immunohistochemistry. Samples for histology were placed in 10% neutral formalin overnight before transfer to 70% ethanol and later embedding in paraffin and cross-sectioned to obtain 5 µm section, and then extra-coated with paraffin to preserve tissue integrity. Glass slide-mounted tissues were scanned with Nanozoomer scanner (Hamamatsu), and images were analyzed using image processing software (ImageJ). Additional sections were stained with antibody specific for CK8-18, and cMYC. Deparaffinization was performed in the BondMax automate (Leica biosystem) in Bond Dewax Solution. Antigen unmasking was performed in the BondMax in EDTA buffer plus detergent pH 8.9-9.1 at 95 °C. Sections were preincubated with H_2O_2 for 5 minutes and then incubated with primary antibodies (CK18, M701029-2, Clone DC 10, Dako; 1:150; cMYC, #ab32072 Active Motif 1:300) according to manufacturer's instructions. After subsequent washes with Bond Wash Solution (Leica), sections were incubated with the secondary antibody for 20 min. After washing, sections were incubated in Mixed-DAB-Solution for 6 min. Images were taken with a Leica Kit- DS9800 microscope (Leica Biosystem) using Plan 20x objectives and analyzed using Image J. For H&E and Sirius Red staining sections were

deparaffinized, rehydrated and processed. For Sirius Red staining: slides were incubated with a 0.1% Sirius Red solution dissolved in aqueous saturated picric acid for 1 h (Sirius Red: Sigma-Aldrich, Direct Red 80, #365548), washed in acidified water (0.5% HCl), dehydrated, and mounted under coverslips with Eukitt mounting medium (Sigma). For HE staining the deparaffinized and rehydrated sections were washed up and stained with Mayer's Hematoxylin solution. Briefly, whole slide images were generated using 3D Panoramic SCAN and uploaded into Ndpi view software. For collagen positive area (CPA) morphometric quantification was performed in the Sirius Red -stained slides and quantified as a percent of the total image analysis area using ImageJ (NIH). Quantification of CPA and HE slides to determine fibrosis and steatosis, respectively were done blindly.

CRISPR/Cas9 gene editing of Huh7.5.1 cells. Lentiviruses expressing single guide RNA (sgRNA) were produced in HEK 293T cells by co-transfection with an envelope plasmid (pMD2.G), a packaging plasmid (psPAX2) and a lentiviral vector expressing the sgRNA (pXPR_BRD016) from the Broad Institute. Co-transfection was performed using the CalPhos Mammalian Transfection Kit (Clontech Laboratories) according to manufacturer's instructions. LX2 stably expressing Cas-9 endonuclease was generated by transduction of a lentiviral vector expressing Cas9 (pXPR_BRD111, Broad Institute). For *control (CTRL)* and *SIRT1* KO, Huh7.5.1-Cas9 cells were then transduced with lentiviruses expressing single guide RNA (sgRNA) CTRL targeting GFP (sgCTRL: 5'-GGT GAA CCG CAT CGA GCT GA) or specific sgRNAs (sgSIRT1-A: 5'-GTT GAC TGT GAA GCT GTA CG; sgSIRT1-B: 5'-CCA GAG AGG CAG TTG GAA GA; and sgSIRT1-C: 5'-ATC CTC CTC ATC ACT TTC AC) designed by the Synthego (California, USA). After 48 h, transduced cells were selected under puromycin treatment (4 µg/mL). KO was determined at genetic level using T7 endonuclease assay (Alt-R

Genome Editing Detection Kit, IDT), according to manufacturer's instructions, and by performing western blot analyses with SIRT1 specific antibody.

Antibodies. ChIP-seq was conducted using antibodies targeting H3K27ac (Active Motif; #39133), H3K9ac (Active Motif, #39137) and IgG (Diagenode, #C15410206). Following antibodies were used for IHC: CK18 (CK18, #M701029-2, Clone DC 10, Dako), and MYC (Abcam, #ab32072). Western blot and immunostaining analyses were performed using: SIRT1 (Abcam, #ab110304), CK18 (Dako, M7010, clone DC10), BMAL1 (Abcam, #ab93806), HCV-Core (Invitrogen, #MA1-080), HCV-NS3 (GenTex, #GTX131276), HCV-NS4 (Invitrogen, #PA1-73108), HCV-NS5A, β-actin (Chemicon, #MAB1501), and β-Tubulin (GenTex, #101279).

Luciferase reporter assay. The *BMAL1*-luciferase (BMAL1-Luc) reporter construct was obtained from the laboratory of Dr. D. Ray (Oxford, UK), and the *SIRT1*-luciferase (SIRT1-Luc) construct was developed in the laboratory of Dr. J. Auwerx (Laussane, Switzerland). Briefly, Huh7.5.1 cells were transiently transfected with 300 ng of either *BMAL1- Luc* or SIRT1-Luc reporter constructs using ViaFect transfection reagent (Promega, #E4982) in Opti-MEM (Gibco, #31985-062) according to the manufacturer's instructions. 500 ng of pEFP plasmid was also added to transfection mixture as transfection control. Luciferase activities were determined by the Bright-Glo or ONE-Glo Reporter Assay System (Promega) and a Mithras LB940 plate reader (Berthold Technologies). The relative luciferase activity values of treated cells were normalized to that of control cells.

Cell line, virus, and viral plasmids: Human hepatoma Huh7.5.1 cells were cultured in Dulbecco's modified Eagle's medium (Gibco, #61965-026) containing 10% fetal bovine serum (Cytiva, #SH30066.03), 0.5% gentamycin (Gibco, #15710-049), and 1% non-essential amino

acids (Gibco, #11140-035). Cells were differentiated in 1% DMSO-containing medium. HCV Jc1 (genotype 2a/2a) were used in viral infection study. Plasmids expressing HCV core-E1-E2 (pEF-JFH1-CE2), core (pEF-JFH1-Core191), NS3 (pEF-JFH1-NS3), NS4 (pEF-JFH1-NS4), NS5B (pEF-JFH1-NS5B), and NS5A (pEF-JFH1-NS5A) were a gift from Dr. Takaji Wakita (Dept. of Virology, National Institute of Infectious Diseases, Tokyo, Japan).

Reagents and chemicals: DMSO were purchased from Sigma. SIRT1 agonist (SRT2183) was from Med Chem Tronica, #HY-19759.

Statistical analysis: For cell culture experiments results are presented as both mean and individual values. The Data were compared by Welch's t-test or Kruskal-Wallis-ANOVA, Bonferroni's test by using GraphPad Prism v 9.5 (GraphPad software). The statistical tests were two-tailed, and significance was set at $P < 0.05$ (*) or < 0.01 (**). For immunoblotting, the band density was quantified by using ImageJ (US National Institutes of Health). Gene expression and pathway enrichments were compared on bulk level for each ZT separately. Gene expressions were compared using DESeq2 and then using Wald test ($p<0.05$). Gene set enrichment scores were drawn online plots and compared applying the Wilcoxon signed-rank test ($p<0.05$) as implemented in R on enrichment scores.

Data availability: Raw files and technical details about the RNA-seq and ChIP-seq data have been deposited in the National Center for Biotechnology Information's Gene Expression Omnibus (GEO) and are accessible through GEO series accession number GSE200812. Previously published data (GSE84346, SRP170244, PRJNA606244, GSE15654) were also used for this work.

Inclusion and ethics: All the authors approved the contents of the manuscript as well as their contributions and order in the author list.

Figure Legends

Fig. 1: The rhythmic transcriptome of human liver cells *in vivo*.

a) Schematic representation for the generation of immunodeficient humanized liver chimeric mice (HLCM). Liver tissues obtained from these HLCM were used to perform circadian RNA-seq and ChIP-seq. PHH: primary human hepatocyte.

b) Circadian expression pattern of core CC-oscillator genes and CC-output regulators in human and mouse liver cells of HLCM in two independent experiments (Series 1 and 2).

c) Expression pattern of genes showing circadian variations in human and mouse hepatocytes in HLCM as predicted by *dryR* (n=5 HLCM/ circadian timepoint). *dryR* identified four models of rhythmic genes in HLCM. A fifth model comprising only non-cycling genes is not shown. Phase distribution of respective models is indicated by radial coordinates, where green-human, pink-mice, and grey-overlap of human and mice.

d) HALLMARK pathways significantly (FDR<0.05) enriched for *dryR* genes displaying circadian variations in human hepatocytes in HLCM (n=5 HLCM/ circadian timepoint) as listed in (c) and their expression in WT mice. Similar pathways show overlapping (±1 ZT step) of peak enrichment scores comparing human and WT mice maximum enrichment scores.

e) Circadian expression pattern of transcription factors (TFs) in human hepatocytes in HLCM (n=5 HLCM/ circadian timepoint), as predicted by *dryR*.

f) Examples for circadian expression pattern of *dryR* identified TFs as listed in (e). The Y-axis represents the DESeq2 normalized reads. Human hepatocytes (red) and in residual murine (green) cells. n=5 HLCM/ circadian timepoint; bars represent SD.

g) ChIP-seq coverage plots indicate circadian variations in H3K27ac levels in *IRF2* in human hepatocytes and WT mice liver.

Fig. 2: HCV disrupts the human liver diurnal transcriptome.

a) Schematic representation for the chronic HCV infection of HLCM. Liver tissues obtained from these infected HLCM were used to perform circadian RNA-seq and ChIP-seq. Two independent infection experiments were performed. PHH: primary human hepatocytes.

b) Heatmap of circadian expression pattern of CC components in human cells in the liver of control (uninfected) and HCV-infected HLCM (n=5 HLCM / group / circadian timepoint).

c) The *dryR*-identifies four categories of CC-disturbed genes according to their expression in control and HCV-infected human liver cells in HLCM (n=5 HLCM / group / circadian timepoint). A fifth model comprising only non-cycling genes is not shown. Phase distribution of indicated models is indicated by radial coordinates, where green: control, pink: HCV, and grey: overlap of control and HCV.

d) Examples for significantly enriched (FDR<0.05) circadian pathways (i.e., HM: HALLMARK, GO: Gene Ontology, BC: BIOCARTA, RE: Reactome gene sets) identified in control human liver cells (Fig. 1d), shown in comparison to HCV-infected livers. n=5 HLCM / group / circadian timepoint. Bold dots inside figure panels represent (p<0.05; Wilcoxon signed-rank test).

Fig. 3. HCV impairs the diurnal remodeling of the circadian enhancers.

a) H3K27ac peaks at different circadian time-points across the gene bodies of cycling genes in control and HCV-infected human liver cells in HLCM. TSS: transcription start site.

b) TSS-associated H3K27ac peaks of cycling genes at different circadian time-points in control and HCV-infected human liver cells in HLCM. TSS: transcription start site.

c) Quantitation of H3K27ac peak numbers at TSS from (b), Mann Whitney test.

d) H3K9ac peaks at different circadian time-points across the gene bodies of cycling genes in control and HCV-infected human liver cells in HLCM. TSS: transcription start site.

e) TSS-associated H3K9ac peaks of cycling genes at different circadian time-points in control and HCV-infected human liver cells in HLCM. TSS: transcription start site.

f) Quantitation of H3K9ac peak numbers at TSS from (e), Mann Whitney test.

g) Promoter-enriched H3K27-acetylation for HCV host factors *XBP1* and *RAF1* in control and HCV-infected human liver cells in HLCM.

h) Percentages of enhancer/TSS-enriched peaks with gain or loss of H3K27-acetylation that overlap with transcription factor binding sites in gene targets as listed in the Jaspar database. Only the 50 top hits are shown for each direction.

i) Circadian expression pattern of indicated SIRTUINs (1, 2, and 3) in human and mouse liver cells in HLCM. The Y-axis represents the DESeq2 normalized reads. n=5 HLCM/ group / circadian timepoint. Bars represent SD. *dryR* identified *SIRT1* as uniquely rhythmic (model 2; Fig. 1c) in human hepatocytes of HLCM liver, while SIRTs 2 and 3 are non-rhythmic (model 1, Supplementary Table 3).

j) Circadian expression pattern of indicated Sirtuins in control and HCV-infected human liver cells in HLCM. The Y-axis represents the DESeq2 normalized reads. n=5 HLCM/ group / circadian timepoint. Bars represent SD. Bold dots inside figure panels represent ($p<0.05$; Wald test).

Fig. 4. HCV NS4 and NS5A perturbs SIRT1 expression in a cell culture model for HCV infection.

a) Western blot (WB) analyses of BMAL1 and its relative quantification in control (*sgCTRL*) and *SIRT1* KO (*sgSIRT1*; clone A) Huh7.5.1 cells.

b) *BMAL1*-luciferase *(BMAL1-Luc*) activity in control (*sgCTRL*) and *SIRT1* KO (*sgSIRT1*) Huh7.5.1 cells, 24 and 48 h post-transfection.

c) *BMAL1-Luc* activity in control (*sgCTRL*) and *SIRT1* KO (*sgSIRT1*) Huh7.5.1 cells, treated with SIRT1 agonist SRT2183, as indicated.

d) Western blot analyses of indicated proteins from Huh7.5.1 cells following either HCV infection or transient expression of individual HCV proteins.

e) *SIRT1*-luciferase (*SIRT1-Luc*) activity in Huh7.5.1 cells following transient expression of individual HCV proteins, as indicated.

f) *BMAL1*-luciferase (*BMAL1-Luc*) activity in Huh7.5.1 cells following transient expression of individual HCV proteins, as indicated.

Fig. 5. HCV perturbed rhythmic pathways are partially reversed following viral cure and stay perturbed in patients progressing to virus-induced cancer.

a) Human liver circadian pathways (as identified in Fig. 1) showing either similar or distinct enrichment (FDR<0.05) in liver of HCV-infected HLCM, and infected patients. Humanized mice: HCV-infected HLCM (Fig. 2b; control vs HCV), Patients: Control vs HCV-Infected (F1-4). Pool: analyses of all the samples in respective groups. Red=significant enrichment, blue=significant negative enrichment, white=no significant enrichment/unchanged.

b) Correlation between induction of the CC-oscillator signatures and clinical demographics and the previously reported Prognostic Liver Signature (PLS) in 216 early-stage (i.e., Child-Pugh class A) HCV cirrhosis patients[18]. The signatures were defined as genes with > 4-fold over-expression in either HCV-infected or control hepatocytes in early or late time points in each of the 4 representative models (i.e., loss, gain, altered, and unaltered). All pair-wise correlations between the variables (correlation matrix) as shown in rows and columns. The correlation matrix was clustered to depict groups of variables sharing similar patterns of correlation. The deep blue, light blue and brown color bars indicate the presence of the 3 major correlation clusters.

c) Classification of the 216 early-stage cirrhosis patients[18] by the CC-oscillator signatures associated with clinical featured related to liver disease severity (Supplementary table 7), Magnitude of gene signature/set modulation is shown in each patient (in each column) in the cohort. The induction/suppression of the circadian clock gene sets (shown in bottom half of the panel) was determined by Gene Set Enrichment Index (GSEI) as previously described[21]. The PLS and HCVclock risk predictions were determined by using Nearest Template Prediction (NTP) algorithm[75]. For the PLS, normalized expression levels or the PLS member genes were used for

the NTP analysis as previously described[18,40]. For the HCV_{CLOCK} risk prediction, the GSEIs of the circadian clock gene sets for the NTP analysis were used. AFP, alpha-fetoprotein; ALT, alanine aminotransferase; AST, aspartate aminotransferase; PLS, prognostic liver signature; SVR, sustained virologic response.

d) Association of the disease-severity-related HCV_{CLOCK} signature-based classification with overall survival in the 216 early-stage HCV cirrhosis patients[18].

e) The Hallmarks of Cancer and the rhythmic pathways in human liver which were altered by HCV-infection, predisposing towards HCC.

Extended Data Fig. 1: Representative images of hematoxylin-eosin (HE) and human hepatocyte-specific CK8-18 staining of control humanized liver chimeric mice (HLCM). Wide field HE images revealed conserved lobular architecture of engrafted human hepatocytes and absence of stress response. Residual murine hepatocytes (smaller and intensely eosinophilic compared to human counterparts) could be seen in the immediate peri-vascular regions. Scale bars: 500 μm.

Extended Data Fig. 2: Unsupervised clustering of control humanized liver mice (HLCM) samples and CC genes from two independent circadian experiments (Series 1 and 2), as indicated. 3 HLCM were used in each circadian time-point. Sample S76790_ZT16 (Series 1; indicated in red) was ignored due to low coverage and low humanization level.

Extended Data Fig. 3: (a) Circadian expression of core CC genes and CC output regulators in human (red lines) and residual murine (green lines) hepatocytes obtained from liver of control HLCM (n=5 HLCM / circadian timepoint). The curves are shown as normalized DESeq2 read counts. Bars represent SD.

(b) Heatmap showing expression of indicated CC genes in human and murine hepatocytes in control HLCM and WT mice livers, as indicated.

Extended Data Fig. 4: (a) *dryR* genes with shared and/or unique circadian gene expression pattern in the cells of human and mouse origin in HLCM livers (Fig. 1c), and genes with circadian expression pattern in wild type mice as published in Reference[6]. (b) Genes with shared and/or unique circadian gene expression pattern in the cells of human origin in HLCM livers (Fig. 1c), and genes with circadian expression in post-mortem human liver (Talamanca et al.). Hypergeometric tests reveal non-significant overlaps.

Extended Data Fig. 5 a, b: Pathways significantly (FDR<0.05) enriched for genes displaying diurnal variations in human hepatocytes in HLCM as listed in (Fig. 1C) and their expression in WT mice. Similar pathways (a) show overlapping (±1 ZT step) of peak enrichment scores comparing human and WT mice maximum enrichment scores. (b) Pathways significantly (FDR<0.05) enriched for genes displaying distinct (separated by more than 1 ZT) diurnal variations in human hepatocytes in HLCM as listed in (Fig. 1C) and their expression in WT mice. BC: BIOCARTA, and KE: KEGG. N=5 HLCM / circadian time point. ZT0: represented twice in each panel to maintain conformity.

Extended Data Fig. 6: Heatmap of *dryR*-identified transcription factors (TFs) with diurnal and unique circadian gene expression pattern in the cells of human and mouse origin in HLCM livers. n=5 HLCM / circadian timepoint. ZT0: represented twice in each panel to maintain conformity.

Extended Data Fig. 7: Examples of transcription factors with diurnal and unique circadian gene expression pattern in the cells of human (red) and mouse (green) origin in HLCM livers (Fig. 1e and

Extended Data Fig. 6). Curves are shown as normalized DESeq2 read counts (n=5 HLCM / circadian timepoint). Bars represent SD.

Extended Data Fig. 8: ChIP-seq coverage plots indicate circadian variations in H3K27ac levels in the CC-output regulator gene *DBP* in human hepatocytes from HLCM and WT mice liver.

Extended Data Fig. 9a, b: Unsupervised clustering of HCV-infected humanized liver mice (HLCM) samples and CC genes from two independent circadian experiments (Experiment series 1- shown in a and series 2- shown in b), as indicated. 3 HLCM were used in each circadian time-point (n=18 HLCM / experiment).

Extended Data Fig. 10: (a) Quantification of human-specific CK8-18 staining reveals degree of humanization in control and HCV-infected HLCM liver. Each data point represents individual HLCM. All the mice from each group were used for staining and quantification. ***: $p<0.001$

(b) Representative images of CK8-18 staining of engrafted human hepatocytes in control and HCV-infected HLCM liver. The black arrows indicate perivascular regions with murine hepatocytes. Scale bar: 250 μm.

(c) Serum levels (ng/mL; $x10^7$) of human albumin in control and HCV-infected HLCM, as indicated. All the mice from each group were used for this measurement (Supplementary Table 1). *: $p<0.05$, **: $p<0.01$, ***: $p<0.001$.

Extended Data Fig. 11: (a) Heatmap of diurnal expression pattern of indicated CC genes in control and HCV-infected HLCM livers in two independent experiments, as indicated.

(b) Circadian expression of indicated CC genes in control and HCV-infected HLCM livers (n=5 HLCM / group / time point). The data represents normalized DESeq2 reads. Bars represent SD. Bold dots ($p<0.05$; Wald test). For BMAL1 (ZT8; p=0.07).

Extended Data Fig. 12: Serum levels of HCV in infected HLCM in different circadian timepoints, as indicated. N=3 HLCM/timepoint.

Extended Data Fig. 13: Percentage of distinct categories of cycling genes (obtained through *dryR* analysis) in circadian pathways in HCV-infected HLCM livers. n=5 HLCM / group / circadian timepoint.

Extended Data Fig. 14: (a, b) Representative images of MYC immunohistochemistry (IHC) from control (a) and HCV-infected HLCM liver

(b) at different circadian time points. Note that MYC (oncogene) is barely expressed in control conditions. Scale bar: 300μm.

(c) Quantification of MYC-immuno-stained cells from control and HCV-infected HLCM liver (as shown in panels a, b). p=0.01557 (Mann-Whitney test).

Extended Data Fig. 15: (a) Representative images of HE stained liver sections of control and HCV-infected HLCM liver, and quantification of the steatosis (inflammation and ballooning). Each point represents an HLCM.

(b) Sirius red staining of the control and HCV-infected HLCM liver at different circadian time points indicate mild fibrosis upon viral infection.

(c) Quantification of the collagen positive area (CPA) from Sirius red-stained (b) in control (n=18) and HCV-infected (n=18) HLCM liver.

Extended Data Fig. 16: (a) Circadian variations in H3K27ac levels in control and HCV-infected HLCM livers for *CRY1*.

(b) Circadian variations in H3K9ac levels in control and HCV-infected HLCM livers, for indicated genes.

Extended Data Fig. 17: (a) Circadian expression pattern of indicated histone acetylases and deacetylases from control and HCV-infected HLCM liver. The Y-axis represents the DESeq2 normalized reads (n=5 HLCM / group / circadian time point). Bars represent SD. Wald test revealed non-significant differences in expression for all the genes.

(b) Circadian expression pattern of indicated Sirtuin genes (SIRTs *2*, *4*, *6,* and *7*) in control and HCV-infected human liver cells in HLCM. The Y-axis represents the DESeq2 counts (n=5 HLCM / group / circadian time point). Bars represent SD. Wald test revealed non-significant differences in expression for all the genes.

(c) Circadian expression of indicated Sirtuin genes (SIRTs *2, 4, 6*, and *7*) in human (red) and mouse (green) liver cells in control HLCM. The Y-axis represents the DESeq2 normalized reads (n=5 HLCM / group / circadian time point). Bars represent SD. *dryR* also identified SIRTs 2, 4, 6, and 7 as non-rhythmic (model 1; Table S3) in human hepatocytes of HLCM.

Extended Data Fig. 18: (a) Human liver circadian pathways (as identified in Fig. 1d) showing either similar or distinct enrichment (FDR<0.05) in liver of indicated HCV-infected and viral-cured groups.

HLCM= humanized mice (HCV-infected vs DAA-cured; Supplementary Table 7), and patients (Pat 1, 2, and 3): HCV-cured paired biopsy (Japanese cohort; Supplementary Table 6). Red= significant enrichment, blue=significant negative enrichment, white=no significant enrichment, unchanged.

(b) Human liver circadian pathways (as identified in Fig. 1d) showing either similar or distinct enrichment (FDR<0.05) in liver of indicated groups. HLCM= humanized mice (HCV-infected vs DAA-cured; Supplementary Table 7), and HCV-infected and cured patients with HCC (Taiwanese cohort, Supplementary Table 6). Red= significant enrichment, blue=significant negative enrichment, white=no significant enrichment, unchanged.

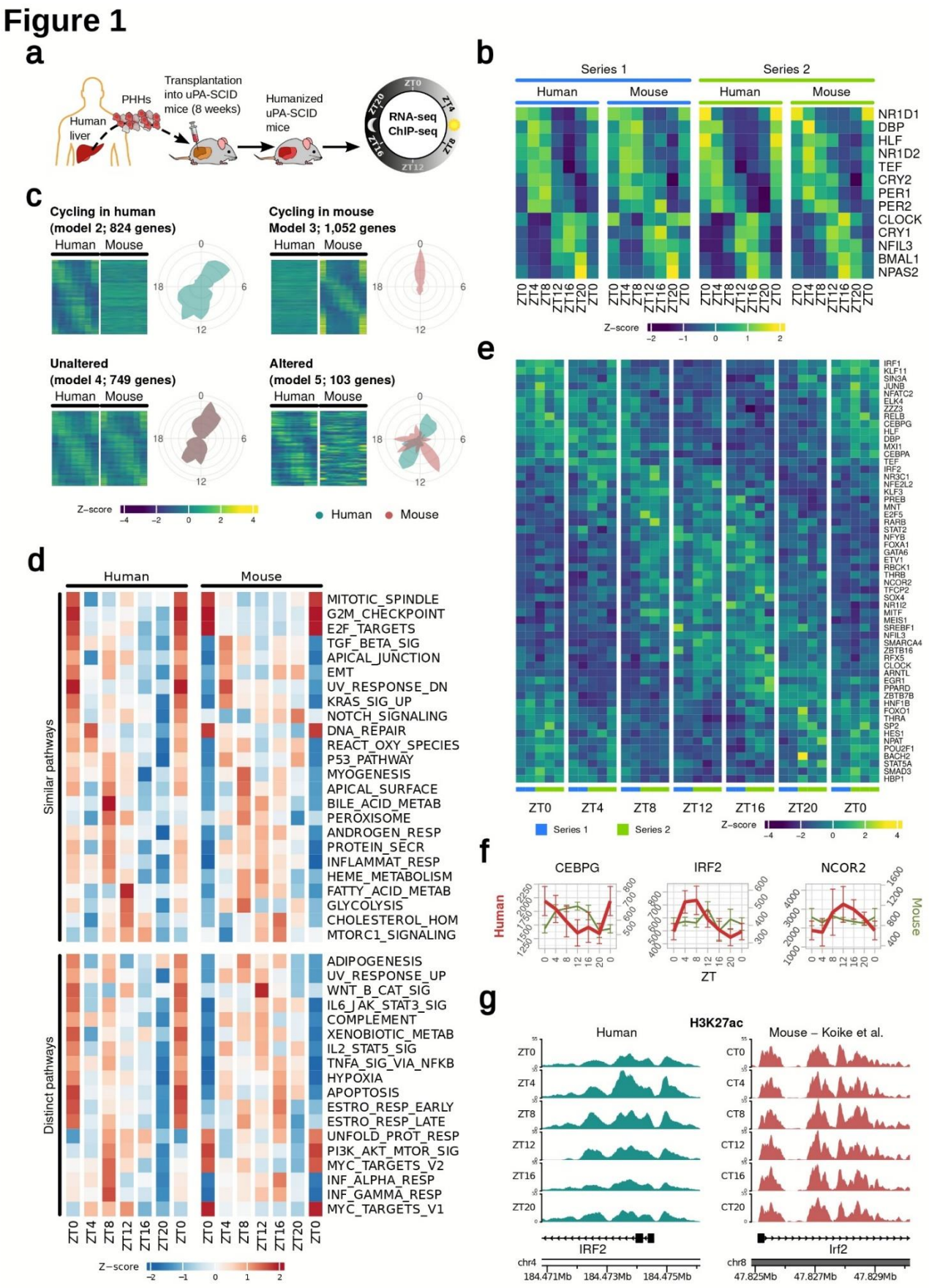
(c) Human liver circadian pathways (as identified in Fig. 1d) showing either similar or distinct enrichment (FDR<0.05) in liver of MASH patients. Red= significant enrichment, blue=significant negative enrichment, white=no significant enrichment, unchanged.

Extended Data Fig. 19: Proportions of circadian genes of the 186-gene prognostic liver signature (PLS) and its perturbations by HCV-infection in HLCM livers.

Extended Data Fig. 20: Full-length WBs for the gels shown in Fig. 4a. Also, shown are full length WBs for three independent (N: 1, 2, and 3) experiments.

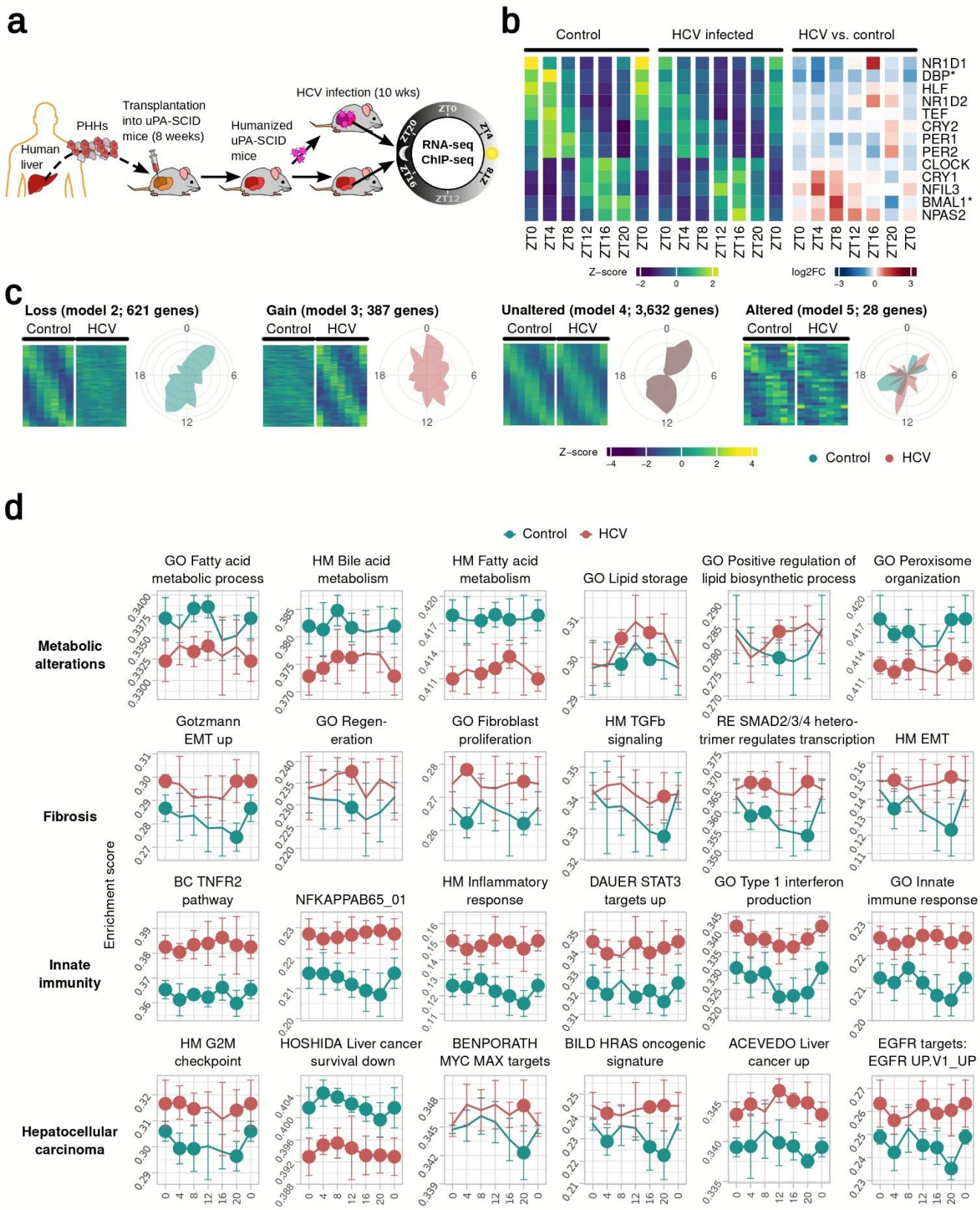
Extended Data Fig. 21: Full-length WBs for the gels shown in Fig. 4d. Also, shown are full length WBs for three independent (N: 1, 2, and 3) experiments.

Article 2: An atlas of the human liver circadian transcriptome and its carcinogenic perturbation by hepatitis C virus infection



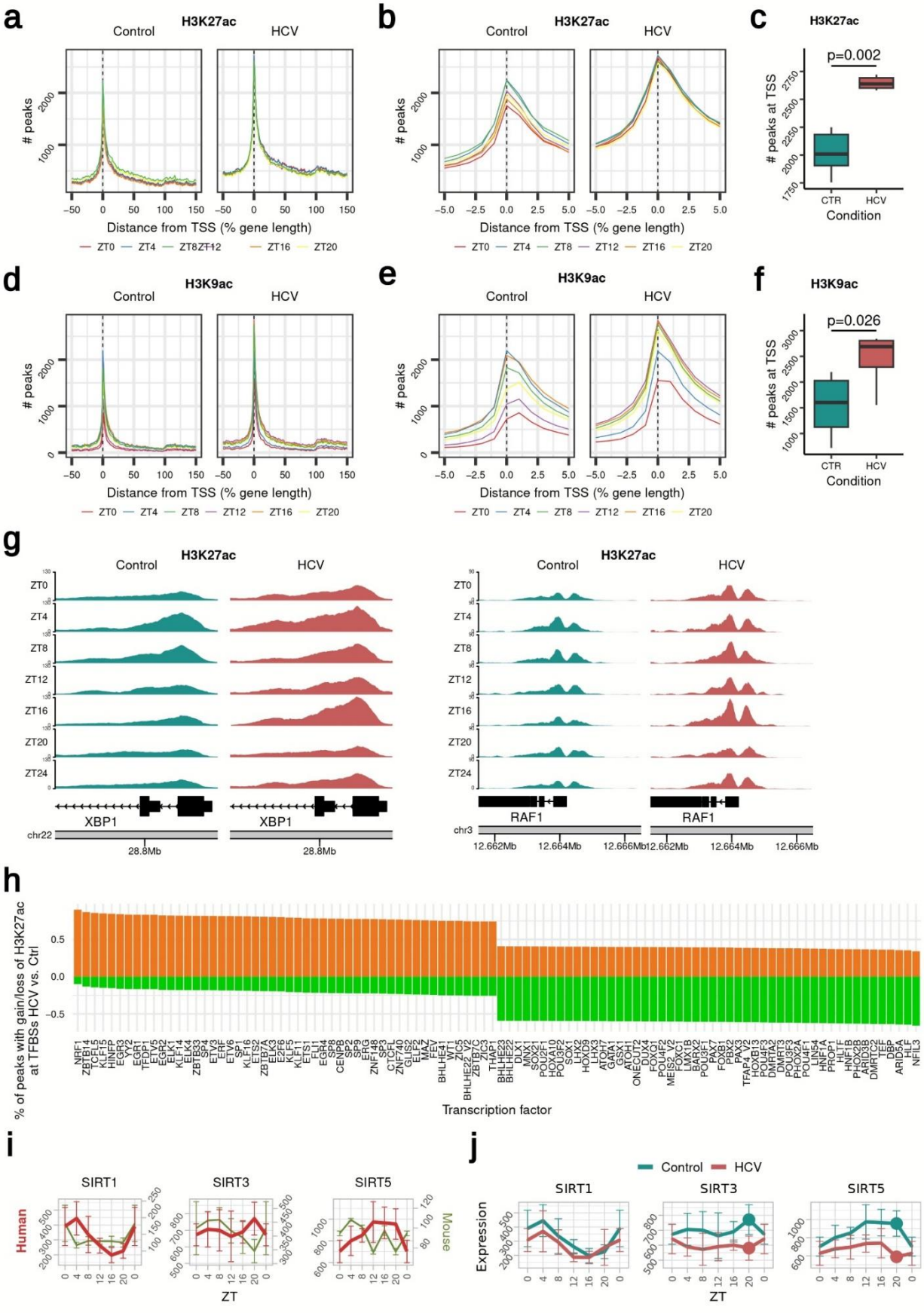
Article 2: An atlas of the human liver circadian transcriptome and its carcinogenic perturbation by hepatitis C virus infection

Figure 2



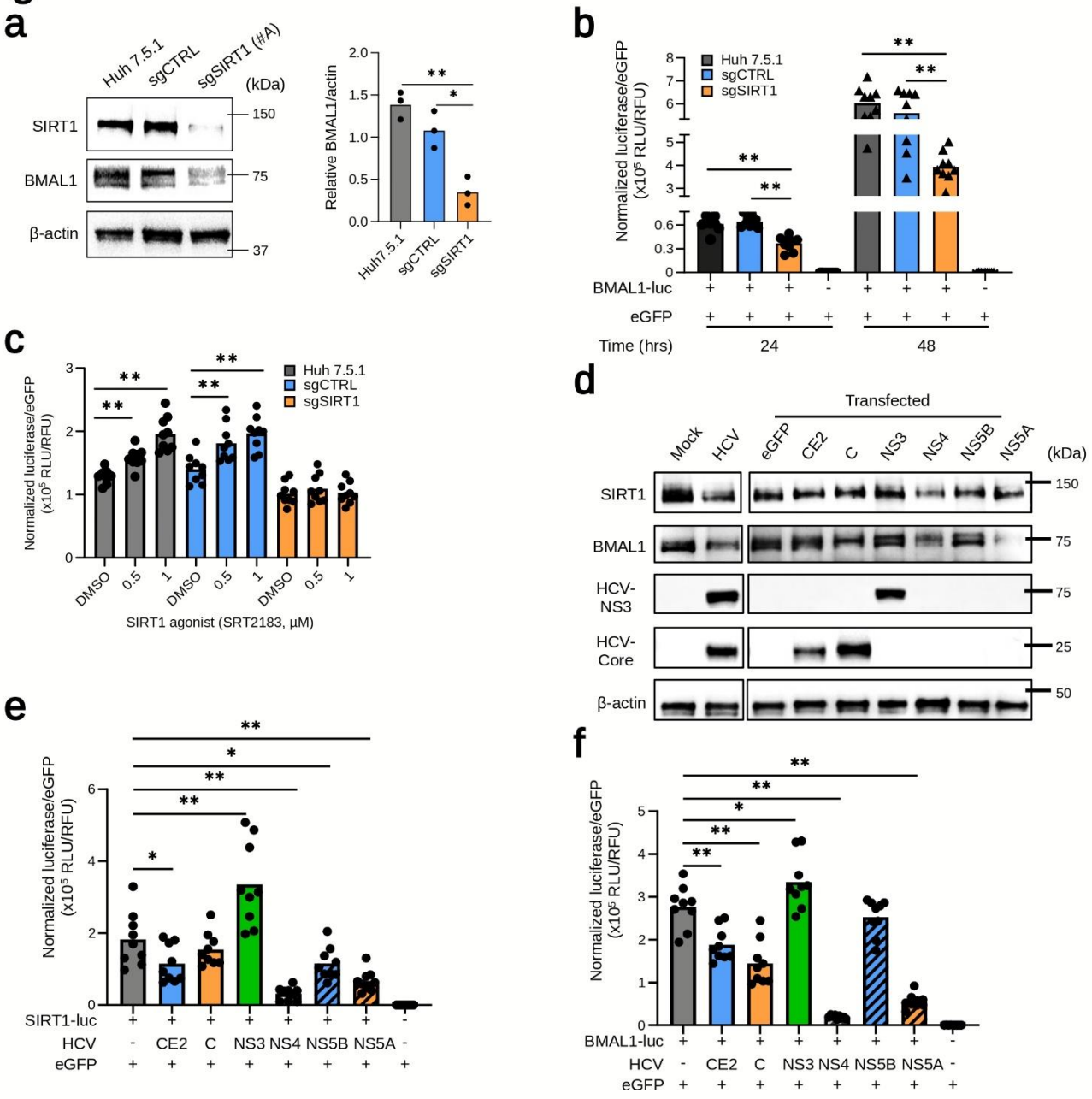
Article 2: An atlas of the human liver circadian transcriptome and its carcinogenic perturbation by hepatitis C virus infection

Figure 3



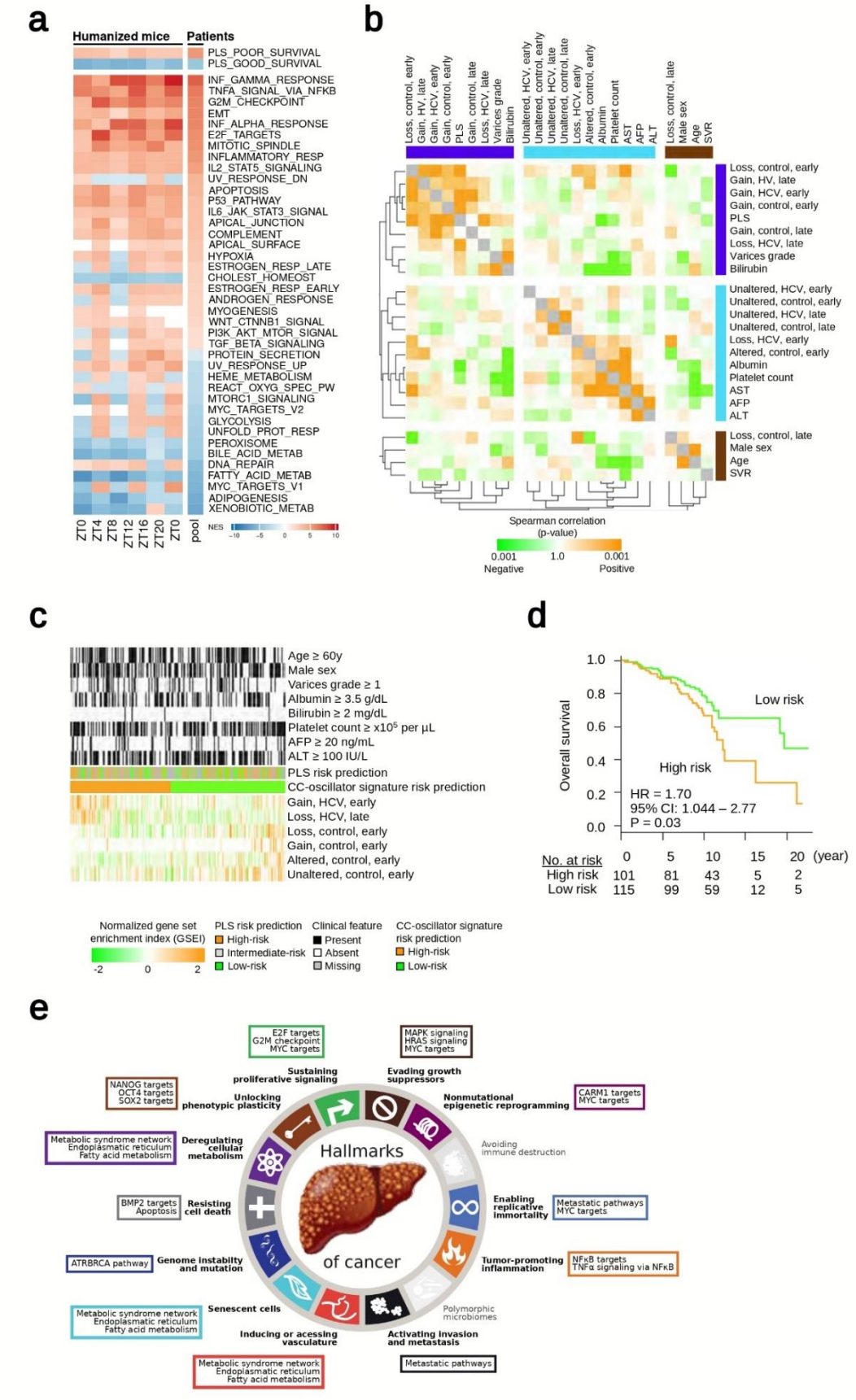
Article 2: An atlas of the human liver circadian transcriptome and its carcinogenic perturbation by hepatitis C virus infection

Figure 4



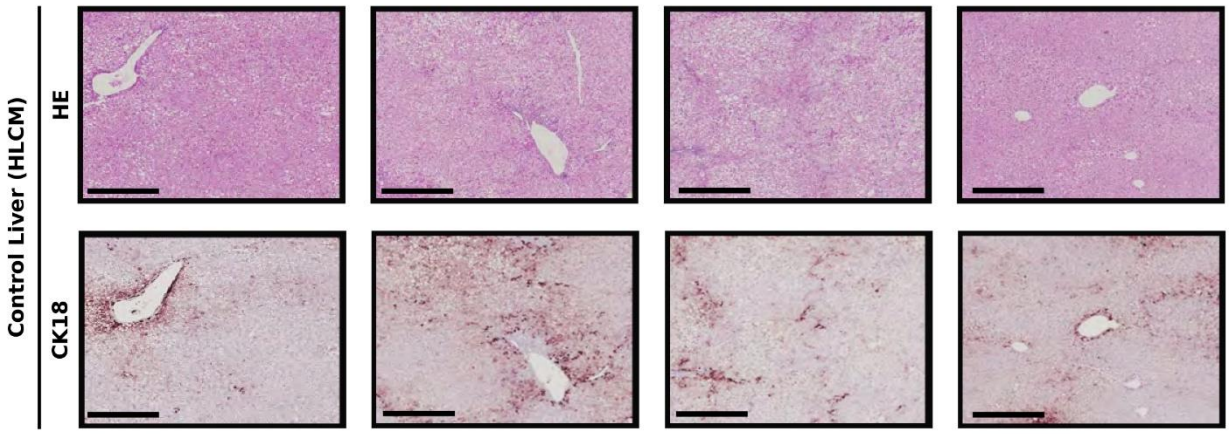
Article 2: An atlas of the human liver circadian transcriptome and its carcinogenic perturbation by hepatitis C virus infection

Figure 5



Article 2: An atlas of the human liver circadian transcriptome and its carcinogenic perturbation by hepatitis C virus infection

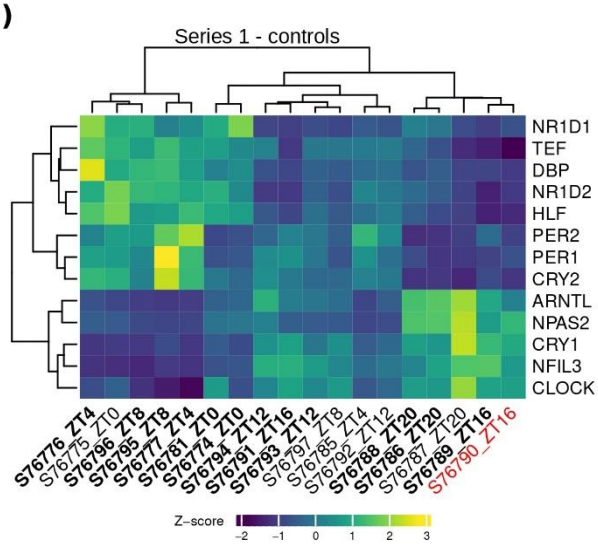
Extended Data Fig. 1:



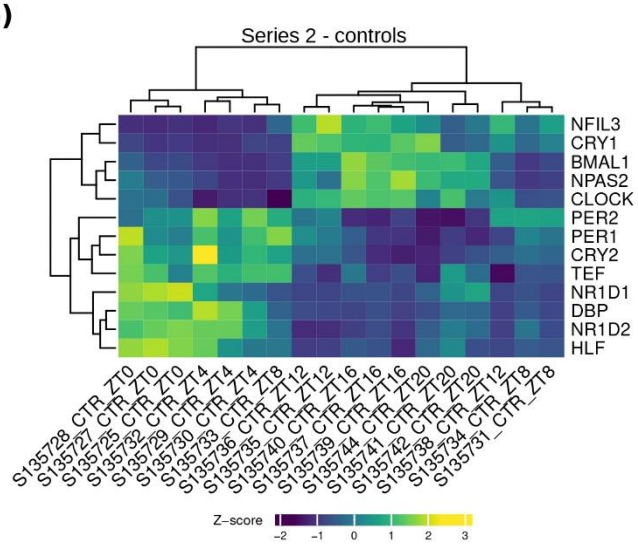
Article 2: An atlas of the human liver circadian transcriptome and its carcinogenic perturbation by hepatitis C virus infection

Extended Data Fig. 2:

a)

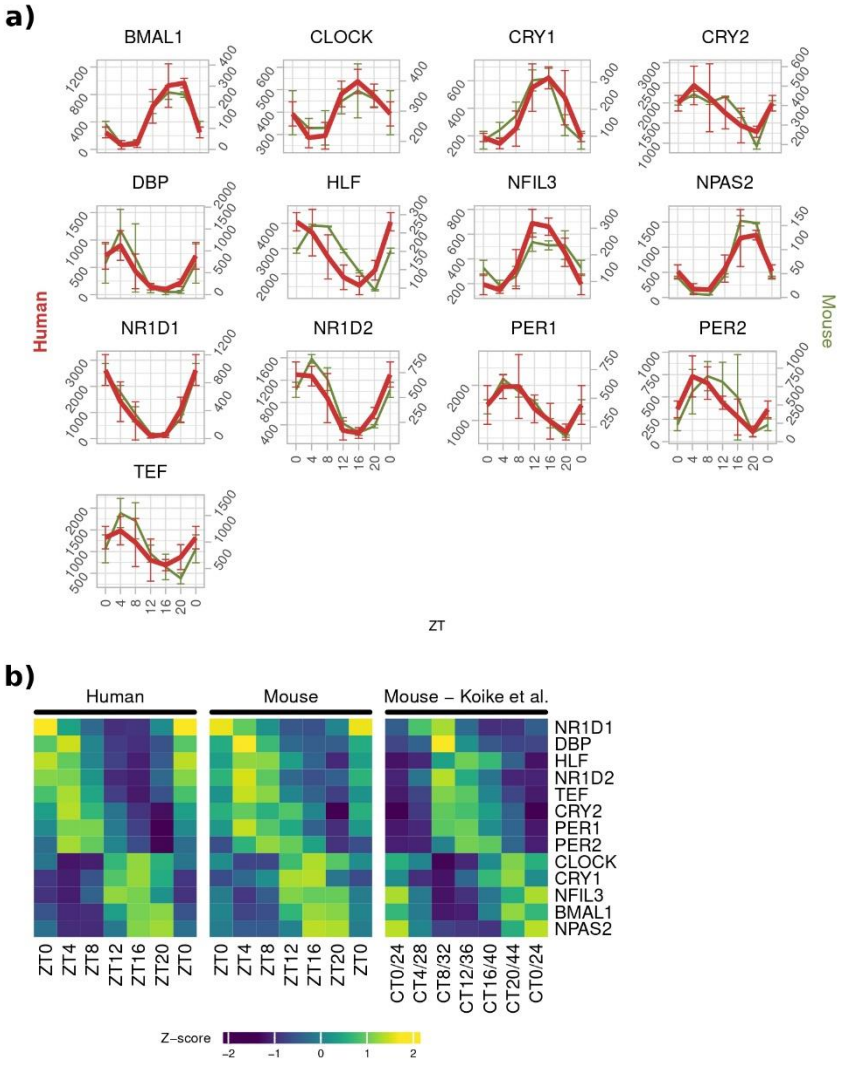


b)



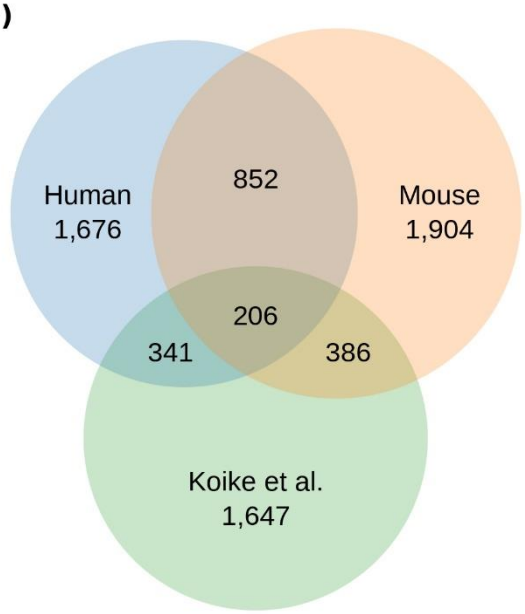
Article 2: An atlas of the human liver circadian transcriptome and its carcinogenic perturbation by hepatitis C virus infection

Extended Data Fig. 3:

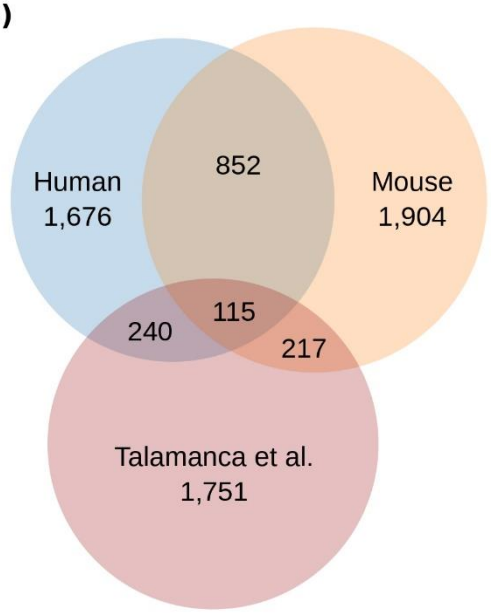


Extended Data Fig. 4:

a)

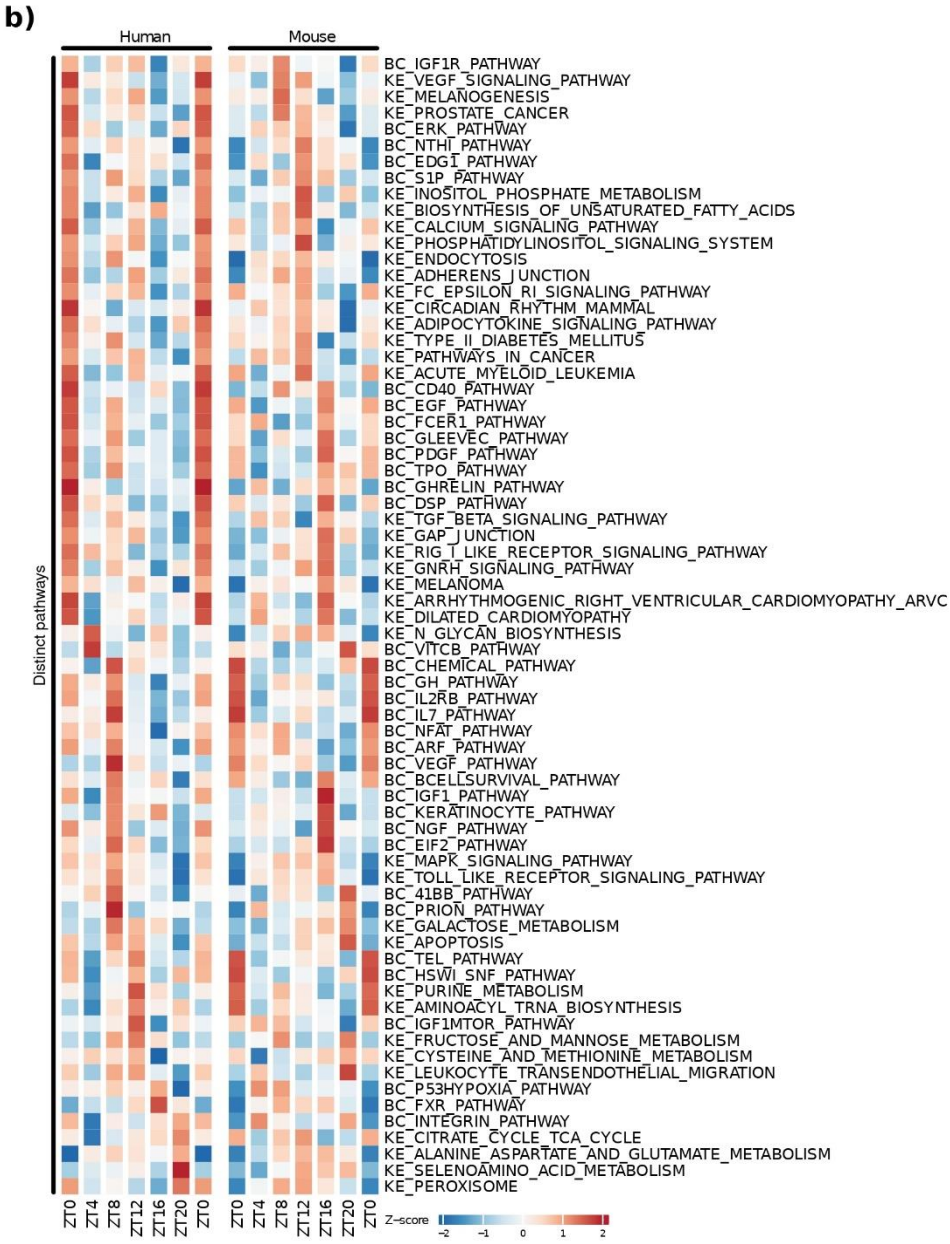


b)

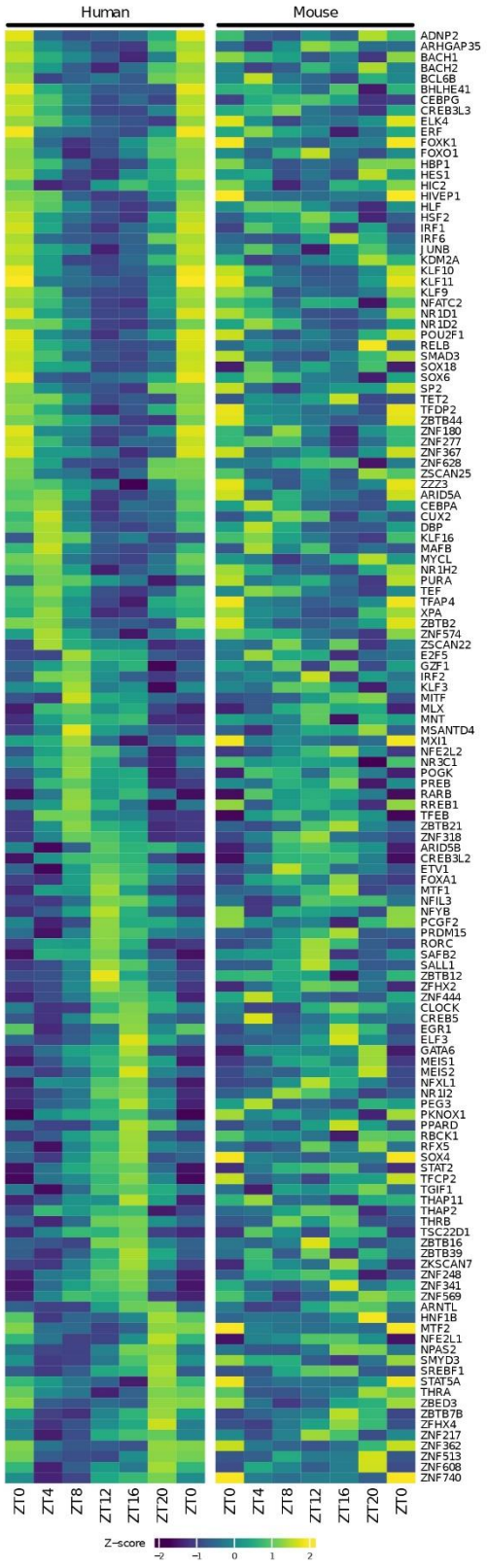


Extended Data Fig. 5:

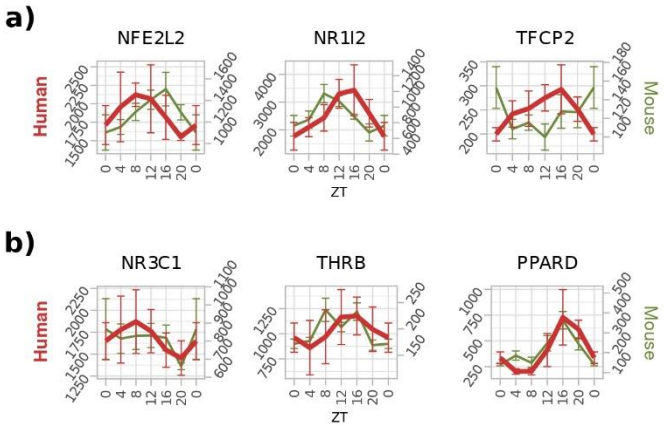
a)


b)
Human
Mouse
Distinct pathways
BC_IGF1R_PATHWAY
KE_VEGF_SIGNALING_PATHWAY
KE_MELANOGENESIS
KE_PROSTATE_CANCER
BC_ERK_PATHWAY
BC_NTHI_PATHWAY
BC_EDG1_PATHWAY
BC_S1P_PATHWAY
KE_INOSITOL_PHOSPHATE_METABOLISM
KE_BIOSYNTHESIS_OF_UNSATURATED_FATTY_ACIDS
KE_CALCIUM_SIGNALING_PATHWAY
KE_PHOSPHATIDYLINOSITOL_SIGNALING_SYSTEM
KE_ENDOCYTOSIS
KE_ADHERENS_JUNCTION
KE_FC_EPSILON_RI_SIGNALING_PATHWAY
KE_CIRCADIAN_RHYTHM_MAMMAL
KE_ADIPOCYTOKINE_SIGNALING_PATHWAY
KE_TYPE_II_DIABETES_MELLITUS
KE_PATHWAYS_IN_CANCER
KE_ACUTE_MYELOID_LEUKEMIA
BC_CD40_PATHWAY
BC_EGF_PATHWAY
BC_FCER1_PATHWAY
BC_GLEEVEC_PATHWAY
BC_PDGF_PATHWAY
BC_TPO_PATHWAY
BC_GHRELIN_PATHWAY
BC_DSP_PATHWAY
KE_TGF_BETA_SIGNALING_PATHWAY
KE_GAP_JUNCTION
KE_RIG_I_LIKE_RECEPTOR_SIGNALING_PATHWAY
KE_GNRH_SIGNALING_PATHWAY
KE_MELANOMA
KE_ARRHYTHMOGENIC_RIGHT_VENTRICULAR_CARDIOMYOPATHY_ARVC
KE_DILATED_CARDIOMYOPATHY
KE_N_GLYCAN_BIOSYNTHESIS
BC_VITCB_PATHWAY
BC_CHEMICAL_PATHWAY
BC_GH_PATHWAY
BC_IL2RB_PATHWAY
BC_IL7_PATHWAY
BC_NFAT_PATHWAY
BC_ARF_PATHWAY
BC_VEGF_PATHWAY
BC_BCELLSURVIVAL_PATHWAY
BC_IGF1_PATHWAY
BC_KERATINOCYTE_PATHWAY
BC_NGF_PATHWAY
BC_EIF2_PATHWAY
KE_MAPK_SIGNALING_PATHWAY
KE_TOLL_LIKE_RECEPTOR_SIGNALING_PATHWAY
BC_41BB_PATHWAY
BC_PRION_PATHWAY
KE_GALACTOSE_METABOLISM
KE_APOPTOSIS
BC_TEL_PATHWAY
BC_HSWI_SNF_PATHWAY
KE_PURINE_METABOLISM
KE_AMINOACYL_TRNA_BIOSYNTHESIS
BC_IGF1MTOR_PATHWAY
KE_FRUCTOSE_AND_MANNOSE_METABOLISM
KE_CYSTEINE_AND_METHIONINE_METABOLISM
KE_LEUKOCYTE_TRANSENDOTHELIAL_MIGRATION
BC_P53HYPOXIA_PATHWAY
BC_FXR_PATHWAY
BC_INTEGRIN_PATHWAY
KE_CITRATE_CYCLE_TCA_CYCLE
KE_ALANINE_ASPARTATE_AND_GLUTAMATE_METABOLISM
KE_SELENOAMINO_ACID_METABOLISM
KE_PEROXISOME
ZT0 ZT4 ZT8 ZT12 ZT16 ZT20 ZT0
ZT0 ZT4 ZT8 ZT12 ZT16 ZT20 ZT0
Z-score
−2 −1 0 1 2

Extended Data Fig. 6:

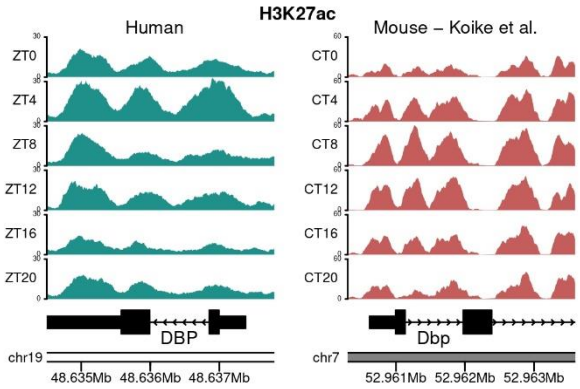


Extended Data Fig. 7:



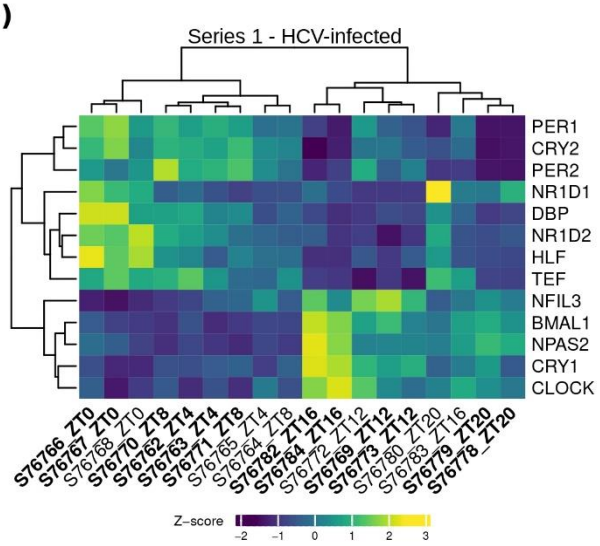
Article 2: An atlas of the human liver circadian transcriptome and its carcinogenic perturbation by hepatitis C virus infection

Extended Data Fig. 8:

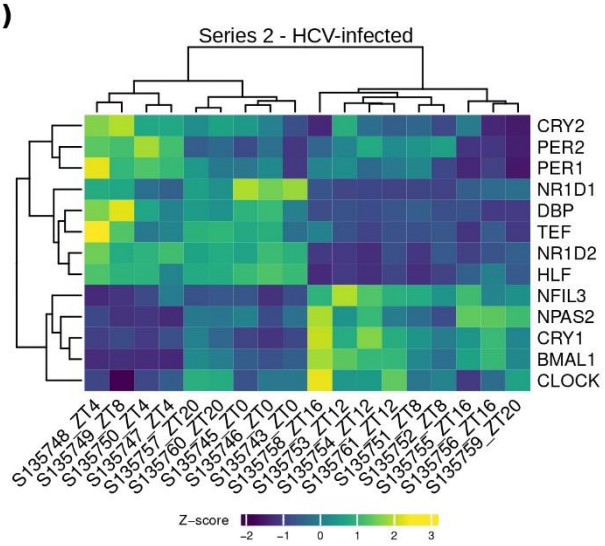


Extended Data Fig. 9:

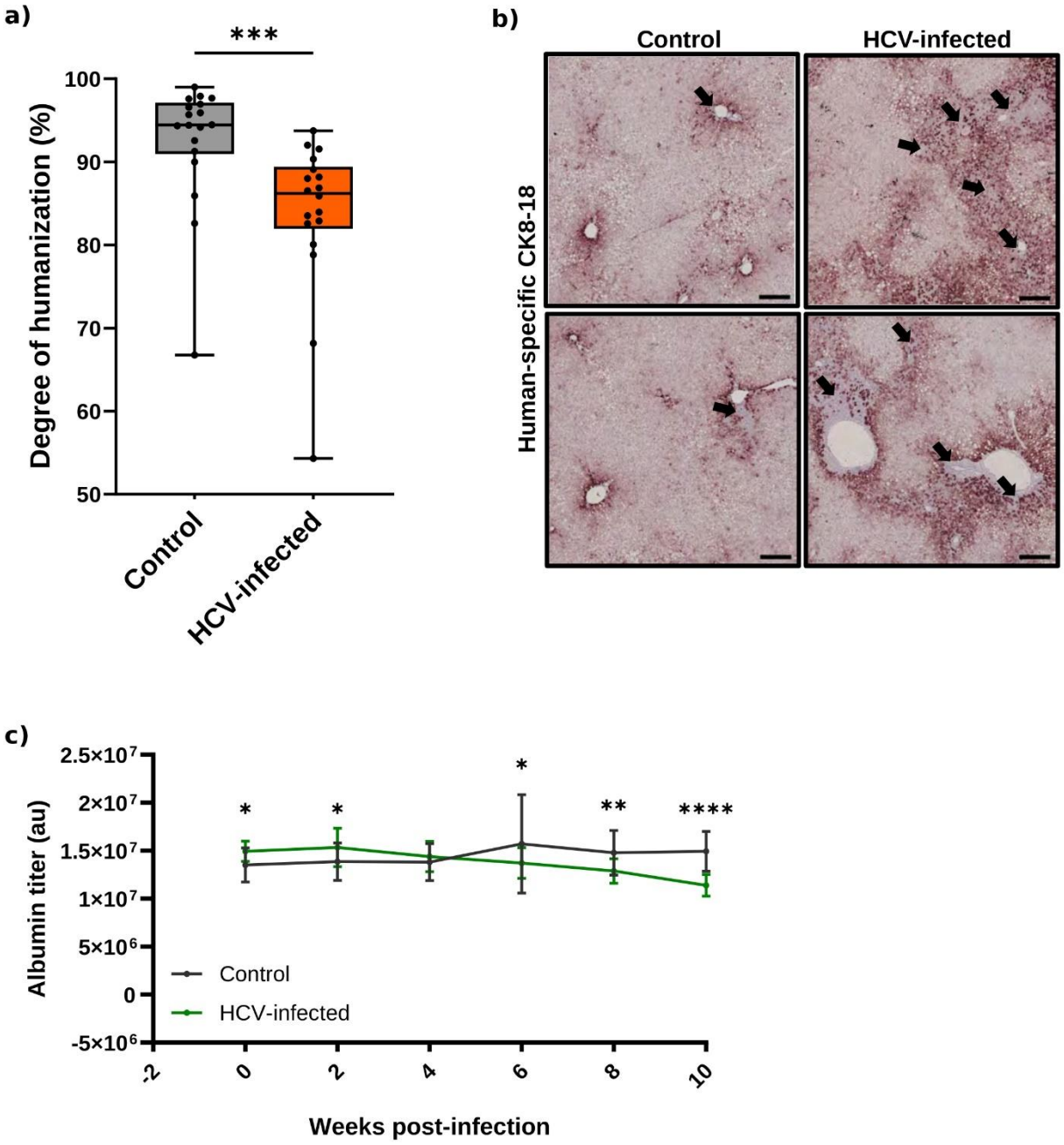
a)



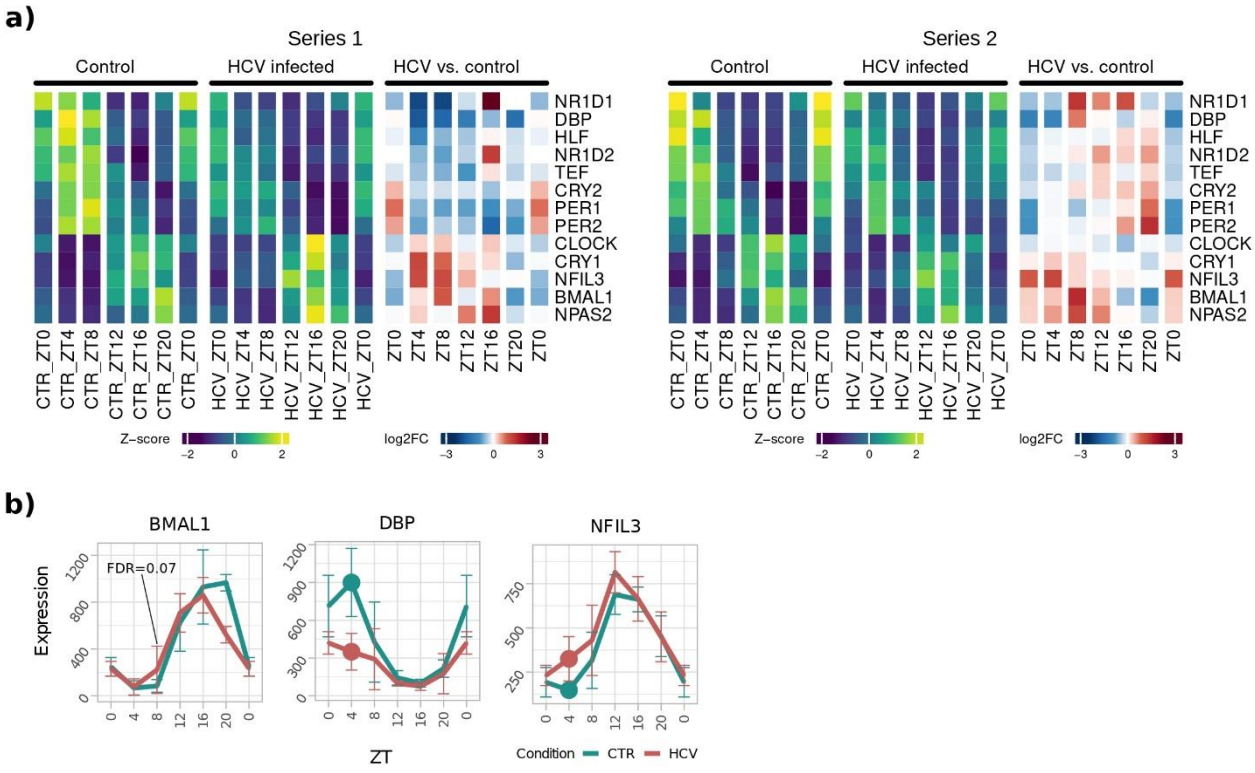
b)



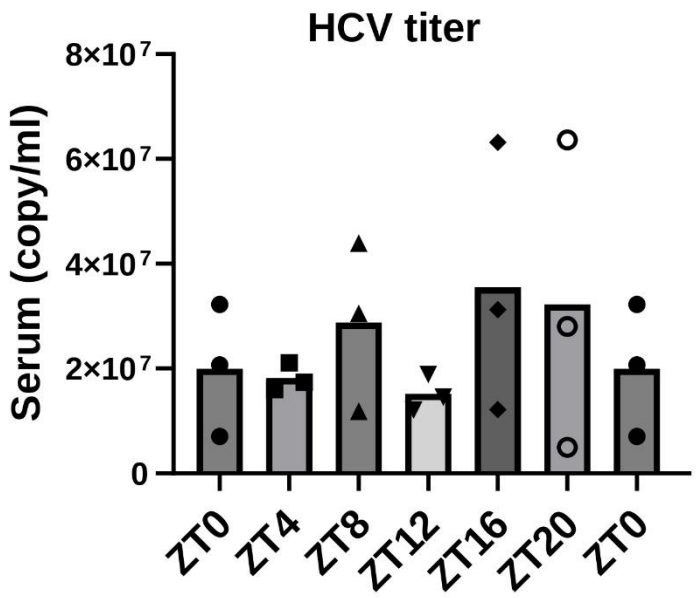
Extended Data Fig. 10:



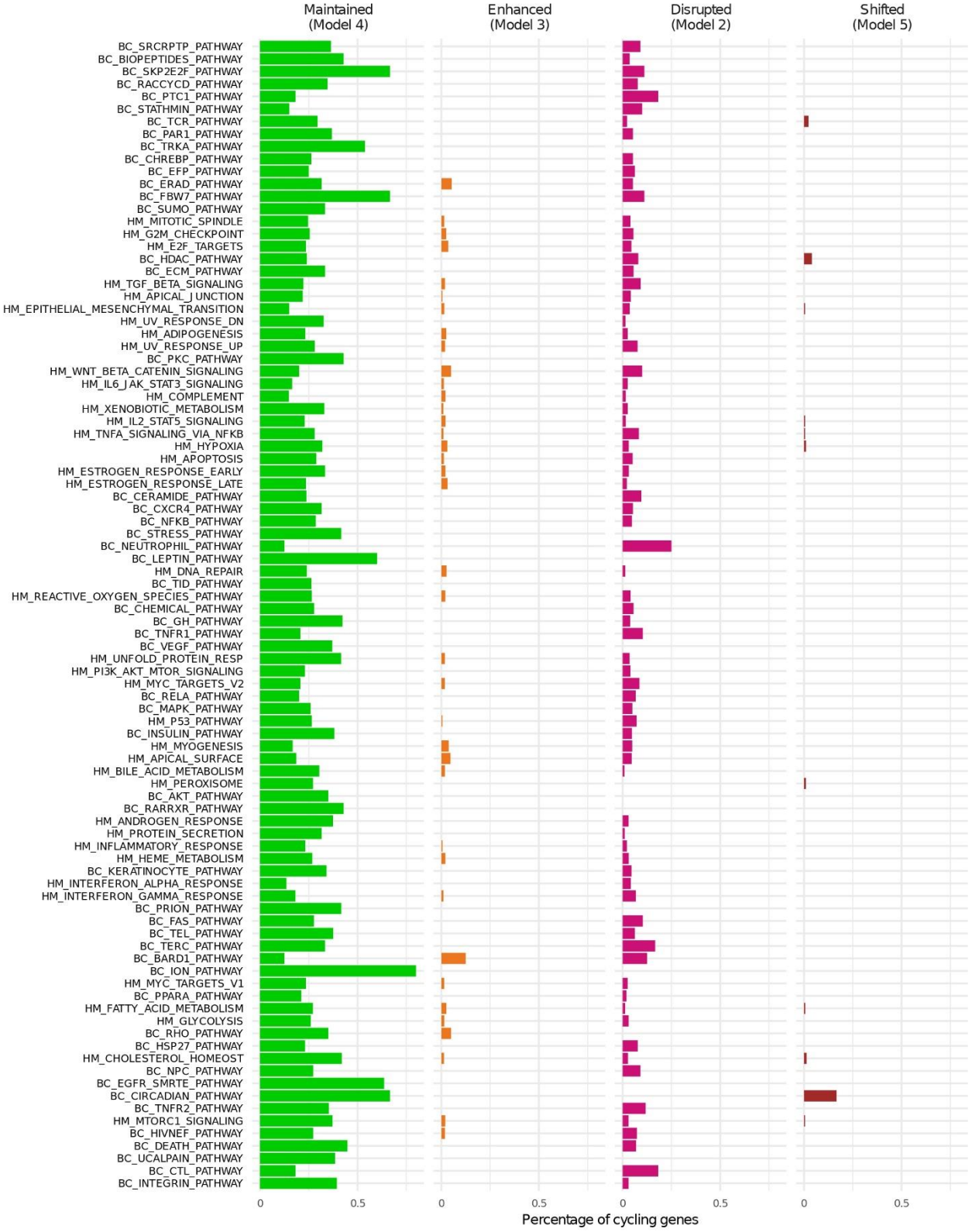
Extended Data Fig. 11:



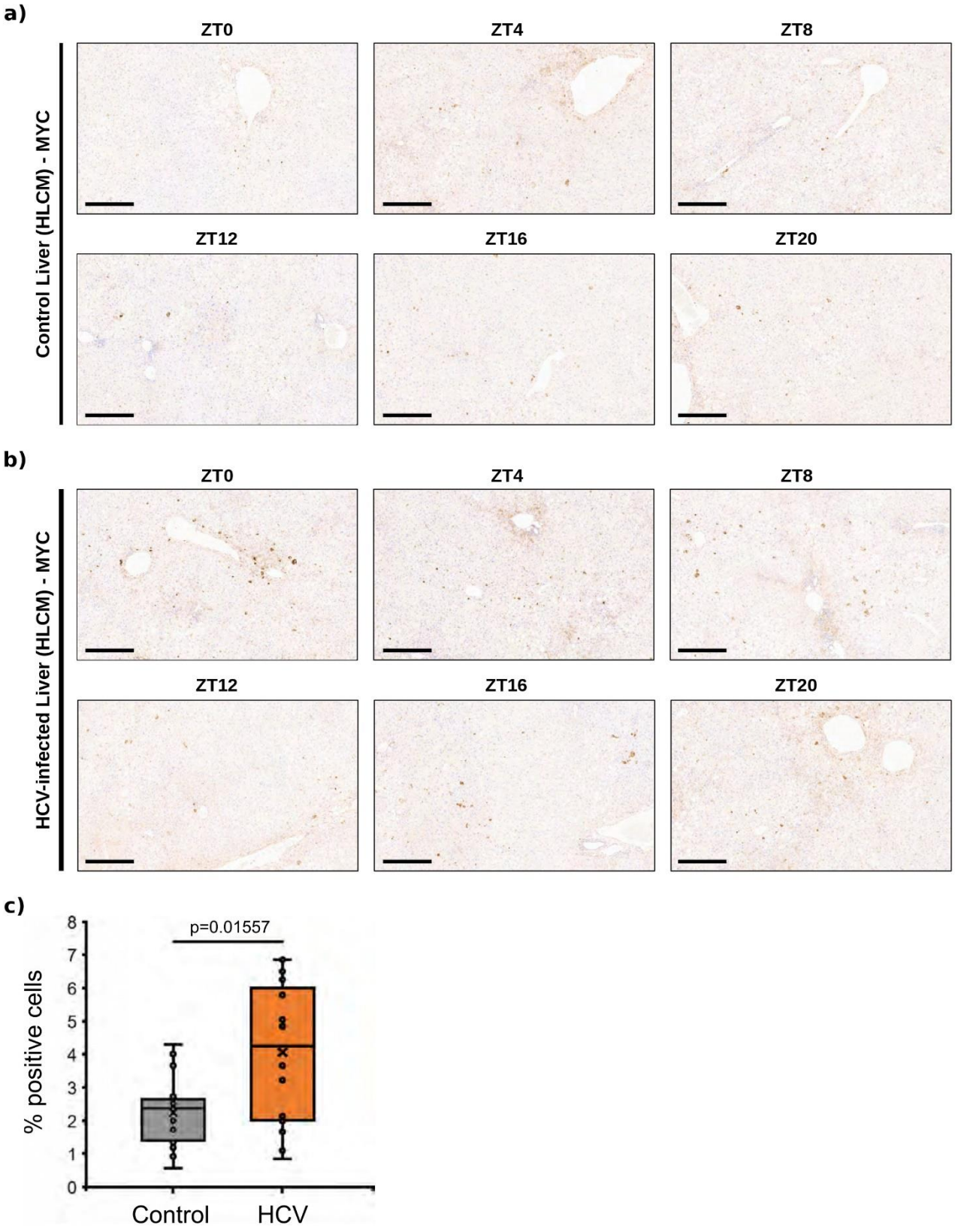
Extended Data Fig. 12:



Extended Data Fig. 13:

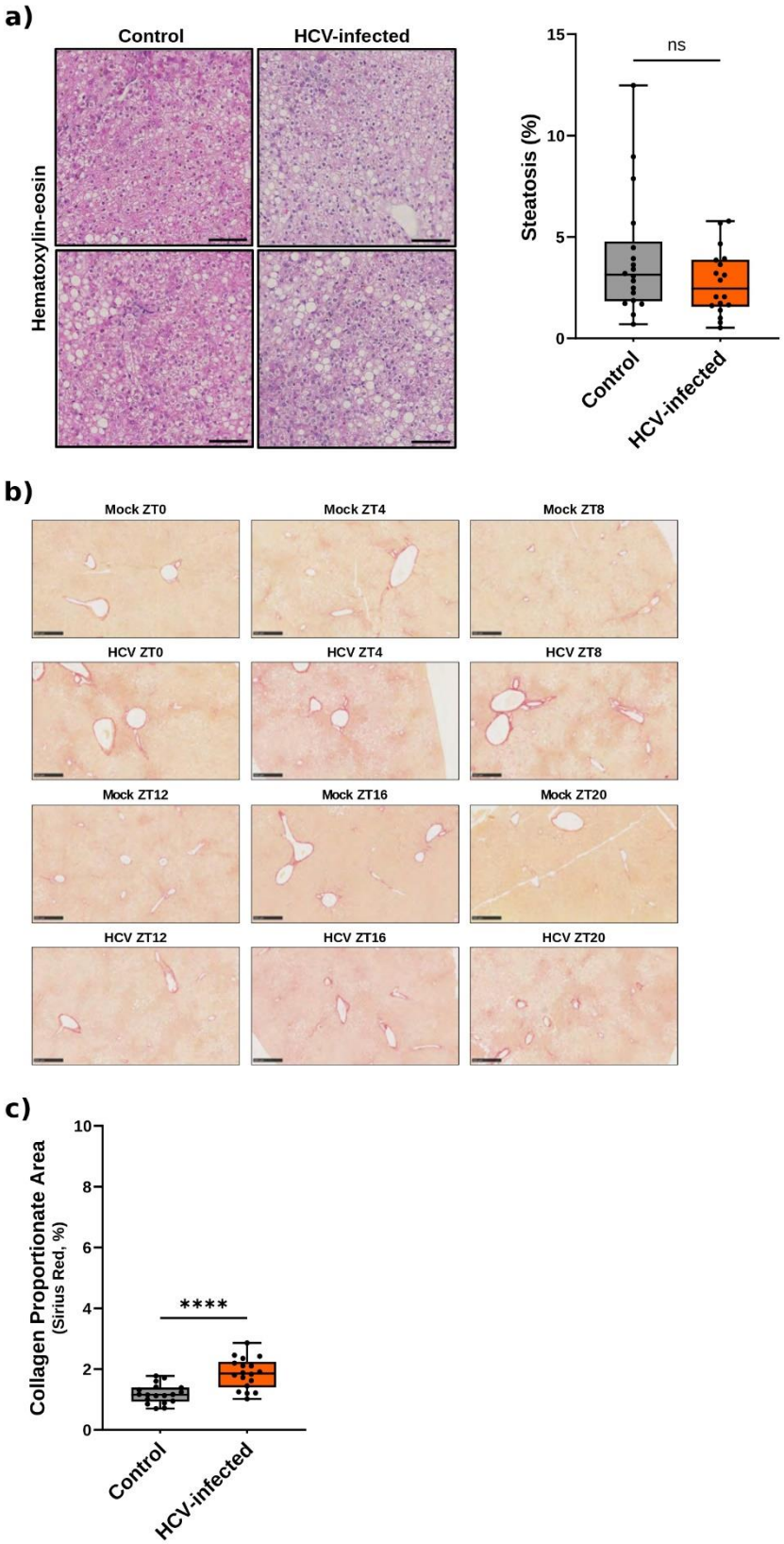


Extended Data Fig. 14:

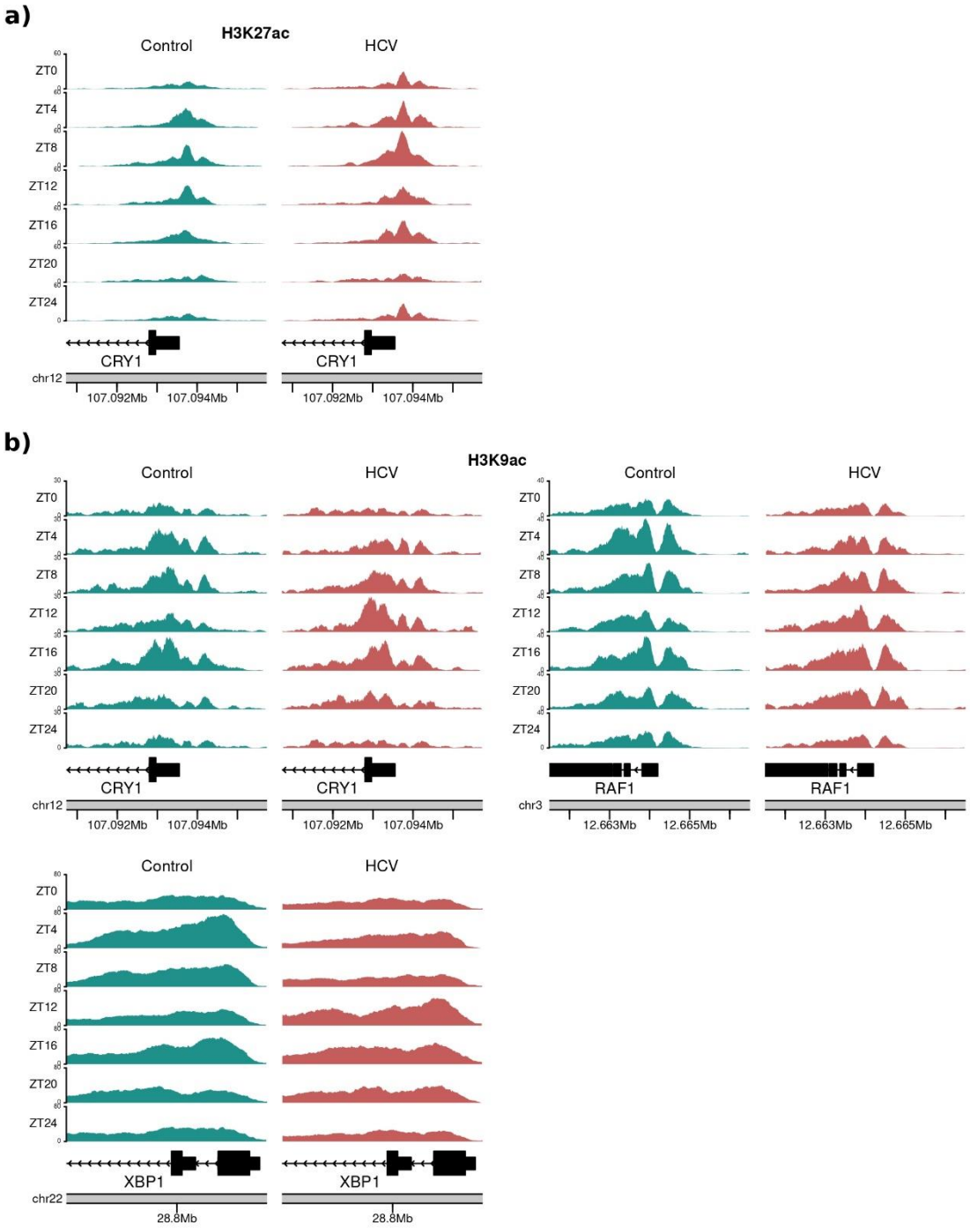


Article 2: An atlas of the human liver circadian transcriptome and its carcinogenic perturbation by hepatitis C virus infection

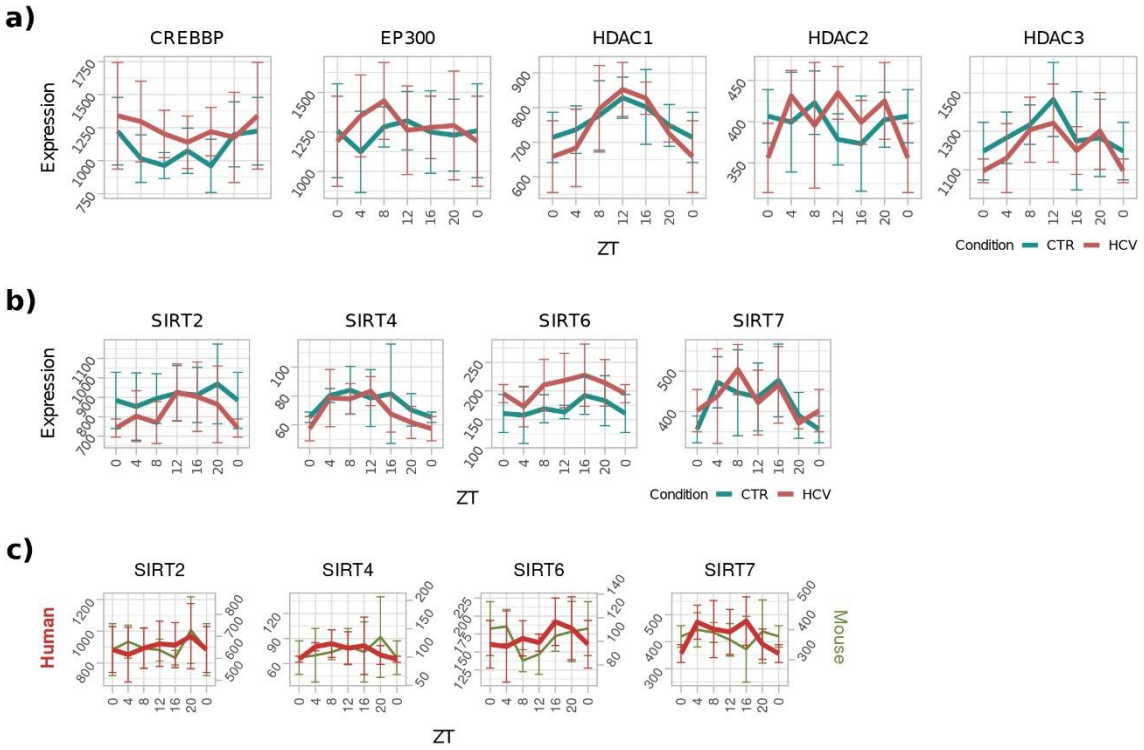
Extended Data Fig. 15:



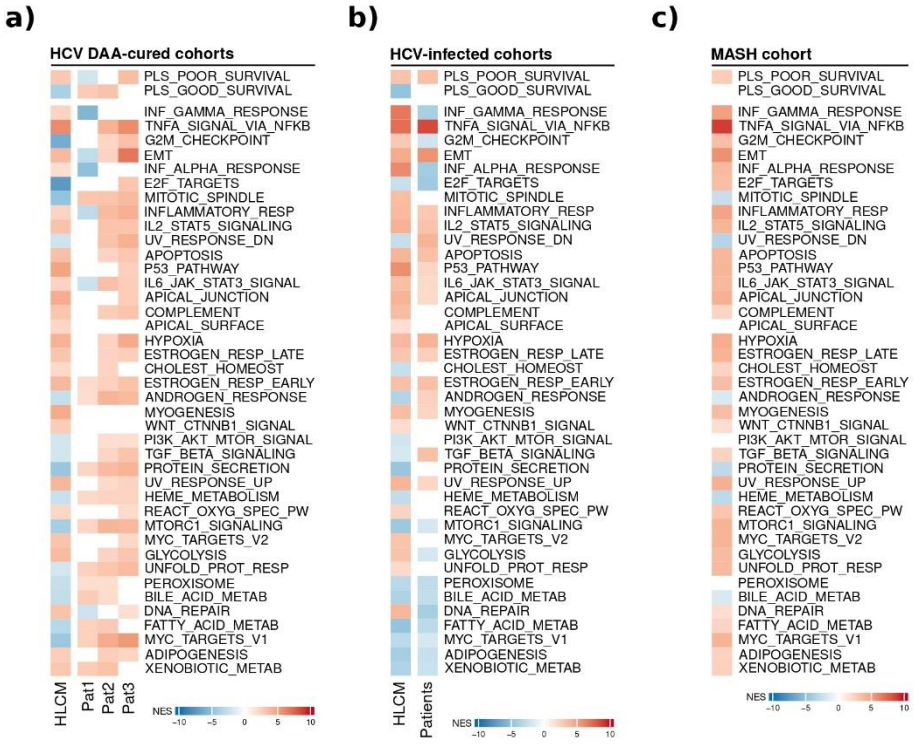
Extended Data Fig. 16:



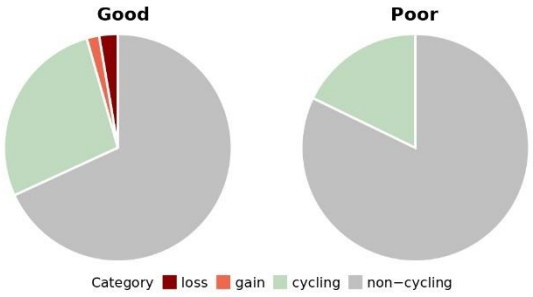
Extended Data Fig. 17:



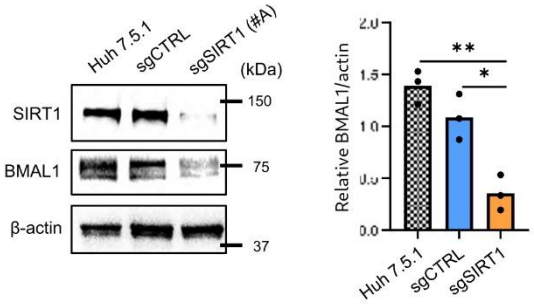
Extended Data Fig. S18:

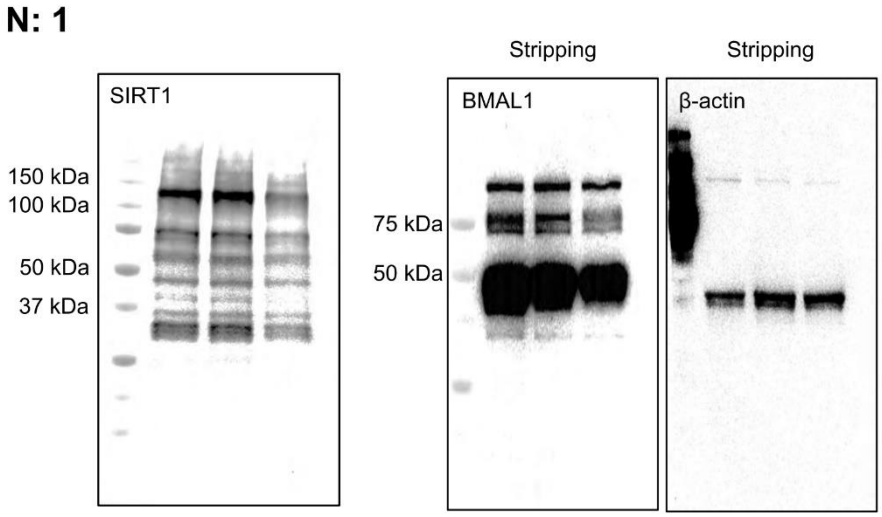


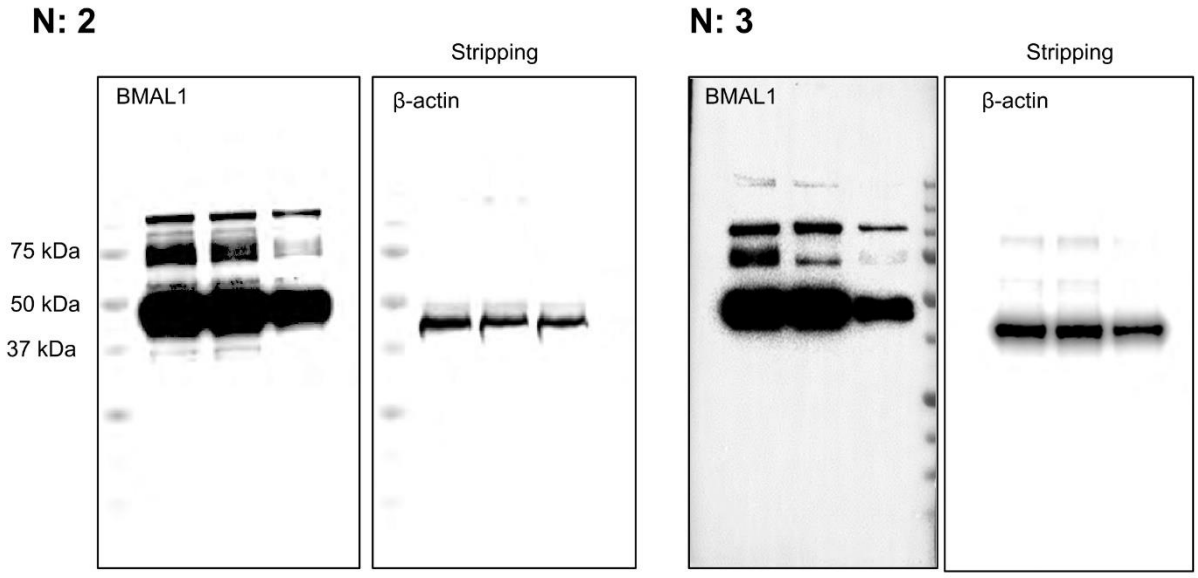
Extended Data Fig. 19:



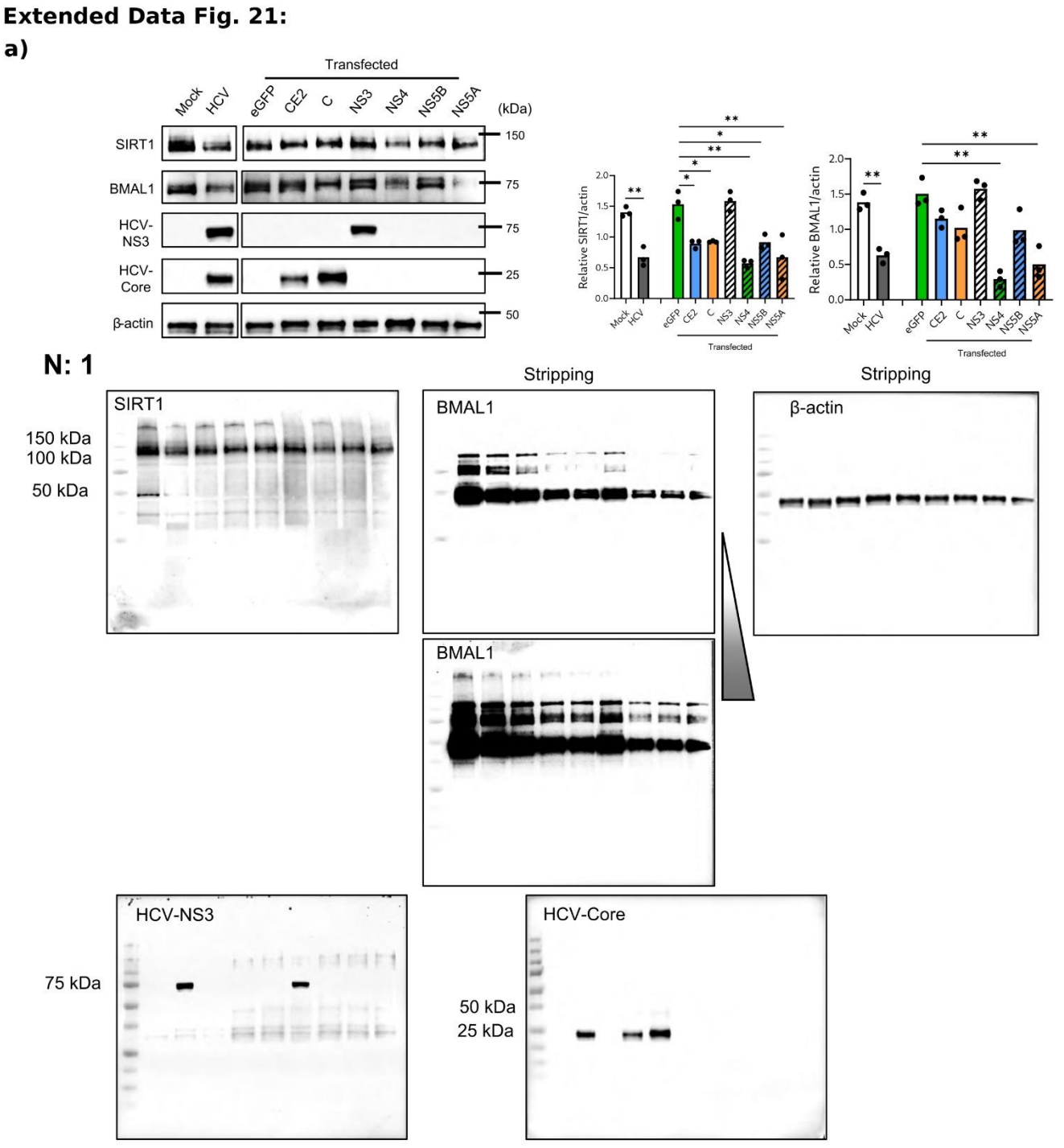
Extended Data Fig. 20:

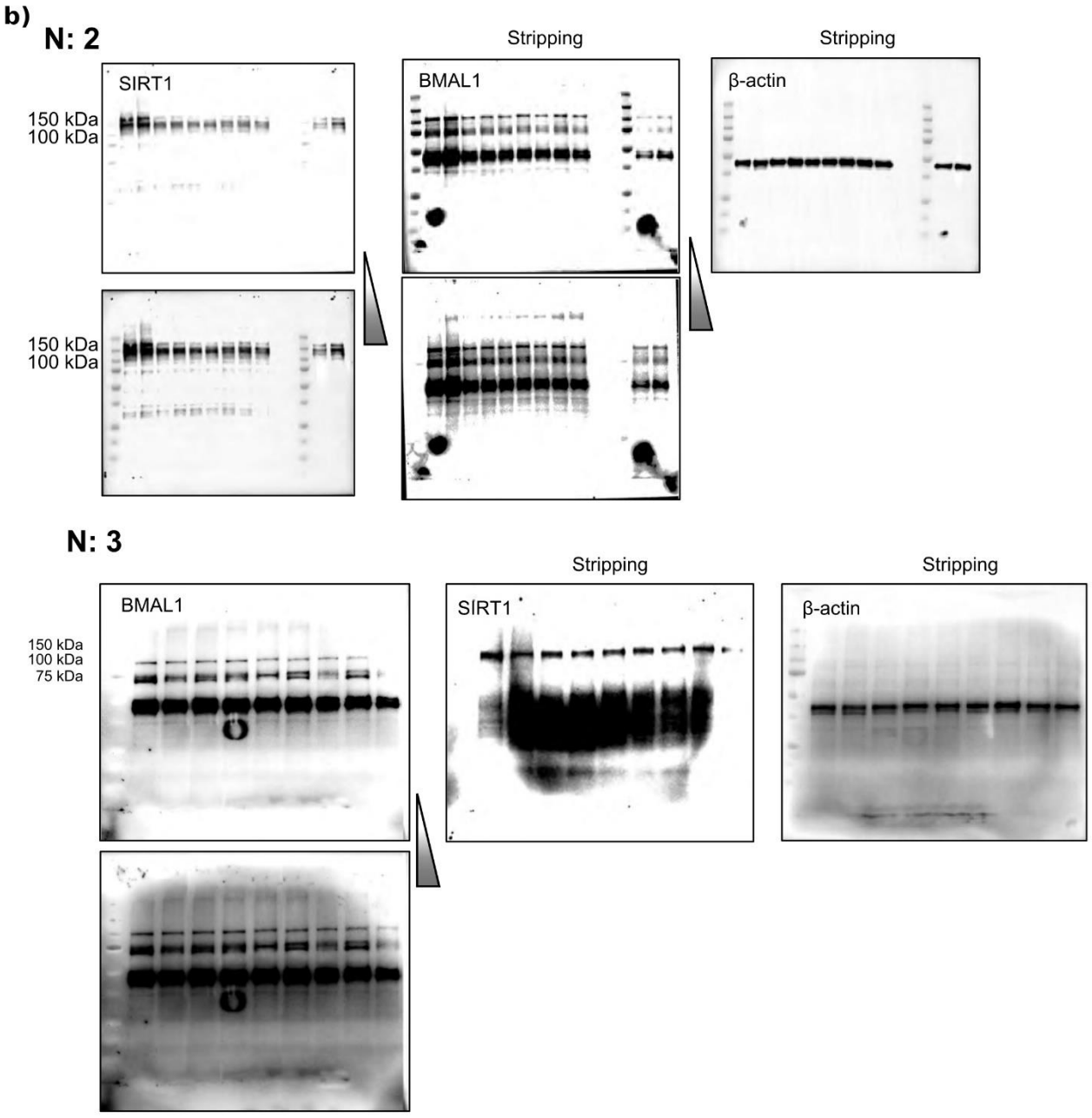






Extended Data Fig. 21:




b)
N: 2
Stripping
Stripping
SIRT1
BMAL1
β-actin
150 kDa
100 kDa
150 kDa
100 kDa
N: 3
Stripping
Stripping
BMAL1
SIRT1
β-actin
150 kDa
100 kDa
75 kDa

Article 3: TFE3 fusion proteins drive oxidative metabolism, ferroptosis resistance and general RNA synthesis in translocation renal cell carcinoma

This article has been submitted to different journals, we are waiting for their responses. It describes the role of TFE3, a transcription factor of the MiT family (with MITF) in renal cell carcinoma displaying TFE3-fusion transcripts with NONO and PRCC (tRCC). These fusion proteins bind elements across the genome, promoting ferroptosis resistance and OXPHOS leading to abnormal cell proliferation.

I participated in the characterization of the TFE3-fusion cell lines sensitivity to RSL3, a GPX4 inhibitor which induces ferroptosis. Strikingly, tRCC cell lines are sensitive to ferroptosis as they show same IC50 for RSL3 than A375 melanoma cells, known to be sensitive to ferroptosis on the opposite of the melanocytic 501mel cells. This result showed that TFE3-fusion proteins drive ferroptosis resistance, which can be overcome by treatment with RSL3.

Article 3: TFE3 fusion proteins drive oxidative metabolism, ferroptosis resistance and general RNA synthesis in translocation renal cell carcinoma

TFE3 fusion proteins drive oxidative metabolism, ferroptosis resistance and general RNA synthesis in translocation renal cell carcinoma

Alexandra Helleux[1,2,3,4], Guillaume Davidson[1,2,3,4], Antonin Lallement[1,2,3,4], Alexandre Haller[1,2,3,4], Isabelle Michel[1,2,3,4], Anas Fadloun[1,2,3,4], Christelle Thibault-Carpentier[1,2,3,4], Xiaoping Su[5], Véronique Lindner[6], Thibaut Tricard[7], Hervé Lang[7], Nizar Tannir[8], Irwin Davidson[1,2,3,4,10] and Gabriel G. Malouf[1,2,3,4,9,10]

1. Institut de Génétique et de Biologie Moléculaire et Cellulaire, BP 163, 67404 Illkirch, France

2. Centre National de la Recherche Scientifique, UMR7104, 67404 Illkirch, France

3. Institut National de la Santé et de la Recherche Médicale, U1258, 67404 Illkirch, France

4. Université de Strasbourg, 67404 Illkirch, France

5. Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

6. Department of Pathology, Strasbourg University Hospital, Strasbourg, France

7. Department of Urology, Strasbourg University Hospital, Strasbourg, France

8. Department of Genitourinary Medical Oncology, The University of Texas MD Anderson Cancer Center Houston, Texas, USA

9. Department of Medical Oncology, Institut de Cancérologie de Strasbourg-Europe, Strasbourg, France

10. These authors should be considered equally as co-senior authors.

Keywords: translocation renal cell carcinoma, TFE3, Metabolism, Ferroptosis, RNA synthesis, cancer associated fibroblasts

Conflict of interest statement: G.M. reported grant from INCA, Fondation ARC and MSD AVENIR. I.D. reported grant from la Ligue Contre le Cancer.

Article 3: TFE3 fusion proteins drive oxidative metabolism, ferroptosis resistance and general RNA synthesis in translocation renal cell carcinoma

Abstract

The oncogenic mechanisms by which TFE3 fusion proteins drive translocation renal cell carcinoma (tRCC) are poorly characterised. Here, we integrated loss and gain of function experiments with multi-omics analyses in tRCC cell lines and patient tumors. High nuclear accumulation of NONO-TFE3 or PRCC-TFE3 fusion proteins promotes their broad binding across the genome at H3K27ac-marked active chromatin, engaging a core set of M/E-box-containing regulatory elements to activate specific gene expression programs as well as promiscuous binding to active promoters to stimulate mRNA synthesis. Within the core program, TFE3 fusions directly regulate genes involved in ferroptosis resistance and oxidative phosphorylation metabolism (OxPhos) increasing functional OxPhos levels. Consequently, human tRCC tumors display high OxPhos scores that persist during their epithelial to mesenchymal transition (EMT). We further show that tRCC tumour aggressiveness is related to their EMT which although enriched in tumours harbouring ASPSCR1-TFE3 fusions, is also a frequent feature of tumours harbouring several other TFE3 fusions. In tRCC tumours, the presence of mesenchymal tRCC cancer cells associates with that of myofibroblast cancer-associated fibroblasts (myCAFs) that are both hallmarks of poor prognostic outcomes. We define tRCC as a novel metabolic subtype of renal cancer and provide unique insights into how broad genomic binding of TFE3 fusion proteins promotes tRCC tumourigenesis by regulating OxPhos and ferroptosis resistance and more generally stimulating RNA synthesis.

Significance.

This study advances understanding of the molecular mechanisms underlying oncogenic transformation by TFE3 fusion proteins in translocation renal cell carcinoma and defines features of tumour cells and their microenvironment that negatively impact patient outcome.

Article 3: TFE3 fusion proteins drive oxidative metabolism, ferroptosis resistance and general RNA synthesis in translocation renal cell carcinoma

Integrative multi-omics analyses reveal how extensive genomic TFE3 fusion protein binding drives oncogenic programs and identify novel potential therapeutic targets.

Article 3: TFE3 fusion proteins drive oxidative metabolism, ferroptosis resistance and general RNA synthesis in translocation renal cell carcinoma

Introduction

TFE3 and TFEB are members of the Microphthalmia-associated Transcription (MiT) family that regulate lysosomal function, nutrient/stress response and drug resistance [1]. Translocation renal cell carcinomas (tRCC) are characterised by genetic translocation events involving one member of the (MiT) family, most frequently *TFE3* (90% of cases) [2,3]. TRCC is a rare subtype of kidney cancer that represents 1-4% of all adults RCC[4,5], but is the most frequent RCC arising in children and adolescents [1,6,7]. The translocations result in expression of chimeric fusion proteins comprising TFE3 and another gene partner, under the control of constitutively active gene promoters[8]. The complexity of tRCC is underscored by the identification of over 20 TFE3-partners, accentuating histological and clinical heterogeneity. Among these, *SFPQ*, *ASPSCR1*, *PRCC*, and *NONO* are the most frequent fusion partners. Aside these translocations, no other highly recurrent mutations have been identified, suggesting that the fusion proteins are the primary drivers of transformation. In agreement with this, all fusion proteins retain the TFE3 DNA binding and dimerization (basic-helix-loop-helix leucine-zipper; bHLH-LZ) domains allowing them to bind to cis-regulatory elements in the genome. Despite recent recognition of tRCC as a distinct RCC subtype, the precise oncogenic mechanisms fueling tRCC tumorigenesis remain elusive. Consequently, there is an urgent medical need to unravel the molecular function of TFE3-fusion proteins to comprehensively understand the molecular basis of tRCC and identify novel therapeutic targets.

To better understand how TFE3 fusion proteins drive transformation in tRCC, we performed gain and loss of function experiments complemented by multi-omics approaches to profile genome occupancy of TFE3 fusions, identify active cis-regulatory elements and characterize the gene expression programs under their control in model cell lines and human tumours. Our findings revealed that NONO-TFE3 and PRCC-TFE3 fusions exhibit broad binding to thousands of sites across the genome, engaging distal and proximal M/E-box-

containing regulatory elements, as well as promiscuously associating with active promoters. Amongst the core regulated genes are those driving oxidative phosphorylation (OxPhos) metabolism, culminating in elevated OxPhos levels in tRCC cells, and those promoting ferroptosis resistance. Analyses of RNA-seq of tRCC patient tumors demonstrated heightened OxPhos scores confirming the observations from the cell lines. Remarkably, elevated OxPhos score persisted in tumors that had undergone epithelial to mesenchymal transition (EMT), a known marker of tumor aggressiveness. Furthermore, using transcriptional signatures derived from clear cell renal carcinoma (ccRCC), we revealed that tRCC tumors with EMT signatures displayed increased levels of myofibroblastic cancer-associated fibroblasts (myCAFs) that correlated with poorer survival. Thus, EMT and the presence of myCAFs emerge as common hallmarks, adversely impacting the prognosis for both ccRCC and tRCC. We further identified a signature of TFE3-fusion regulated genes, strongly associated with poor patient outcome, including novel potential therapeutic targets.

Results

Abundant constitutive nuclear expression of TFE3 fusion proteins is essential for tRCC cell line viability

To address the role of TFE3 fusion proteins in tRCC, we used 4 cell lines expressing either NONO-TFE3 (lines UOK109, TF1) or PRCC-TFE3 (UOK120, UOK146). Each fusion protein presents a different breakpoint within TFE3, but conserve its bHLH-LZ domain maintaining their ability to specifically bind DNA. To assess expression of TFE3 fusion proteins compared to native TFE3, we performed immunoblots on extracts from the above lines as well as 2 non-transformed renal cell lines (RPTEC, HEK293T), 4 clear-cell (cc)RCC lines (RCC4, UOK121, A-498, 786-O) and 2 papillary (p)RCC lines (UOK112, ACHN). High expression of NONO-TFE3 and lower expression of PRCC-TFE3 fusions in the corresponding

tRCC lines was observed, while expression of native TFE3 in the other cell lines was variable, but considerably lower than the fusion proteins **(Fig. S1A)**. In contrast, no native TFE3 protein was visible in tRCC lines consistent with the *TFE3* gene location on the X chromosome and hence in cells of male origin, UOK109, TF1 and UOK120, the single copy of *TFE3* is subject to translocation[9–11] whereas in the female UOK146 line expression of the non-translocated allele transcript is silenced[10]. In addition, biochemical fractionation of the nuclear and cytoplasmic compartments showed strong enrichment of TFE3 fusion proteins in the nucleus similar to what is observed with the control BRG1, the catalytic subunit of the SWI/SNF chromatin remodelling complex **(Figs. S1B)**. Thus, the tRCC cell lines abundantly express the expected nuclear TFE3-fusion proteins.

To characterize the gene expression programs of the tRCC cell lines, triplicate RNA-seq was performed for each line. More than 80% of the 3000 most highly expressed protein-coding genes were common to all 4 cell lines **(Fig. S1C)** and were enriched for pathways previously associated with function of native TFE3 such as lysosome function and endocytosis (**Fig. S1D**). In addition, various metabolic pathways were enriched notably those linked to the electron transport chain, oxidative phosphorylation (OxPhos), reactive oxygen species (ROS) and carbon metabolism **(Fig. S1D)**. Each line displayed hybrid EMT-type markers **(Fig. S1E)**. For example, while TF1 and UOK146 showed highest expression of the epithelial marker *EPCAM*, they overexpress the mesenchymal *PRXX1* and *SNAI1/2* genes, respectively. Of note, the EMT regulators *SNAI1/2* were mostly expressed in the PRRC-TFE3 lines as compared to *TWIST1* or *TWIST2* most expressed in the NONO-TFE3 lines. Overall, UOK120 appeared to be the most mesenchymal line with the lowest expression of most epithelial markers, and higher expression of multiple mesenchymal markers and/or regulators **(Fig. S1E)**.

To assess the role of TFE3 fusion proteins in the transformed phenotype, their expression was silenced using siRNA specifically directed against the TFE3 C-terminal region

and compared with non-targeting control siRNA. RT-qPCR indicated efficient knockdown of TFE3 in the 2 PRCC-TFE3 lines (UOK120, UOK146; >80% reduction) as compared to control. Less efficient silencing was observed in the NONO-TFE3 lines (UOK109, TF1) perhaps due to the higher fusion expression. (**Fig. 1A)**. Similar effects were seen on immunoblots with strongly reduced levels of fusion protein in UOK120 and UOK146 lines, with less attenuated effects in UOK109 and TF1 lines **(Fig. 1B)**. TFE3 fusion silencing dramatically reduced colony forming ability with a reduction reaching 90% for both NONO-TFE3 cell lines as compared to 60% and 80% for UOK120 and UOK146 lines, respectively **(Fig. 1C)**. Thus, high levels of fusion protein were essential for tRCC cell viability irrespective of their epithelial/mesenchymal states.

TFE3 fusions drive expression of genes involved in OxPhos

To identify genes regulated by TFE3 fusions, RNA-seq was performed on each cell line after siRNA-control (siCTR) or siRNA-TFE3 (siTFE3). Following silencing, the number of reads at the fusion junction and read coverage over TFE3 exons present in both NONO- and PRCC-TFE3 fusions were strongly decreased **(Figs. S2A-B)**. Following TFE3-fusion protein silencing, we observed up- or down-regulation of hundreds of genes (range: ~200-500) in each tRCC cell line using a standard cut-off (log2 fold change > +/- 1, and adjusted p-value <0.05, **Fig. S2B and Dataset S1**).

Gene Set Enrichment Analyses (GSEA) using the Hallmarks gene sets showed consistent down-regulation of the Oxidative Phosphorylation (OxPhos) pathway in all tRCC lines. In addition, we observed enrichment for down-regulation of oncogenic pathways in several cell lines such as MYC targets and pathways related to cell cycle and proliferation (E2F targets, G2M checkpoint; **Figs. 1D and S2C)**. Limited pathways were enriched for up-regulated genes (**Fig. 1D**). Among the 200 genes comprising the GSEA OxPhos signature (**Dataset S2**), between 115 and 139 genes were down-regulated depending on the line, with 54 genes commonly down-regulated in all 4 lines (**Fig. 1E).** These genes encode multiple subunits of the

mitochondrial respiration chain complexes I (*NDUF* (NADH:ubiquinone oxidoreductase), II (*SDHA/B*), III *UQC* (ubiquitol-cytochrome c oxidoreductase) and IV (*COX* cytochrome c-oxidase) (**Fig. 1F**). In addition, multiple subunits of the V-type proton ATPase, a complex responsible for acidification of intracellular organelles notably lysosomes, also comprised in the GSEA OxPhos pathway, were down-regulated by siTFE3. These observations suggested a direct role of TFE3 fusions in regulating expression of genes involved in OxPhos metabolism.

To evaluate whether these gene expression changes affected OxPhos, mitochondrial activity was measured by profiling the Oxygen Consumption Rate (OCR) using the Agilent SeaHorse (**Fig. 1G**). Comparison of OCR activity showed higher basal and maximum respiration capacities in 3 of the tRCC lines compared to the 786-O and UOK121 ccRCC lines, the exception being TF1 with higher basal, but not maximal OCR levels (**Fig. 1H)**.

To determine the contribution of the TFE3 fusions to the elevated OxPhos capacity, SeaHorse experiments were performed in siTFE3 versus siCTR cells 72h after transfection. Although basal OCR levels were unaffected, both maximum respiration and reserve capacity were significantly decreased after siTFE3 in UOK109, UOK146 and UOK120, with a tendency observed for TF1 line (**Fig. 1I)**. These data showed that TFE3-fusions drive expression of genes involved in OxPhos resulting in a higher contribution of OxPhos to the metabolism of tRCC compared to ccRCC cell lines.

Ectopic expression of TFE3 fusions leads to loss of renal identity and stimulates OxPhos.

To determine whether TFE3 fusions can directly activate OxPhos, we performed gain of function experiments using HEK293T embryonic kidney cells. HEK293T cells were infected with lentiviral particles containing a Doxycycline (Dox)-inducible expression vector for either native TFE3 (T), NONO-TFE3 (NT), PRCC-TFE3 (PT) or GFP control. 24 hours after Dox

addition, a strong induction was seen at both the mRNA (**Fig. 2A)** and protein expression levels with NONO-TFE3 accumulating to lower levels than PRCC-TFE3 (**Fig. 2B)**. Under these conditions, ectopic TFE3, NONO-TFE3 and PRCC-TFE3 proteins all accumulated in the nucleus (**Fig. 2C**).

To identify genes regulated by ectopic protein expression, we performed RNA-seq 24 hours after Dox induction. Ectopic TFE3 expression deregulated over 1000 genes as compared to the GFP control line using standard cut-off (log2FC>|1| and adj.p.val<0.05, **Fig. 2D and Dataset S3**). Most genes deregulated by native TFE3 expression were also deregulated using fusion proteins (**Figs. S3A-B)**, whereas additional gene sets either commonly or specifically deregulated by both fusion proteins were observed. Overall, 558 genes were up-regulated in all TFE3-expressing lines and linked to functions previously ascribed to native TFE3 such as lysosome, mTORC1-associated pathway or transport of small molecules through ion channels (**Fig. S3A**). Although fewer (71) common down-regulated genes were found, ontology showed they were linked to morphogenesis and development processes including kidney development suggesting that ectopic fusion protein expression may lead to loss of epithelial and kidney identity (**Fig. S3B)**.

While hallmark GSEA analysis showed few significant pathways for down-regulated genes, several pathways were enriched amongst up-regulated genes such mTORC1 signalling and immune signalling (TNFa signalling via NFkB, IFNg response; **Fig. 2E)**. Importantly and in line with loss of function experiments, OxPhos was up-regulated in all conditions (**Figs. 2E and S3C)**. Among the 200 genes in the OxPhos pathway, between 84 to 119 genes were up-regulated with 59 genes common to all 3 lines including those encoding subunits of complexes of the electron transport chain and the lysosomal V-ATPase, but also other genes involved in mitochondrial homeostasis and function (**Figs. 2F-G)**. In line with these observations, ectopic expression of TFE3, and to a greater extent TFE3 fusions enhanced basal and maximal OCR

respiration as well as reserve capacity compared to the GFP control line (**Figs. 2H-I)**. Thus, both loss and gain of function converged to show that TFE3 fusion proteins activated genes of the OxPhos pathway leading to enhanced OxPhos metabolism.

TFE3 fusions bind promiscuously at active promoters and stimulate RNA synthesis

To better understand how the fusion proteins regulate gene expression, we performed Cut&Tag (C&T) in all 4 tRCC cell lines to profile TFE3 fusion protein binding along with H3K27ac ChIP-seq to identify active cis-regulatory elements. Read density analyses of the non-redundant TFE3 peaks from all lines revealed a large number of low occupied sites that did not coincide with H3K27ac (Cluster 14 in **Fig. S4A**). We therefore focussed our analyses on the 84 765 non-redundant sites strongly occupied in at least one of the 4 lines (**Fig. 3A)**. Using this set of highly occupied sites, we found that number of TFE3 peaks was around 2-fold higher (65 000-78 000) in both NONO-TFE3 expressing lines compared to PRCC-TFE3 (33 000-42 000) lines, whereas comparable numbers (44 000-64 000) of H3K27ac ChIP-seq peaks were observed in each cell line (**Fig. S4B)**.

MEME-ChIP-analyses of the top 1000 TFE3 peaks in the UOK120 and UOK146 lines revealed strong enrichment in the M/E-box motif, that was the only enriched motif (**Fig. S4C**). In contrast, analogous analyses of the UOK109 and TF1 TFE3 peaks revealed enrichment of GC-rich SP1/KLF-type motifs in addition to the M/E-box motif. Analyses of motif locations showed that the M/E-box motif protected at the peak centre, whereas the other motifs were spread across the peaks. Similar results were found using RSAT software, where the M/E-box motif was highly represented in the UOK120 and UOK146 peaks, but was lower ranked in the other lines, where in addition the GC-rich SP1/KLF and NFY motifs were also detected (**Fig. S4D**). Thus, the fusion proteins bind to their cognate M/E-box motif, but additional indirect binding was seen, in particular of NONO-TFE3.

The GC-rich SP1/KLF and NFY-binding CAAT-box motifs are prevalent in proximal promoters with the CATT-box located around 70 nucleotides upstream of the transcription start site (TSS) [12]. Read density analyses showed that many of the 84 765 non-redundant were associated with H3K27ac either in a common or cell-specific manner (**Fig. 3A**) and consistent with the presence of SP1 and CATT-motifs, strong TFE3 fusion occupancy of ≈8 700 active H3K27ac-marked promoters was seen (**Fig. 3B**). TFE3 fusion binding was also seen at 19 040 promoter distal regions (**Fig. 3C**) with a set of sites commonly occupied in all 4 lines (clusters 1-14), but showing differential association with H3K27ac (clusters 8-14). The remaining sites showed cell-specific occupancy such as cluster 15 preferentially occupied by NONO-TFE3 or clusters 16-19 corresponding to sites selectively occupied and/or marked by H3K27ac in a cell line-specific manner.

To consolidate these findings, we performed TFE3 C&T in the HEKT cells described above. As an additional control, these experiments were performed with an independent TFE3 antibody that was used in parallel in a replicate C&T in UOK109 cells. Comparison of the UOK109 data with the 2 TFE3 antibodies revealed high concordance of the strongly bound sites, with only the weak binding sites showing less signal with the second antibody (**Fig. S5**). In the HEK-GFP cells, binding of endogenous TFE3 was seen at only 1058 sites, with 615 at proximal promoters (+/- 500 nt from TSS) associated with genes involved in lysosome and mTOR signalling as previously reported [13] [14] [15] [16]. Interestingly, OxPhos and autophagy [17] [18] were also represented (**Figs. S6A-B**). MEME-ChIP analyses of endogenous TFE3 sites revealed enrichment only of the M/E-box motif (**Fig. S6C**). In contrast, 17-28-fold more sites were seen in HEKT cells expressing native TFE3 or PRCC-TFE3 fusions with the NONO-TFE3 showing the highest number of peaks (**Fig. S6A**). MEME-ChIP revealed enrichment of the M/E-box in the TFE3 and PRCC-TFE3 expressing lines together with the SP1/KLF motif, whereas in the NONO-TFE3 expressing cells, the SP1/KLF and NFY-binding CAAT-box motifs were the

most enriched (**Fig. S6C**). Again, the M/E-box was protected at the peak centre whereas the other motifs were spread across the peaks. While endogenous TFE3 was present at only 615 proximal promoters, the ectopically expressed proteins bound much more promiscuously at promoters, most notably NONO-TFE3 (**Fig. S6A**). Thus, high nuclear accumulation of native TFE3, PRCC-TFE3 and, in particular of NONO-TFE3, in HEKT cells led to promiscuous promoter binding.

The broad genomic binding seemed inconsistent with the limited effects on gene expression seen by RNA-seq upon fusion protein silencing. It was previously reported that MYC can bind promiscuously to active promoters and stimulate mRNA synthesis [19 20]. As such an effect cannot be readily detected by standard RNA-seq, we performed EU incorporation to determine if TFE3 fusion silencing reduced nascent mRNA synthesis. In each cell line, siTFE3 reduced EU incorporation compared to the siCTR and hence reduced levels of mRNA synthesis **(Fig. 3D)**. The reduction was most prominent in the NONO-TFE3-expressing lines in particular an almost 45% reduction in UOK109 compared to the milder (15-20%) reduction in PRCC-TFE3 expressing lines. Together with the C&T results, these data indicated that TFE3 fusion proteins can both activate specific gene expression programs and, via binding to active promoters, more generally stimulate levels of mRNA synthesis.

TFE3 fusions directly regulate OxPhos and ferroptosis genes

To define the core program of TFE3-fusion driven genes, we identified ≈12 534 sites commonly bound in all 4 tRCC cell lines comprising an M/E-box of which 3900 were located in the proximal promoter (+/- 500 nucleotides from TSS, **Dataset S4**). KEGG ontology of the 3313 associated protein coding genes showed enrichment in previously defined TFE3 functions (lysosome, mTOR pathway, autophagy and mitophagy) but also cell cycle and metabolic pathways including Oxphos and ferroptosis (**Fig. 3E)**. Of the 235 OxPhos signature genes

(comprising GSEA-OxPhos and KEGG-OxPhos genes **Dataset S2**), 145 were associated with the presence of TFE3-bound sites comprising an M/E-box within a distance of +/- 30kB of the TSS marked by H3K27ac (**Fig. S6D**). Among the 54 OxPhos genes down-regulated by siTFE3 in all 4 cell lines, 33 displayed a TFE3 bound site, while of the 59 genes up-regulated in the 3 modified HEK293T cells, 44 displayed TFE3 binding (**Fig. S6D)**.

Examples of TFE3 occupancy of M/E-box-containing sites at genes of mitochondrial electron transport chain complexes and the lysosomal V-ATPase in the tRCC lines are shown in **Figs. S7A-C**. Many of these genes were also bound by ectopic native TFE3 and the NONO- and PRCC-TFE3 fusions in HEKT cells along with endogenous TFE3 binding in the GFP line. For example, the COX5A and ATP6V0E1 loci displayed fusion protein binding to the same sites in the tRCC and HEKT cells (**Figs. S7B-C**). Similarly, *PPARGC1A* (PGC1A) a master regulator of mitochondrial biogenesis regulated by MITF in melanoma [21] also displayed multiple TFE3 fusion binding sites at the promoter and putative upstream and downstream regulatory elements in the tRCC lines (**Fig. S7C**). In contrast, no endogenous TFE3 was seen at the PPARGC1A locus, and only a subset of the sites seen in the tRCC lines were bound in HEKT cells expressing ectopic TFE3.

We noted that ferroptosis was represented in the core program and that several critical genes were regulated by the TFE3 fusions (**Fig. 3E**). GPX4 and the ferroptosis modulators GCLC and HMOX1 were upregulated by ectopic fusion protein expression in HEKT cells and both loci showed prominent binding of fusion proteins in the tRCC lines and HEKT cells (**Fig. S8A-B and Dataset S3**). Moreover, the GPX4 locus displayed a broad H3K27ac signal reminiscent of a super-enhancer in all 4 tRCC lines indicating that it may be a critical gene in these cells. In line with this, all 4 tRCC lines showed low IC50 values for the GPX4 inhibitor RSL3 between 2.6 nM for UOK109, and 58 nM for TF1 (**Fig. 3F**). These values were similar to that of A375 melanoma cells known to be ferroptosis sensitive and much lower than for

501Mel melanoma cells (IC50> 2 mM, not shown) reported to be ferroptosis resistant [22]. RSL3-induced cell death was efficiently rescued in the presence of the ferroptosis inhibitor ferristatin-1. These data showed that TFE3 directly regulated key ferroptosis genes in tRCC cells that consequently display marked sensitivity to GPX4 inhibition.

Human tRCC tumours retain elevated OxPhos gene expression during EMT

To investigate whether the elevated TFE3-driven OxPhos in the 4 tRCC cell lines was also seen in human patient samples, we integrated clinical and RNA-seq data of 37 TFE3-tRCC primary samples and 2 metastasis samples (from the same patient, therefore 38 patients and 39 samples) from a cohort of patients from American and French institutes and TCGA whose clinical and pathological features are summarized in **Figs. S9A-B.** This cohort comprised 66% (n=25) females and 34% (n=13) males (ratio 1.92:1 female:male), with a median age of 34 years (range: 3-65), amongst which 8 were under 18. All stages were broadly represented with 53% cases presenting stage I-II and 47% stage III-IV, with at least 10 cases already spread in lymph nodes and 1 case was poly-metastatic with two different metastatic samples collected. *SFPQ* (n=12), *PRCC* (n=12), *ASPSCR1* (n=5) were the most frequent fusion partners. The cohort also comprised the p34 subunit of general transcription factor TFIIH (GTF2H3) as a novel fusion partner of TFE3 (**Fig. S9C**). At the end of follow-up, 14 (36%) patients died with median overall survival of 47 months (range: 8.3-231.1 months).

To evaluate general differences in the gene expression signatures between tRCC and NAT, we normalized the bulk RNA-seq of 18 normal adjacent tissue (NAT) from the TCGA KIRC and the 39 TFE3-tRCC samples. 2856 genes were up-regulated in tumour samples and 2146 in NAT (**Fig. S9D-E**). Ontology analysis of genes up-regulated in tumours revealed enrichment in cell proliferation (regulation of cell activation, ribosome, mitotic cell cycle) and immune pathways (regulation of immune effector process, positive regulation of immune response and cytokine production). In contrast, genes up-regulated in NAT were enriched in

kidney structure and function (inorganic ion transmembrane transport, kidney development) and several metabolic pathways (organic hydroxy compound metabolic process, monocarboxylic acid metabolic process). Hence, transformation was associated with loss of epithelial and kidney identity and function, but gain of proliferation and immune markers.

To refine these analyses, we performed unsupervised clustering of gene expression identifying 3 main clusters, each enriched in specific fusion partners (**Fig. 4A**). Cluster 1 was enriched in PRCC-TFE3 fusions (8/12 cases, p.val=2.4e-3) with 3 of the remaining 4 in cluster 3. Cluster 2 was enriched in SFPQ- or NONO-TFE3 fusions (9/14, p.val=3.7e-4) with 4 of the remaining 5 in cluster 3 (**Fig. 4B**). SFPQ and NONO cases were combined as they belong to the same RNA-binding protein family and are known to form a protein complex[23]. Cluster 3 encompassed all 5 ASPSCR1-TFE3 fusions (enrichment p.val=2.4e-3), but comprised a majority (10 of 16) of non-ASPCR1 fusions. Hallmark GSEA analyses of genes differentially expressed between clusters revealed the mesenchymal characteristics and hot-immune profile of Cluster 3 (**Fig. 4B)**. Conversely, clusters 1-2 were more epithelial with enrichment in OxPhos and other metabolic pathways.

The immune profiles were consolidated by MCP-counter analysis showing clusters 1 and 2 to be mostly composed of cold tumours (22/23), whereas cluster 3 displayed immune infiltration (**Fig. S10A**). Similarly, use of a signature gene set confirmed the EMT gradient between clusters 1-3 (**Fig. S10B**). Combining these analyses further revealed the higher fibroblast signature score of mesenchymal tumours. Correlation with patient survival showed that cluster 3 (HR=5.84, Log-rank p.val=0.034) and fibroblast score (HR=3.46, Log-rank p.val=0.027) were strongly associated with poor survival (**Fig. 4C)**.

We next examined the OxPhos signature score using the 200 genes of the GSEA-OxPhos Hallmark pathway. Clusters 1 and 2 displayed higher scores than Cluster 3 (**Fig. 4D**). To compare the tRCC OxPhos scores to those of ccRCC, we stratified ccRCC tumours based

on our previously reported single cell RNA-seq signatures that defined an EMT gradient associated with poor survival and an OxPhos to glycolysis switch [24]. Based on deconvolution of the TCGA-KIRC collection with the ccRCC tumour cell signatures, we selected 20 tumours of the ccRCC.epi (epithelial), ccRCC.int (intermediate) and ccRCC.mes (mesenchymal) states and scored their OxPhos signatures (**Fig. 4D**). Comparison with the tRCC clusters showed that although ccRCC.epi had a comparable OxPhos score to tRCC clusters 1 and 2, tRCC tumours showed higher OxPhos score than the ccRCC.int and ccRCC.mes. Cluster 3 in particular showed a higher OxPhos score than ccRCC.mes. Thus, while OxPhos score strongly declined upon EMT in ccRCC, it remained elevated during EMT in tRCC consistent with the ability of TFE3 fusions to drive OxPhos gene expression.

EMT in tRCC tumours associates with presence of myCAFs and poor outcome

We investigated whether we could use the ccRCC tumour cell signatures to identify cell populations in tRCC. CibersortX deconvolution of the tRCC cohort showed that the intermediate ccRCC.int signature was present in almost all tumours (**Figs. 5A-B**). In contrast, cluster 3 tumours showed enrichment in ccRCC.mes, whereas ccRCC.epi was associated with clusters 1 and 2. Deconvolution with the ccRCC myofibroblast (my)CAF and antigen-presenting (ap)CAF signatures revealed enrichment of myCAFs in cluster 3 and hence strong correlation with ccRCC.mes, whereas apCAFs were more broadly associated with tumours of each cluster (**Figs. 5A-C**). Both ccRCC.mes (HR 5.06 Log-rank p=0,005) and myCAFs (HR 4.25, Log-rank p=0,005) were strongly associated with poor tRCC outcome (**Fig. 5D**).

We confirmed these findings in an independent tRCC cohort whose composition and clinical characteristics were described (**Fig. S11A** and [25]). Unsupervised clustering defined 3 main clusters, each enriched in specific fusion partners (**Fig. S11B**) with cluster 3 comprising the 9 ASPSCR1-TFE3 fusions and displaying analogous features to cluster 3 of the above-described cohort (**Fig. S11C**). Cluster 3 showed strongest association with poor survival as did

the EMT and fibroblast scores (**Fig. S11D**). Deconvolution with the ccRCC tumour cell signatures again revealed the strong myCAF-mes association and both cell populations strongly associated with poor survival (**Figs. S11E-F**). Furthermore, scoring their OxPhos signatures again revealed that the tRCC EMT cluster 3 retained a higher OxPhos core compared to ccRCC.mes tumours (**Fig. S11G**).

The data from these 2 cohorts showed that tRCC tumours displayed a high OxPhos score that was maintained during EMT. Moreover, tRCC EMT was associated with enrichment in mesenchymal cancer cells and myCAFs that both strongly associated with poor survival.

Lastly, we integrated the loss and gain of function with the C&T data to define a set of genes whose expression was diminished by TFE3-fusion silencing in tRCC cells, induced by ectopic expression in HEKT cells and in human tumours samples and whose promoters comprised a M/E-box containing TFE3 binding site associated with H3K27ac (**Fig. 6A**). This set of 71 genes fulfil the criteria of those directly regulated by TFE3 fusions both in tRCC cells and in human tumours. From these, we selected the 10 genes most expressed in the cluster 3 EMT tumours (**Fig. 6B**) and found that in both patient cohorts, high expression of this signature was strongly associated with poor outcome (**Fig. 6C**). These TFE3 fusion-regulated genes defined a signature for aggressive tRCC, several of which (*TRPV2*, *EEF2K; GRN* and *VPS18*) are novel potential therapeutic targets.

Together, our data suggest that TFE3 tRCC represent a spectrum of tumours ranging from epithelial to mesenchymal, with high expression of oxidative phosphorylation and ferroptosis resistance driven by a TFE3 core program (**Fig. 6D**).

Discussion

TFE3-fusion driven gene expression.

Here, we performed loss and gain of function experiments in cultured cells alongside multi-omics analyses conducted in both tRCC cell lines and patient tumors. Through this integrated and comprehensive approach, we profiled the genomic binding of TFE3 fusion proteins identifying a core set of target genes including those involved in OxPhos and ferroptosis. TFE3-fusions directly regulated OxPhos genes leading to higher functional OxPhos levels and drove ferroptosis resistance, with cells highly sensitive to RSL3-induced ferroptotic cell death. Consequently, tRCC tumours displayed elevated OxPhos scores that persisted during their EMT defining a novel metabolic subtype of renal carcinoma. Furthermore, we uncovered a correlation between EMT, the presence of mesenchymal tRCC cells and myCAFs that emerged as hallmarks of poor prognostic outcome. This association underscores the clinical relevance of EMT and the presence of myCAFs in the overall progression and outcome of tRCC.

An important observation of our study is the broad binding of NONO-TFE3 and PRCC-TFE3 over the genome. Previous studies on a limited set of tRCC lines reported only small numbers of TFE3 fusion binding sites (around 3032 sites in UOK146 and SFPQ-TFE3 xenograft lines [14], 4129 in UOK109 and 1584 in UOK120 [16]) and they provided no information on association with active H3K27ac-marked regulatory elements. These observations, obtained by ChIP-seq using antibodies distinct from those used here, are inconsistent with the abundant nuclear fusion protein expression. The comprehensive Cut&Tag profiling we report here using 2 independent antibodies, identified more than 12 000 M/E-box-containing sites occupied in all 4 tRCC lines, with 3900 located in proximal promoters of genes many of which were associated with functions previously found to be regulated by native TFE3, such a lysosome and mTOR pathway, but also novel pathways pertinent to oncogenic transformation such as cell cycle,

metabolism and ferroptosis. Our observations are complementary to those of Sicinska et al., [26] who showed that ASPRC1-TFE3 fusions bound broadly to active H3K27ac-marked chromatin regions across the genome in alveolar soft-part sarcoma cells. While these authors did not distinguish M/E-box containing elements from E/M-box independent binding, their data together with that reported here underscore the conclusion that promiscuous genome occupancy is a feature of multiple TFE3-fusion proteins in both sarcomas and tRCC.

While endogenous TFE3 bound only a restricted number of sites in HEKT cells, consistent with its lower expression and limited nuclear presence, ectopically expressed native TFE3, and the NONO- and PRCC-TFE3 fusion proteins bound many more sites and promiscuously at promoters. In particular, the NONO-TFE3 fusion displayed broadest binding in both tRCC and HEK293T cells. The 1000 most occupied PRCC-TFE3 sites in HEKT cells and UOK120 and UOK146 tRCC cells were strongly enriched in the E/Mbox motif, indicating that most sites were directly bound. In contrast, analogous analyses of NONO-TFE3 revealed enrichment for the E/Mbox, but also promoter SP1/KLF and NFY motifs, indicating that unlike PRCC-TFE3, NONO-TFE3 bound indirectly at promoter sites with high affinity. While, this may be explained by the much higher nuclear accumulation of NONO-TFE3 in the tRCC lines, this is not the case in the HEKT cells, rather suggesting that increased indirect promoter binding is an intrinsic property of the NONO-TFE3 fusion.

Several mechanisms may account for the ability of the fusions to associate with active promoters. The NONO-SFPQ complex was previously shown to interact with the C-terminal domain (CTD) of the largest subunit of RNA polymerase II (Pol II) and bind to proximal promoters [27] [28] [29]. The NONO-TFE3 fusion protein may therefore also be recruited promiscuously to promoters in an analogous manner via the Pol II CTD. Although less is known concerning the splicing factor PRCC, the TFE3 fusion partners are invariably proteins involved in RNA splicing or metabolism. It is now well established that such factors assemble on the

CTD to coordinate transcription and splicing [30 31 32]. For example, the TFE3 partner ZC3H4 is not involved in splicing, but restriction of non-coding transcription [33 34 35] and was shown to be recruited to active promoters. This idea is further underlined by the observation that TFE3 can be fused with Mediator subunit MED15 or the p34 subunit (GTF2H3) of TFIIH, general transcription factor complexes integral to the Pol II preinitiation complex. It is therefore probable that recruitment to active promoters is a general feature of TFE3 fusions. In addition, it was reported that high NONO-TFE3 accumulation may result from its oligomerization and formation of phase separated condensates [36]. The presence of the fusion partners in nuclear speckles or other phase separated structures coordinating transcription and RNA splicing/metabolism may further contribute to promiscuous TFE3-fusion binding to promoters and other regulatory elements.

Nevertheless, broad promoter binding was also seen with ectopically expressed native TFE3 indicating that while the fusion partner may contribute to this, it is not absolutely necessary. Promoter binding may therefore be mediated either through the presence of the fusion partner and/or by TFE3 tethering via SP1-family or NFY transcription factors whose motifs were abundant at TFE3-bound promoter sites. The ability of native TFE3 to bind promoters when accumulated in the nucleus may be pertinent to tRCC where TFEB is either amplified or fused with the non-coding RNA MALAT1 [8 37 38]. In both situations, native TFEB accumulates in the nucleus suggesting that promiscuous promoter binding may also be a facet of its oncogenic function.

Here we found that mRNA synthesis was diminished upon TFE3 fusion silencing consistent with its broad promoter recruitment. Moreover, consistent with the increased propensity of NONO-TFE3 to bind active promoters, the strongest reductions in EU incorporation were seen in UOK109 and in TF1, despite its poorer transfection efficiency. It was previously reported that high MYC accumulation in cancers led to its promiscuous

promoter recruitment causing transcriptional amplification [19] [20]. Similarly, oncogenic RAS stimulates global mRNA synthesis when used to transform immortalized fibroblasts [39] and RAS-mediated oncogenesis is associated with increased TATA-binding protein expression and increased transcription that promote oncogenic transformation [40] [41]. Together, these data suggest that global stimulation of mRNA synthesis is a feature shared by several oncogenic proteins and that contributes to oncogenic transformation. Hence TFE3 fusions may act in a bipartite manner, activating specific genes via binding to M/E-box containing regulatory elements and more generally stimulating mRNA synthesis via promoter binding.

TFE3 fusion-driven OxPhos and ferroptosis resistance as mechanisms driving tRCC tumorigenesis.

Loss and gain of function experiments revealed that TFE3 fusions directly drive a large set of OxPhos genes. Down-regulation of OxPhos gene expression correlated with reduced maximal and reserve OxPhos capacities following siTFE3 silencing, but we did not observe the reduced basal OxPhos previously reported in UOK109 cells [34, 38]. Ectopic fusion protein expression in HEK293T cells rapidly activated numerous OxPhos genes leading to increased functional OxPhos levels. Therefore, both loss and gain of function experiments converged to show how the NONO- and PRCC-TFE3-fusions directly regulated OxPhos related genes driving high OxPhos levels in tRCC lines. This comprehensive set of genes went beyond the previously reported *PPARGC1A* and *NMRK2* genes [43] [44].

These observations from cell lines were extended to two independent patient tumour cohorts. In line with the ability of TFE3 fusions to drive OxPhos gene expression and as previously suggested [45] [13] [46], tRCC tumours displayed globally elevated OxPhos scores compared to ccRCC. Furthermore, by stratifying ccRCC tumours of different EMT status, we showed that OxPhos score was strongly reduced during EMT in ccRCC, but maintained at

higher levels in tRCC. This function of TFE3 in tumours is reminiscent of its role in metabolism and mitochondrial function in normal tissues [47]. High OxPhos has been associated with an increased propensity for oncogenic transformation and resistance to apoptosis [48 49]. Similarly, maintaining elevated OxPhos may promote tumour growth in the stressful cellular environments encountered in primary tumours and during metastatic dissemination [50]. The TFE3-related transcription factor MITF was shown to drive OxPhos in melanoma through regulation of PPARGC1A, but also the non-coding RNAs SAMMSON and LENOX [21 51 52 53]. High OxPhos in melanoma is associated with resistance to oxidative stress and resistance to MAP kinase inhibition. The fact that numerous pathways converge to regulate OxPhos further highlights the important role of this pathway in tumour development. Together these observations support the idea that TFE3 fusion driven OxPhos may be an essential aspect of the initiation, survival and progression of tRCC tumours and a mechanism shared amongst MiT family-dependent tumours.

Another mechanism that may contribute to tRCC tumorigenesis is ferroptosis resistance. Ferroptosis-related genes were found in the core program of TFE3-fusion proteins, with *GPX4*, *HMOX1*, *GCLC* and *ASCL1* upregulated by TFE3 fusion protein expression in HEKT cells. Notably, *GCLC* and *HMOX1* have been described as NFE2L2 (NRF2) targets [54 55], a pathway reported to be active in tRCC [13]. Nevertheless, TFE3-fusions bind to M/E-box motifs in the *HMOX1* and *GCLC* proximal promoters and upregulated their expression identifying them as direct target genes. Similarly, the proximal promoters of the X_C-transporter system genes *SLC3A2* and *SLC7A11* also displayed TFE3-fusion binding to M/E-box motifs with *SLC7A11* expression upregulated upon fusion protein expression in HEKT cells. Accordingly, tRCC lines showed marked sensitivity to GPX4 inhibition with IC50 values in the nanomolar range similar to renal medullary carcinoma lines that we previously showed were highly sensitive to GPX4 inhibition [56]. Although NFE2L2 can regulate antioxidant and ferroptosis response in different

cellular contexts, in tRCC TFE3 fusions may bypass this requirement to directly regulate key ferroptosis genes such as *GPX4* and *GCLC* promoting ferroptosis resistance and protection from oxidative stress both of which enhance tumour growth. These observations are reminiscent of those reported in melanoma where TFE3 drives GPX4 inhibitor sensitivity in MITF-low melanoma cells [22]. Direct TFE3 fusion regulation of genes involved in OxPhos and ferroptosis therefore constitute two complementary mechanisms contributing to tRCC tumorigenesis. Moreover, the observation of ferroptosis resistance in tRCC underlines this process as a common mechanism for RCC development [57].

EMT and myCAFs, markers of poor outcome common to tRCC and ccRCC.

Analyses of the tRCC cohorts revealed that EMT and presence of myCAFs both associated with poor outcome, characteristics shared with ccRCC [24]. In each tRCC cohort, we stratified tumours that differed in EMT status, hot-immune microenvironment and presence of CAFs. As previously observed [25], ASPSCR1-TFE3 tumours displayed a strong propensity for EMT, but tumours with other fusion partners also underwent EMT. Interestingly, tumours with PRCC or SFPQ fusions displayed either epithelial or mesenchymal signatures highlighting the variability among their phenotypes. EMT of tRCC cancer cells was further consolidated by deconvolution with signatures for epithelial and mesenchymal ccRCC cancer cells. Tumours with EMT signatures were enriched in mesenchymal cancer cells showing that the bulk EMT signature was not only due to presence of fibroblasts or other components of the TME. Similarly, deconvolution revealed that EMT tumours were enriched in myCAFs and that enrichment in both mesenchymal tumour cells and myCAFs were strongly associated with poor survival. Thus, despite the distinct oncogenic mechanisms, both ccRCC and tRCC tumours display EMT and infiltration by myCAFs that constitute hallmarks of poor disease outcome. These results might explain why tRCC patients are highly resistant to immune checkpoint inhibitors [58 59]. Future studies will clarify the similarities and differences in the myCAF across

RCC subtypes and whether this could be used to personalize treatments of patients using immune checkpoint inhibitors.

By defining direct TFE3-fusion target genes strongly expressed in aggressive tRCC, we identified several with therapeutic potential. TRPV2 (Transient Receptor Potential Vanilloid type-2) is a Ca^{2+} permeable ion channel associated with cell survival, proliferation, and metastasis in various cancers [60] [61]. Several activating (Probenecid) and inhibiting (Tranilast) molecules have been characterized and have been shown to have clinical potential for several cancers [62] [63]. Interestingly Tranilast has further been shown to act on CAFs, reducing TGFβ signaling, fibrosis and myCAF numbers, an observation of particular significance given the role of myCAFs in poor outcome of tRCC [64] [65]. EEF2K is an atypical protein kinase that phosphorylates and inactivates EEF2 to modulate translation [66]. EEF2K is phosphorylated and regulated by MTORC1 and hence is regulated by cellular nutrient levels like TFE3 and plays a role in autophagy. EEF2K may play a role in tumour development and several inhibitory molecules are being developed to block its function [67].

VPS18 is an E3 ligase involved in the homotypic fusion and vacuole protein sorting (HOPS) complex and regulates autophagosome-lysosome fusion [68]. VPS18 expression is elevated in several cancers and is strongly up-regulated in response to drug treatment to elicit drug resistance [69]. In contrast, its silencing blocks xenograft tumour growth. Moreover, its inhibition by the lysosomotropic agent RDN, an amino-methylated derivative of naturally occurring bi-sbibenzyl riccardin D [70] that binds the VPS18 RING domain disrupting the VPS18-VPS16 interaction, impairs lysosomal function and lysosomal mediated drug resistance. Consequently, RDN reverses TFE3-driven lysosomal mediated multidrug resistance [69].

Progranulin encoded by the *GRN* gene is a secreted lysosomal chaperone involved in inflammatory and immune response, regulation of growth and invasion of several cancers, and is an inhibitor of ferroptosis [71] [72]. Tumour derived GRN can modulate the microenvironment

and promote immune evasion and in a mouse model of pancreatic cancer, anti-GRN antibody blockade inhibited tumour initiation and progression [73]. Moreover, the GRN-targeting antibody AGO1 is currently in phase 1 clinical trial [74]. Each of the above proteins therefore represent novel potential therapeutic targets for tRCC.

Methods

Cell Culture

UOK109, UOK120, UOK121, UOK112 and UOK146 were generous gift from Dr Martson Linehan (NIH, Bethesda, USA) while TF1 line was a generous gift from Dr Gilles Pagès (Nice, France). These cell lines were grown in DMEM medium supplemented with glucose (4.5g/L), 10% heat-inactivated Foetal Calf Serum (FCS), Non-essential amino acid (AANE) and 1% penicillin/streptomycin. The remaining cell lines were purchased from ATCC. The embryonic kidney HEK293T cell line and the derived lines expressing ectopic TFE3 fusions as well as UOK121, UOK112 and RCC4 cells were grown in DMEM medium supplemented with glucose (1g/L), 10% FCS and 1% penicillin/streptomycin. The primary kidney cell line RPTEC was grown in DMEM/F12 medium supplemented with 2.5mM L-Glutamine, 15mM HEPES, hTERT Immortalized RPTEC growth kit-supplement A&B (ATCC ACS-4007), gentamycin (40µg/mL) and G418 (100µg/mL). 786-O ccRCC cells were grown in RPMI 1640 medium supplemented with glucose (2.5g/L), HEPES (10mM), 10% FCS, sodium pyruvate (1mM) and gentamycin (40µg/mL). A-498 ccRCC cells were grown in MEM medium (LifeTechnologies 31095-029) supplemented with 10% heat-inactivated FCS and gentamycin (40µg/mL). The papillary carcinoma renal cell line ACHN was grown in MEM medium (ATC 30-2003) supplemented with 10% FCS and gentamycin (40µg/mL). All cells were cultured at 37°C, 5% CO_2.

Cell line transfection with siRNA

For siRNA knockdown experiments, 400 000 cells were plated in 6-well plates 18 hours before transfection. Transfection was performed in OPTI-MEM medium using lipofectamine RNAiMAX (13778150 – Invitrogen) with 25nM siRNA (mix of 3 individual siRNA for knockdown of TFE3 – J-009363 -06, -08, -09, or with a pool of 4 non-targeting siRNA as control – D-0018-10 – Dharmacon). After 7 hours of incubation, OPTI-MEM was replaced by complete medium. siRNAs sequences are listed in the Table S1.

Plasmid production, lentiviral particle production, cell line transduction and expression

TFE3, NONO-TFE3 and PRCC-TFE3 cDNAs were synthesized by GenScript and subsequently cloned into the all-in-one doxycycline inducible lentiviral vector pCW57-GFP-P2A-MCS vector (Addgene plasmid #71783) in place of GFP. Lentiviral particles were produced to express independently each of the 3 genes of interest or GFP control by transfection of HEK293T cells with packaging plasmids. Lentiviral particles were purified by filtration and ultracentrifugation and resuspended in PBS. After titration, HEK293T cells were infected at multiplicity of infection of 1 using polybrene and selected by puromycin addition to the media (1µg/mL). After establishment of the stable cell lines, expression of the proteins of interest was induced by adding of Doxycycline (Dox) in the medium (10µg/mL).

RNA extraction and quantitative PCR

Total mRNA isolation was performed using Trizol (TR 118) and isopropanol precipitation. RNA was then treated with DNAseI following the TurboDNAse kit instructions (AM1907 – ThermoFisher) and 1µg of RNA was reversed transcribed using SuperScript IV (18090050 – ThermoFisher) following manufacturer instructions. qPCR (program: 95°C, 15min; (95°C, 10s; 55°C, 15s; 72°C, 10s) x60; (95°C, 5s; 70°C, 1min05; 95°C, 5 min) x1; cool down) was performed using SYBR Green I (Roche) and monitored by a LightCycler 480 (Roche). Target gene expression was normalized using the geometric means of *TBP*, *HMBS*,

GAPDH, *RPLP0* as reference genes. Primers for RT-qPCR were designed using the OligoAnalyzer™Tool from IDT and are listed in Table S2.

Total protein extraction

Whole cell extracts were prepared by 4 freeze-thaw cycles in LSDB 0.5M buffer (0.5M KCl, 25mM Tris pH7.9, 10% glycerol v/v, 0.05% NP-40 v/v, 16mM DTT and protease inhibitor cocktail, 1861280 – Thermo Fisher). After 20 minutes on ice, extracts were centrifuged for 10 minutes at 13 000g, 4°C and supernatant was retrieved.

Differential extraction of nuclear and cytoplasmic proteins

Protein extracts from nuclear and cytoplasmic compartments were prepared using manual lysis and sucrose buffer. Briefly, cells were resuspended in hypotonic buffer (1mL buffer / g of cells – 10mM Tris pH7.65, 1.5mM $MgCl_2$, 10mM KCl), incubated on ice for 5 min and transferred to a loose Dounce. After 15 strokes, sucrose buffer (0.33mL/g cells – 20mM Tris pH7.65, 15mM KCl, 60mM NaCl, 0.34M sucrose) and PMSF (0.2mM) were added to the lysed solution and centrifuged for 10min at 4500g, 4°C. The supernatant was recovered (cytosolic extract) and the pellet resuspended in sucrose buffer (1mL/g cells). Hypertonic buffer (20mM Tris pH7.65, 25% glycerol, 1.5mM $MgCl_2$, 0.2mM EDTA, 900mM NaCl) was added dropwise under slight agitation (0.35mL/g cells), tubes were incubated for 30min at 4°C on a wheel and centrifuged for 10min at 4700g, 4°C. The supernatant formed the soluble nuclear extract.

Immuno-blots

Proteins were quantified using the Bradford protein quantification assay (BioRad) and 20µg of extract were used to perform SDS-polyacrylamide gel electrophoresis. Proteins were transferred to a nitrocellulose membrane. Membranes were incubated overnight at 4°C with primary antibody diluted in PBS + 5% dry-fat milk + 0,01% Tween-20. After washing steps, membranes were incubated with HRP-conjugated secondary antibodies for 1 hour at room

temperature and revealed using the ECL detection system (GE Healthcare). Antibodies and used dilutions are listed in Table S3.

Colony forming analyses by Crystal violet staining

Colony formation was assessed by plating 5 000 cells in 6-well plates, incubating them for 14 days and finally fixing in 4% paraformaldehyde and staining with 0,05% Crystal Violet solution (Sigma Aldrich). Modified HEK293T lines used for Dox-inducible gain of function experiments were plated on poly-D-lysine pre-coated dishes and Dox was renewed every 48 hours.

RNA-sequencing and analysis

Cells were harvested by scraping 72 hours after siRNA transfection and total mRNA was extracted with Trizol as described above. For HEKT cells, RNA was prepared 24 hours after Doxycycline (Dox) addition. Following QC and library preparation, RNA-seq was performed on an Illumina Hiseq4000 as 2*100bp paired end sequencing following Illumina's instructions. After sequencing raw reads were pre-processed in order to remove adapter and low-quality sequences (Phred quality score below 20) using cutadapt version 1.10 and reads shorter than 40 bases were discarded. Reads were mapped to rRNA sequences using bowtie version 2.2.8 and matching reads removed. Remaining reads were then mapped onto the hg19 assembly of Homo sapiens genome using STAR version 2.5.3a. Gene expression quantification was performed from uniquely aligned reads using htseq-count version 0.6.1p1, with annotations from Ensembl version 75 and "union" mode. Only non-ambiguously assigned reads were retained for further analyses. Read counts were normalized across samples with the median-of-ratios method. Comparisons of interest were performed using the Wald test for differential expression and implemented in the Bioconductor package DESeq2 version 1.16.1. Genes with high Cook's distance were filtered out and independent filtering based on the mean of normalized counts was performed. P-values were adjusted for multiple testing using the

Benjamini and Hochberg method. Deregulated genes were defined as genes with log2(fold-change) >1 or <-1 and adjusted p-value <0.05. Gene set enrichment analyses were done with the GSEA software v3.0 using the hallmark gene sets of Molecular Signature Database v6.2. Gene Ontology analysis was performed using the Database for Annotation, Visualization and Integrated Discovery (http://david.abcc.ncifcrf.gov/). Gene list intersections were computed and Venn diagrams represented using the web tool Venny (https://bioinfogp.cnb.csic.es/tools/venny/). Gene set variation analysis (GSEA) was computed using the Bioconductor R package GSVA v1.48.3.

Measure of Oxygen Consumption Rate (OCR) in living cells

OCR in living cells was measured using the MitoStress kit and the XF96 extracellular analyser (Agilent) following manufacturer's instructions. Cells were seeded 48 hours prior to measurement in SeaHorse microplates (Agilent) in order to have confluent cells on the day of measurement: 30 000 cells per well for translocation cell lines, 40 000 cells for Dox-inducible HEK293T lines, 10 000 cells for 786-O and 20 000 cells for UOK121. Dox-inducible HEK293T cells were treated with Dox 24 hours prior to measurement. For siRNA-mediated knockdown, 30 000 UOK109 cells were seeded per well, after 7 hours of incubation with siRNA and 65 hours prior to measurement, whereas 20 000 cells were seeded for TF1, UOK120, UOK146. For each line, 6 wells were considered for technical replicates per experiment and up to 5 biological replicates were performed. The day before measurement, the cartridge was hydrated. One hour before measurement, medium on cells was changed for XF base medium supplemented with 1mM pyruvate, 2mM glutamine and 10mM glucose and cells were incubated at 37°C with no CO_2. After calibration of the cartridge, cells were sequentially exposed to 1µM oligomycin, 1µM carbonyl cyanide-4-(trifluorome-thoxy)-phenylhydrazone (FCCP), and 0.5µM rotenone/antimycin A, with 13 min between each drug. After measurement, cells were fixed with 4% PFA, permeabilized with 0.2% Triton and stained with DAPI (1µg/µL). Number of cells per well was determined by microscopy at the IGBMC High

Throughput Screening Facility and used to normalize obtained measurements per well. Calculation of basal, maximal and reserve capacity were performed as defined in the Agilent User Guide.

ChiP-seq and Cut&TAG

For H3K27ac chromatin immunoprecipitation, cells were grown in 15cm petri dishes and fixed with 0,4% PFA for 10min at room temperature with agitation. Crosslinking was quenched by adding glycine (2M, pH8) for 10min with agitation. Fixed cells were washed once with PBS, harvested by scraping, pelleted and resuspended in lysis buffer (2-times pellet volume – 50mM TrisHCl pH8, 1% SDS, 10mM EDTA, 0,1M PMSF, protease inhibitor cocktail). Lysates were transferred to Covaris tubes and sonicated in a Covaris sonicator to obtain DNA fragments (median size 200-500 bp). After sonication, lysates were centrifuged at 13 000g, 4°C for 10min. Chromatin sonication was verified on agarose gels after decrosslinking (overnight incubation at 65°C with 0.3M NaCl, RNAseA and 1h incubation with proteinase K at 42°C) and phenol-chloroform extraction. 100µg of chromatin were diluted with ChIP dilution buffer (16.7mM Tris HCl pH8, 167nM NaCl, 1.2mM EDTA, 1.1% Triton X-100, 0.01% SDS) and precleared with Protein G DNA magnetic beads (30µL/mL lysate) for 1h, 4°C. After bead removal, chromatin extracts were incubated with anti-H3K27ac (1µg/20µg of chromatin) at 4°C, overnight. To recover antibody-chromatin complexes, Protein G DNA magnetic beads were added to each sample and incubated at 4°C for 1h. Beads were then washed with 4 different buffers, two times each (Low Salt Buffer: 20mM Tris HCL pH8, 150mM NaCl, 2mM EDTA, 1% Triton X-100, 0.1% SDS; High Salt Buffer: 20mM Tris HCL pH8, 500mM NaCl, 2mM EDTA, 1% Triton X-100, 0.1% SDS; LiCl Buffer: 10mM Tris HCl pH8, 1mM EDTA, 1% sodium deoxycholate, 1% NP40, 0.25mM LiCl; TE Buffer: 10mM Tris HCl pH8, 1mM EDTA). Finally, beads were re-suspended twice in DNA elution buffer (0.1M $NaHCO_3$, 1%SDS) and incubated 15min at room temperature. Decrosslinking and phenol-chloroform

extraction were then performed and extracted DNA was resuspended in water. After QC and library preparation, DNA was sequenced on Illumina HiSeq 4000 sequencer as 1*50bp single-read following Illumina's instructions.

For Cut&TAG, cells were plated in 10 cm petri-dish for 48h and harvested using CellStripper solution and washed 2 times in PBS 1X. 500 000 cells were used for Cut&TAG experiments following the manufacturer's instructions (Cut&TAG-IT™ Assay kit – n°53160 – ActiveMotif). Libraries were prepared using the Nextera™-Compatible Multiplex Primers (96-plex – n°53155 – ActiveMotif). Libraries were verified on a BioAnalyzer before 2 rounds of purification using SPRI beads as described in the Cut&TAG kit. Antibodies used are listed in Table S3.

Analysis of ChIP-seq and Cut&TAG

Following sequencing, reads were mapped to the Homo sapiens genome assembly hg19 using Bowtie with the following arguments: -m 1 --strata --best -y -S -l 40 -p 2. Peak detection was performed using the MACS software v1.4.3 for transcription factors and v2.1.1 for chromatin marks. Peaks were annotated with the 'annotatePeaks' command from HOMER v4.11 using the GTF annotation file from ENSEMBL v75 (http://homer.ucsd.edu/homer/ngs/annotation.html). Global clustering analyses with quantitative comparisons and representations of read density heatmaps were performed using seqMINER. Gene ontology analyses from peak annotations were performed with the Bioconductor R package clusterProfiler v4.8.2 and the metascape web tool (https://metascape.org). Motif discovery from ChIP-seq peak sequences was performed using the MEME CHIP and RSAT peak-motifs algorithms (http://rsat.sb-roscoff.fr/peak-motifs_form.cgi). Peak intersections were computed using bedtools v2.26.0 and for multiple samples, the R package DiffBind v3.8.4. Visualization of ChIP-seq signal at specific gene loci of interest was achieved using the UCSC genome browser (https://genome.ucsc.edu/).

RNA-sequencing and analysis of patient tumour samples

RNA sequencing for TFE3 tRCC was available for 39 samples from 38 patients (one patient presented with metastases resulting in two samples) including in house cohort of 26 samples and 13 samples from TCGA KIRC and KIRP cohorts. Identification of fusion transcripts for 17 of those samples have been previously reported [75] [76] and nine samples were novel. Library preparation, sequencing and fusion transcript detection for those novel samples was performed as previously described [76]. After sequencing, raw read counts were adjusted for batch-effect correction using ComBat-seq by providing the sequencing batch as the 'batch' argument. For analyses requiring integration with TCGA-KIRC samples, an additional covariate specifying the type of sample (tRCC, normal adjacent tissue or ccRCC) was provided as the 'group' argument. The adjusted raw-count matrices were normalized by sequencing depth using DESeq2 size-factors and then gene-counts were divided by median transcript lengths.

Consensus clustering of TFE3-fusion samples was performed using the R package ConsensusClusterPlus v3.17 following standard procedure. In short, matrices are filtered in order to keep only coding genes based on their biotype annotation. The 5000 most variable genes were retained with the mad() function and the matrices were median centered with the sweep(), apply() and median() functions. Consensus clustering was done with the ConsensusClusterPlus() function using base parameters. The appropriate numbers of clusters were chosen based on the curve of cumulative distribution function. Differential gene expression analyses between the different groups were done with DESeq2 with functional analyses and visualization performed as described in the previous section for cellular models.

Translocation tumour samples were subjected to deconvolution analysis using the R-package MCP-counter v1.2.0. Then, hierarchical clustering was performed based on the MCP-counter population scores using the hclust() function with "ward.D2" linkage in order to define "hot" and "cold" tumour groups.

Translocation tumour samples were scored according to their mesenchymal/epithelial signature ratios in the following way: epithelial and mesenchymal genes derived from ccRCC samples described in Liang et al. [77]were subjected to unsupervised clustering using tRCC samples then these genes were assigned into mesenchymal and epithelial categories based on their expression patterns in tRCC in order to adapt the original signatures. Some genes such as *VEGFA* and *ANGPTL4* showed different expression patterns in tRCC compared to what was described in ccRCC and were re-assigned to their proper categories while some genes such as *CA9* had no clear expression patterns in tRCC and were removed. Lastly, mesenchymal genes and proximal tubule markers identified from single-cell analyses were added to complete the new mesenchymal and epithelial signatures in order to have the 22 representative genes in each category. The 'SignatureRatio' value was computed as the geometric mean of the mesenchymal signature divided by the geometric mean of the epithelial signature.

Deconvolution of tRCC samples using the ccRCC single cell RNA-seq signatures we previously described was performed using CIBERSORTx algorithm [24]. To stratify epithelial, intermediate and mesenchymal ccRCC samples, tumor samples from the TCGA-KIRC were ordered by the ccRCC.epi/ccRCC.mes ratio of their CIBERSORTx score, then the top 20 were chosen as epithelial, the bottom 20 as mesenchymal and the 20 around the median as intermediate.

Scoring of a considered list of genes to determine patient survival was defined as the geometrical mean of representative genes of the transcriptional signature.

Patient survival was computed in R using packages survival v3.2 and survminer v0.4.9. Patients were stratified using the surv_cutpoint() function and Kaplan–Meier curves were represented with the survfit() and ggsurvplot() functions. Hazard ratios were determined by univariate Cox proportional-hazards models with the "coxph()" function.

Labelling nascent RNA with 5-ethynyluridine (EU)

EU incorporation experiments were conducted using the Click-iT™ RNA Alexa Fluor™ 594 Imaging kit from Invitrogen (C10330), following the supplier's recommendations. Cells were transfected with siCTR or siTFE3 in 6-well plates before replating 25000 cells in 8-well cell culture slides (CCS8 – Mattek) and growing for 24h. Fresh media containing 1 mM of the cell-permeable uridine analog, 5-ethynyluridine (EU) and a final concentration of 50 ng/mL Actinomycin D were added for one hour prior to fixation. Fixation was performed with a 3.7% Formaldehyde solution in PBS for 15 minutes, followed by permeabilization with a 0.5% Triton-X PBS solution for an additional 15 minutes before continuing with the EU labelling in accordance with the protocol's instructions. After one wash step with PBS 1x, cells were fixed with 3.7% formaldehyde for 15min at RT, washed once with PBS and permeabilized with 0.5% Triton X-100 in PBS 1x for 15min and washed again with PBS. Cells were then incubated with Click-IT® reaction cocktail as described in the kit for 30 min at RT protected from light. The cell nuclei were counterstained with Hoechst 33342. After two washes in PBS 1x, microscope slides were mounted using DAPI-ProLongGold (P36935 – ThermoFisher). Images were capture with a SP8-X confocal microscope and the EU fluorescence signal from the nucleus was quantified using the Hoechst signal as a mask for determining the mean fluorescence of the EU signal in each nucleus by analyses using a FlowJo macro program.

Treatment with RSL3 and determination of IC50

10 000 cells were plated in 96-well plate and cultured overnight. Cells were treated with 1, 5, 10, 25, 50, 100, or 250 nM of RSL3 with or without 1uM ferrostatin-1 (SelleckChem, #S7243) and cultured for 48 hours. Each condition was performed in triplicate. Complete medium was replaced by a solution of 1:1 complete medium/PrestoBlue™ (A13261 – Invitrogen) and incubated for 1h at 37°C, 5% CO_2. Viability was then evaluated by measuring the absorbance at 590 nm. Dose response curves and the half maximal inhibitory concentration (IC50) were calculated using GraphPad Prism software.

Statistics.

The statistical tests used in each analysis, mainly paired t-test, 1-way Anova (with Dunett's correction) or Wilcoxon test, are described in the corresponding figure legends with the p-values.

Table S1 – siRNA sequences

siRNA	Sequence (5' -> 3')
siCTR - pool	UGGUUUACAUGUCGACUAA, UGGUUUACAUGUUGUGUGA, UGGUUUACAUGUUUUCUGA, UGGUUUACAUGUUUUCCUA
siTFE3 #6	GGAAUCUGCUUGAUGUGUA
siTFE3 #8	CGCAGGCGAUUCAACAUUA
siTFE3 #9	CAGAGCAGCUGGACAUUGA

Table S2 – Primers sequences – Primers used for RT-qPCR to verify TFE3 expression level in cell lines

Primer name	Sequence (5' -> 3')
HMBS_F	GGCAATGCGGCTGCAA
HMBS_R	GGGTACCCACGCGAATCAC
GAPDH_F	ACAACTTTGGTATCGTGGAAGG
GAPDH_R	GCCATCACGCCACAGTTTC
RPLP0_F	AGTGTTCTGGGCTGGCACA
RPLP0_R	GAAACTGCTGCCTCATATCC
TFE3_F	GCCATCACTGTCAGCAACTC
TFE3_R	GATGAGAGTGCCCAGTTCCT

Table S3 – Used antibodies for Western-blot and Immunofluorescence analysis

Target	Host	Application	Dilution	Reference
TFE3	Rabbit Rabbit	WB; Cut&Tag (tRCC) Cut&Tag (UOK109, HEKT)	1/2500; 1µg 1µg	HPA-023881 (Sigma) Lot. 000009169 Ab93808 (Abcam) Lot 1003520-27
VCL	Mouse	WB	1/1000	V4505 (Sigma Aldrich) Lot 099M4850V
BRG1	Rabbit	WB	1/10 000	Ab110641 (Abcam) Lot : 1000647-1
GAPDH	Rabbit	WB	1/1000	In house
Histone H3	Rabbit	WB	1/1000	Ab1791 (Abcam) Lot 150891-1
H3K27ac	Rabbit	ChIP-seq	10µg	39133 (ActiveMotif) Lot 31521015-13
IgG	Rabbit	Cut&Tag Control	10µg	Ab171870 ((Abcam) Lot 3353004-2

Bibliography

1. Raben, N. & Puertollano, R. TFEB and TFE3: Linking Lysosomes to Cellular Adaptation to Stress. *Annu. Rev. Cell Dev. Biol.* **32**, 255–278 (2016).

2. Bakouny, Z. *et al.* Integrative molecular characterization of sarcomatoid and rhabdoid renal cell carcinoma. *Nat. Commun.* **12**, 808 (2021).

3. Caliò, A., Segala, D., Munari, E., Brunelli, M. & Martignoni, G. MiT Family Translocation Renal Cell Carcinoma: from the Early Descriptions to the Current Knowledge. *Cancers* **11**, (2019).

4. Ellati, R. T. *et al.* Clinicopathologic Features of Translocation Renal Cell Carcinoma. *Clin. Genitourin. Cancer* **15**, 112–116 (2017).

5. Moch, H., Cubilla, A. L., Humphrey, P. A., Reuter, V. E. & Ulbright, T. M. The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs-Part A: Renal, Penile, and Testicular Tumours. *Eur. Urol.* **70**, 93–105 (2016).

6. Geller, J. I. *et al.* Characterization of adolescent and pediatric renal cell carcinoma: A report from the Children's Oncology Group study AREN03B2. *Cancer* **121**, 2457–2464 (2015).

7. Sukov, W. R. *et al.* TFE3 rearrangements in adult renal cell carcinoma: clinical and pathologic features with outcome in a large series of consecutively treated patients. *Am. J. Surg. Pathol.* **36**, 663–670 (2012).

8. Kauffman, E. C. *et al.* Molecular Genetics and Cellular Characteristics of TFE3 and TFEB Translocation Renal Cell Carcinomas. *Nat. Rev. Urol.* **11**, 465–475 (2014).

9. Grépin, R. *et al.* The Relevance of Testing the Efficacy of Anti-Angiogenesis Treatments on Cells Derived from Primary Tumors: A New Method for the Personalized Treatment of Renal Cell Carcinoma. *PLoS ONE* **9**, e89449 (2014).

10. Clark, J. *et al.* Fusion of splicing factor genes PSF and NonO (p54 nrb) to the TFE3 gene in papillary renal cell carcinoma. *Oncogene* **15**, 2233–2239 (1997).

11. Sidhar, S. The t(X;1)(p11.2;q21.2) translocation in papillary renal cell carcinoma fuses a novel gene PRCC to the TFE3 transcription factor gene. *Hum. Mol. Genet.* **5**, 1333–1338 (1996).

12. Suske, G. NF-Y and SP transcription factors - New insights in a long-standing liaison. *Biochim. Biophys. Acta Gene Regul. Mech.* **1860**, 590–597 (2017).

13. Bakouny, Z. *et al.* Integrative clinical and molecular characterization of translocation renal cell carcinoma. *Cell Rep.* **38**, 110190 (2022).

14. Damayanti, N. P. *et al.* Therapeutic targeting of TFE3/IRS-1/PI3K/mTOR axis in translocation renal cell carcinoma. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **24**, 5977–5989 (2018).

15. Perera, R. M., Di Malta, C. & Ballabio, A. MiT/TFE Family of Transcription Factors, Lysosomes, and Cancer. *Annu. Rev. Cancer Biol.* **3**, 203–222 (2019).

16. Yin, X. *et al.* TFE3 fusions escape from controlling of mTOR signaling pathway and accumulate in the nucleus promoting genes expression in Xp11.2 translocation renal cell carcinomas. *J. Exp. Clin. Cancer Res. CR* **38**, 119 (2019).

17. Settembre, C. *et al.* TFEB Links Autophagy to Lysosomal Biogenesis. *Science* **332**, 1429–1433 (2011).

18. Pastore, N. *et al.* Nutrient-sensitive transcription factors TFEB and TFE3 couple autophagy and metabolism to the peripheral clock. *EMBO J.* **38**, (2019).

19. Lin, C. Y. *et al.* Transcriptional Amplification in Tumor Cells with Elevated c-Myc. *Cell* **151**, 56–67 (2012).

20. Nie, Z. *et al.* c-Myc Is a Universal Amplifier of Expressed Genes in Lymphocytes and Embryonic Stem Cells. *Cell* **151**, 68–79 (2012).

21. Vazquez, F. *et al.* PGC1α Expression Defines a Subset of Human Melanoma Tumors with Increased Mitochondrial Capacity and Resistance to Oxidative Stress. *Cancer Cell* **23**, 287–301 (2013).

22. Dias, D., Louphrasitthiphol, P. & Goding, C. R. TFE3 promotes ferroptosis in melanoma. *Pigment Cell Melanoma Res.* **37**, 286–290 (2024).

23. Knott, G. J., Bond, C. S. & Fox, A. H. The DBHS proteins SFPQ, NONO and PSPC1: a multipurpose molecular scaffold. *Nucleic Acids Res.* **44**, 3989–4004 (2016).

24. Davidson, G. *et al.* Mesenchymal-like Tumor Cells and Myofibroblastic Cancer-Associated Fibroblasts Are Associated with Progression and Immunotherapy Response of Clear Cell Renal Cell Carcinoma. *Cancer Res.* **83**, 2952–2969 (2023).

25. Sun, G. *et al.* Integrated exome and RNA sequencing of TFE3-translocation renal cell carcinom. (2021) doi:10.21203/rs.3.rs-296261/v1.

26. Sicinska, E. *et al.* ASPSCR1::TFE3 Drives Alveolar Soft Part Sarcoma by Inducing Targetable Transcriptional Programs. *Cancer Res.* (2024) doi:10.1158/0008-5472.CAN-23-2115.

27. Lim, Y. W., James, D., Huang, J. & Lee, M. The Emerging Role of the RNA-Binding Protein SFPQ in Neuronal Function and Neurodegeneration. *Int. J. Mol. Sci.* **21**, 7151 (2020).

28. Emili, A. *et al.* Splicing and transcription-associated proteins PSF and p54nrb/nonO bind to the RNA polymerase II CTD. *RNA N. Y. N* **8**, 1102–1111 (2002).

29. Kameoka, S., Duque, P. & Konarska, M. M. p54(nrb) associates with the 5' splice site within large transcription/splicing complexes. *EMBO J.* **23**, 1782–1791 (2004).

30. Saldi, T., Cortazar, M. A., Sheridan, R. M. & Bentley, D. L. Coupling of RNA polymerase II transcription elongation with pre-mRNA splicing. *J. Mol. Biol.* **428**, 2623–2635 (2016).

31. Custódio, N. & Carmo-Fonseca, M. Co-transcriptional splicing and the CTD code. *Crit. Rev. Biochem. Mol. Biol.* **51**, 395–411 (2016).

32. Shenasa, H. & Bentley, D. L. Pre-mRNA splicing and its cotranscriptional connections. *Trends Genet. TIG* **39**, 672–685 (2023).

33. Hughes, A. L. *et al.* A CpG island-encoded mechanism protects genes from premature transcription termination. *Nat. Commun.* **14**, 726 (2023).

34. Estell, C. *et al.* A restrictor complex of ZC3H4, WDR82, and ARS2 integrates with PNUTS to control unproductive transcription. *Mol. Cell* **83**, 2222-2239.e5 (2023).

35. Rouvière, J. O. *et al.* ARS2 instructs early transcription termination-coupled RNA decay by recruiting ZC3H4 to nascent transcripts. *Mol. Cell* **83**, 2240-2257.e6 (2023).

36. Wang, B. *et al.* The positive regulation loop between NRF1 and NONO-TFE3 fusion promotes phase separation and aggregation of NONO-TFE3 in NONO-TFE3 tRCC. *Int. J. Biol. Macromol.* **176**, 437–447 (2021).

37. Napolitano, G. & Ballabio, A. TFEB at a glance. *J. Cell Sci.* **129**, 2475–2481 (2016).

38. Argani, P. *et al.* TFEB-amplified Renal Cell Carcinomas: An Aggressive Molecular Subset Demonstrating Variable Melanocytic Marker Expression and Morphologic Heterogeneity. *Am J Surg Pathol* **40**, 1484–1495 (2016).

39. Kotsantis, P. *et al.* Increased global transcription activity as a mechanism of replication stress in cancer. *Nat. Commun.* **7**, 13087 (2016).

40. Johnson, S. A. *et al.* Increased expression of TATA-binding protein, the central transcription factor, can contribute to oncogenesis. *Mol Cell Biol* **23**, 3043–51. (2003).

41. Johnson, S. A., Mandavia, N., Wang, H. D. & Johnson, D. L. Transcriptional regulation of the TATA-binding protein by Ras cellular signaling. *Mol Cell Biol* **20**, 5000–9 (2000).

42. Wang, B. *et al.* PRCC-TFE3 fusion-mediated PRKN/parkin-dependent mitophagy promotes cell survival and proliferation in PRCC-TFE3 translocation renal cell carcinoma. *Autophagy* **0**, 1–19

(2020).

43. Chen, Y. *et al.* Up-regulation of NMRK2 mediated by TFE3 fusions is the key for energy metabolism adaption of Xp11.2 translocation renal cell carcinoma. *Cancer Lett.* **538**, 215689 (2022).

44. SALMA, N., SONG, J. S., ARANY, Z. & FISHER, D. E. Transcription Factor Tfe3 Directly Regulates Pgc-1alpha in Muscle. *J. Cell. Physiol.* **230**, 2330–2336 (2015).

45. Qu, Y. *et al.* Proteogenomic characterization of MiT family translocation renal cell carcinoma. *Nat. Commun.* **13**, 7494 (2022).

46. TCGA, N. COMPREHENSIVE MOLECULAR CHARACTERIZATION OF CLEAR CELL RENAL CELL CARCINOMA. *Nature* **499**, 43–49 (2013).

47. Pastore, N. *et al.* TFE3 regulates whole-body energy metabolism in cooperation with TFEB. *EMBO Mol. Med.* **9**, 605–621 (2017).

48. Bonnay, F. *et al.* Oxidative Metabolism Drives Immortalization of Neural Stem Cells during Tumorigenesis. *Cell* **182**, 1490-1507.e19 (2020).

49. Spurlock, B. *et al.* Fine-tuned repression of Drp1-driven mitochondrial fission primes a 'stem/progenitor-like state' to support neoplastic transformation. *eLife* **10**, e68394 (2021).

50. Bergers, G. & Fendt, S.-M. The metabolism of cancer cells during metastasis. *Nat. Rev. Cancer* **21**, 162–180 (2021).

51. Haq, R. *et al.* Oncogenic BRAF regulates oxidative metabolism via PGC1alpha and MITF. *Cancer Cell* **23**, 302–15 (2013).

52. Leucci, E. *et al.* Melanoma addiction to the long non-coding RNA SAMMSON. *Nature* **531**, 518–22 (2016).

53. Gambi, G. *et al.* The lncRNA LENOX interacts with RAP2C to regulate metabolism and promote resistance to MAPK inhibition in melanoma. *Cancer Res.* CAN-22-0959 (2022) doi:10.1158/0008-5472.CAN-22-0959.

54. Yan, R. *et al.* NRF2, a Superstar of Ferroptosis. *Antioxid. Basel Switz.* **12**, 1739 (2023).

55. Yang, Z. *et al.* ACTL6A protects gastric cancer cells against ferroptosis through induction of glutathione synthesis. *Nat. Commun.* **14**, 4193 (2023).

56. Vokshi, B. H. *et al.* SMARCB1 regulates a TFCP2L1-MYC transcriptional switch promoting renal medullary carcinoma transformation and ferroptosis resistance. *Nat. Commun.* **14**, 3034 (2023).

57. Yang, L., Fan, Y. & Zhang, Q. Targeting ferroptosis in renal cell carcinoma: Potential mechanisms and novel therapeutics. *Heliyon* **9**, e18504 (2023).

58. Alhalabi, O. *et al.* Immune Checkpoint Therapy Combinations in Adult Advanced MiT Family Translocation Renal Cell Carcinomas. *The Oncologist* **28**, 433–439 (2023).

59. Boilève, A. *et al.* Immune checkpoint inhibitors in MITF family translocation renal cell carcinomas and genetic correlates of exceptional responders. *J. Immunother. Cancer* **6**, 159 (2018).

60. Siveen, K. S. *et al.* TRPV2: A Cancer Biomarker and Potential Therapeutic Target. *Dis. Markers* **2020**, 8892312 (2020).

61. Pumroy, R. A. *et al.* Structural insights into TRPV2 activation by small molecules. *Nat. Commun.* **13**, 2334 (2022).

62. Osman, S. *et al.* Anti-cancer effects of Tranilast: An update. *Biomed. Pharmacother. Biomedecine Pharmacother.* **141**, 111844 (2021).

63. Fury, M. G. *et al.* A phase I clinical pharmacologic study of pralatrexate in combination with probenecid in adults with advanced solid tumors. *Cancer Chemother. Pharmacol.* **57**, 671–677 (2006).

64. Ochi, K. *et al.* Drug repositioning of tranilast to sensitize a cancer therapy by targeting cancer-associated fibroblast. *Cancer Sci.* **113**, 3428–3436 (2022).

65. Panagi, M. *et al.* Polymeric micelles effectively reprogram the tumor microenvironment to potentiate nano-immunotherapy in mouse breast cancer models. *Nat. Commun.* **13**, 7165 (2022).

66. Wang, X., Xie, J. & Proud, C. G. Eukaryotic Elongation Factor 2 Kinase (eEF2K) in Cancer. *Cancers* **9**, 162 (2017).

67. Klupt, K. A. & Jia, Z. eEF2K Inhibitor Design: The Progression of Exemplary Structure-Based Drug Design. *Mol. Basel Switz.* **28**, 1095 (2023).

68. Zhang, S., Li, L., Liu, X. & Zhong, Q. The hookup model of the HOPS complex in autophagosome-lysosome fusion. *Autophagy* 1–2 (2023) doi:10.1080/15548627.2023.2291938.

69. Niu, H. *et al.* Targeting of VPS18 by the lysosomotropic agent RDN reverses TFE3-mediated drug resistance. *Signal Transduct. Target. Ther.* **6**, 224 (2021).

70. Wang, Y. *et al.* A novel derivative of riccardin D induces cell death through lysosomal rupture in vitro and inhibits tumor growth in vivo. *Cancer Lett.* **329**, 207–216 (2013).

71. Ventura, E. *et al.* Progranulin Oncogenic Network in Solid Tumors. *Cancers* **15**, 1706 (2023).

72. Chen, T. *et al.* Progranulin released from microglial lysosomes reduces neuronal ferroptosis after cerebral ischemia in mice. *J. Cereb. Blood Flow Metab. Off. J. Int. Soc. Cereb. Blood Flow Metab.* **43**, 505–517 (2023).

73. Cheung, P. F. *et al.* Progranulin mediates immune evasion of pancreatic ductal adenocarcinoma through regulation of MHCI expression. *Nat. Commun.* **13**, 156 (2022).

74. Tkaczuk, K. R. *et al.* Abstract CT240: On-going phase 1A clinical trial for AG01 an first-in-class anti-progranulin (GP88) monoclonal antibody in patients with advanced malignancies. *Cancer Res.* **83**, CT240–CT240 (2023).

75. Classe, M. *et al.* Incidence, clinicopathological features and fusion transcript landscape of translocation renal cell carcinomas. *Histopathology* **70**, 1089–1097 (2017).

76. Malouf, G. G. *et al.* Next-generation sequencing of translocation renal cell carcinoma reveals novel RNA splicing partners and frequent mutations of chromatin-remodeling genes. *Clin Cancer Res* **20**, 4129–40 (2014).

77. Liang, J. *et al.* The Correlation Between the Immune and Epithelial-Mesenchymal Transition Signatures Suggests Potential Therapeutic Targets and Prognosis Prediction Approaches in Kidney Cancer. *Sci. Rep.* **8**, 6570 (2018).

Article 3: TFE3 fusion proteins drive oxidative metabolism, ferroptosis resistance and general RNA synthesis in translocation renal cell carcinoma

Article 3: TFE3 fusion proteins drive oxidative metabolism, ferroptosis resistance and general RNA synthesis in translocation renal cell carcinoma

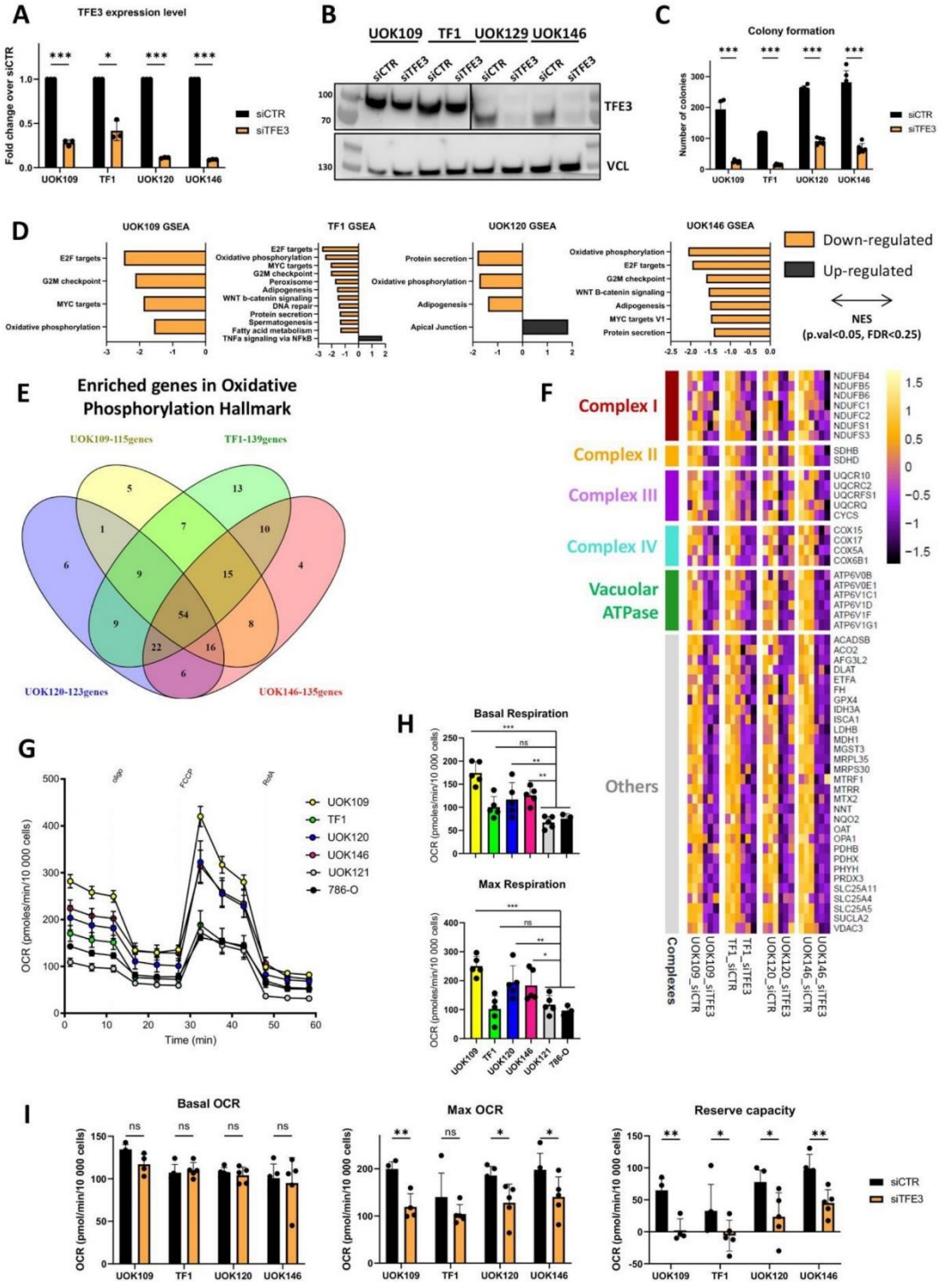
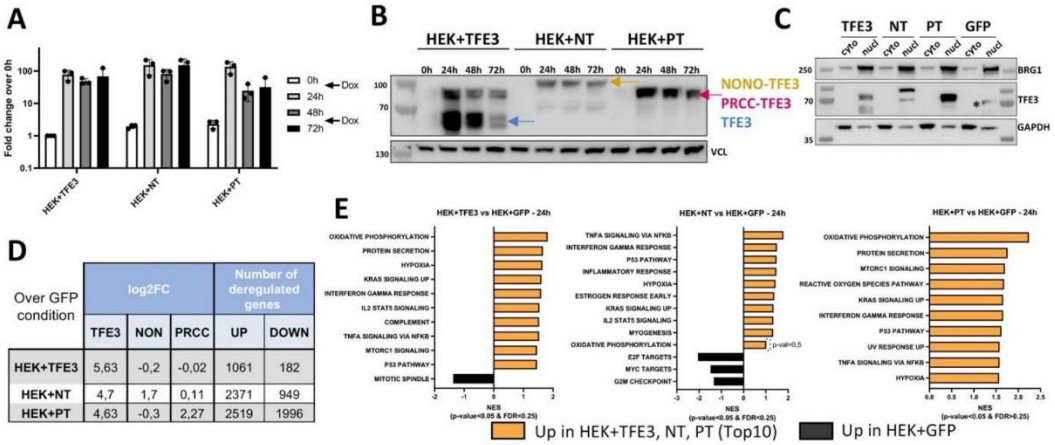


Figure 1

Figure 1. TFE3 fusion proteins regulate colony forming capacity and OxPhos gene expression. A. TFE3-fusion protein expression was measured by RT-qPCR 72 hours after siRNA transfection and expressed as fold change over siCTR normalized to housekeeping gene RPLP0 (n=3). Conditions were compared by paired t-test (*: p.val<0.05, ***: p.val<0.001). **B.** Immunoblot of total protein extracts after siCTR or siTFE3 treatment. TFE3 blots for UOK109/TF1 and UOK120/UOK146 lines were exposed separately for 1 second and 10 seconds respectively. VCL was used as a loading control. **C.** Cells were treated with siCTR or siTFE3 and 5000 cells seeded and cultured for 14 days, fixed and colored with crystal violet to assess colony forming ability (n=6). Conditions were compared by paired t-test (***: p.val<0.001). **D.** Gene Set Enrichment Analysis (GSEA) results comparing siCTR with siTFE3 using the Hallmark gene sets. Significantly deregulated pathways (p.val<0.05, FDR<0.25) are represented for each of the 4 lines. **E.** Venn diagram showing overlap between significantly deregulated OxPhos genes in each cell line as defined by GSEA. **F.** Heatmap showing expression of the 54-commonly down-regulated OxPhos genes as a Z-score. **G.** Measure of OCR in tRCC (UOK109, TF1, UOK120, UOK146) and ccRCC (UOK121, 786-O) cell lines (n=5). OCR values measured in tRCC and ccRCC lines were normalized by the number of cells in each well. **H.** OCR values were used to measure the basal and the maximal respiration levels and compared these levels between each tRCC and the two ccRCC lines by 1-way Anova (Dunnett's multiple comparison test, ns: non-significant – p.val>0.05, *: p.val<0.05, **: p.val<0.01, ***: p.val<0.001). **I.** Measurement of OCR in tRCC lines 72 hours after treatment with siCTR or siTFE3 (n=5). OCR values were used to measure the basal and the maximal respiration levels and reserve respiration capacity. siCTR and siTFE3 conditions were compared by paired t-test (ns: non-significant p.val>0.05, *: p.val<0.05, **: p.val<0.01). For Max OCR: p.val UOK109 = 0.0047; TF1 = 0.19; UOK120 = 0.027; U0K146 = 0.048

Article 3: TFE3 fusion proteins drive oxidative metabolism, ferroptosis resistance and general RNA synthesis in translocation renal cell carcinoma



Over GFP condition	log2FC			Number of deregulated genes	
	TFE3	NON	PRCC	UP	DOWN
HEK+TFE3	5,63	-0,2	-0,02	1061	182
HEK+NT	4,7	1,7	0,11	2371	949
HEK+PT	4,63	-0,3	2,27	2519	1996

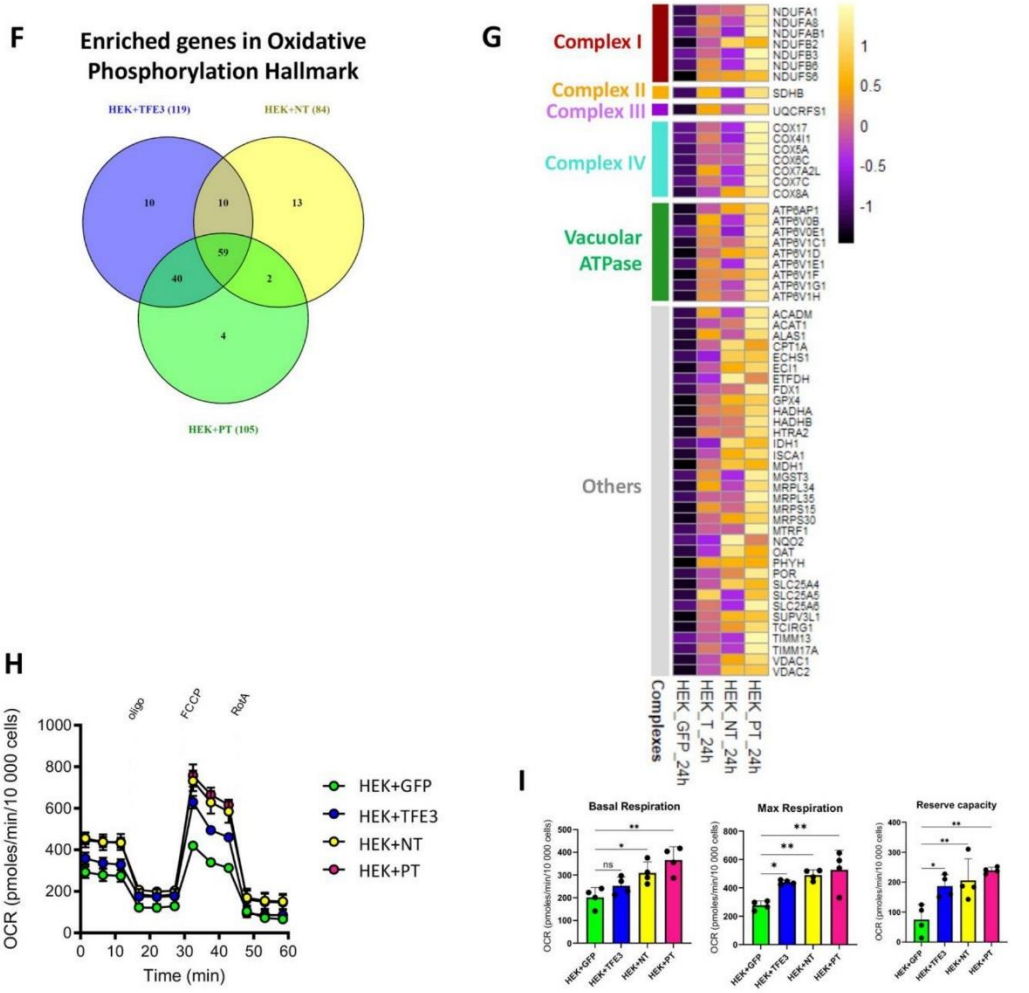


Figure 2

Figure 2. Enhanced OxPhos following ectopic expression of TFE3 fusions in HEK293T cells. A. Fold change of TFE3 expression compared to no-Dox treatment was measured by RT-qPCR at the indicated times after Dox addition (n=3). **B.** Immunoblots of total protein extracts after the indicated time of culture in presence of Dox. VCL was used as loading control. Blue arrow indicates native TFE3 protein, orange arrow NONO-TFE3 (NT) and pink arrow PRCC-TFE3 (PT). **C.** Western blot analysis of cytoplasmic (cyto) and nuclear (nucl) protein extracts of indicated HEK293T cell lines. BRG1 was used as positive control for the nuclear extract, GAPDH for cytoplasmic extract. * Indicates a non-specific signal in the GFP control nuclear extract. **D.** Log2FC values (over GFP) of TFE3, NONO and PRCC in each line and number of genes deregulated 24 hours after Dox addition (log2FC > ±1; adj p.val<0.05 over GFP) in each line (n=3). **E.** GSEA ontology analysis of differentially regulated genes in each cell line over GFP-line control. Significantly deregulated pathways (p.val<0.05, FDR<0.25) are represented for each of the 3 lines, restricted to the top10 pathways for HEK+PT line. **F**. Venn diagram showing overlap between significantly up-regulated OxPhos genes according to GSEA. **G.** Heatmap showing relative expression of each of the 59-commonly up-regulated OxPhos genes as a Z-score. **H.** Measurement of the OCR values following ectopic TFE3, NT or PT expression (n=4). OCR values measured in each HEK line were normalized by the number of cells in each well and used to measure the basal and the maximal respiration levels and the reserve respiration capacity. Experimental conditions were compared over the GFP-condition by 1-way Anova (Dunnett's multiple comparison test, ns: non-significant p.val>0.05, *: p.val<0.05, **: p.val<0.01).

Article 3: TFE3 fusion proteins drive oxidative metabolism, ferroptosis resistance and general RNA synthesis in translocation renal cell carcinoma

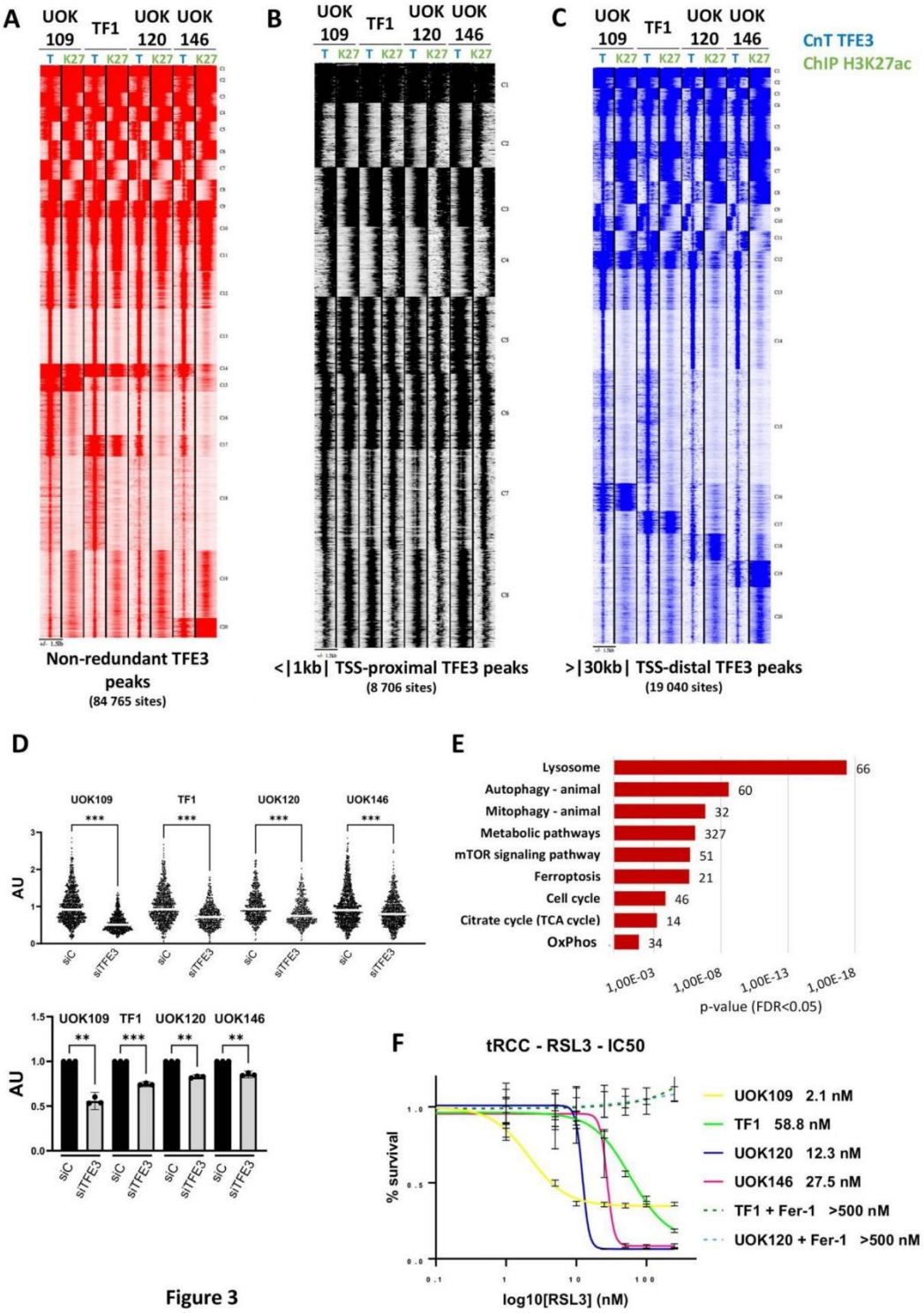


Figure 3

Figure 3. Profiling genome occupancy of TFE3 fusion proteins. A. Read density maps for TFE3 fusion protein occupancy (T) and H3K27ac (K27) at the 84 765 non-redundant sites. **B-C**. Read density maps for the non-redundant TFE3 fusion protein and H3K27ac sites at proximal promoters (<1 kb from the TSS, B) or putative distal enhancers (>30 Kb from the TSS, C). **D**. Level of EU-incorporation in each tRCC line 48 hours after siCTR or siTFE3. The upper panel shows the combined quantification of the relative EU intensity signal in the nucleus for each cell line from the 3 independent experiments normalized by the mean intensity of the control condition. Student T-test: pvalue 0.033 (*), 0,002 (**), < 0,001 (***). The lower panel shows results for the 3 independent experiments with errors bar representing 95% confidence interval. Paired Student T-test pvalue 0.033 (*), 0,002 (**), < 0,001 (***). **E.** KEGG ontology of genes associated with TFE3 fusion protein occupancy of M/E-box containing sites +/- 500 bp of the TSS. Each indicated KEGG pathway presented an FDR < 0.05 and number of genes found in each pathway were noted to the right of the bar. **F.** RSL3 IC50 values of tRCC lines in presence or absence of 1 mM ferristatin-1 (Fer-1) as indicated.

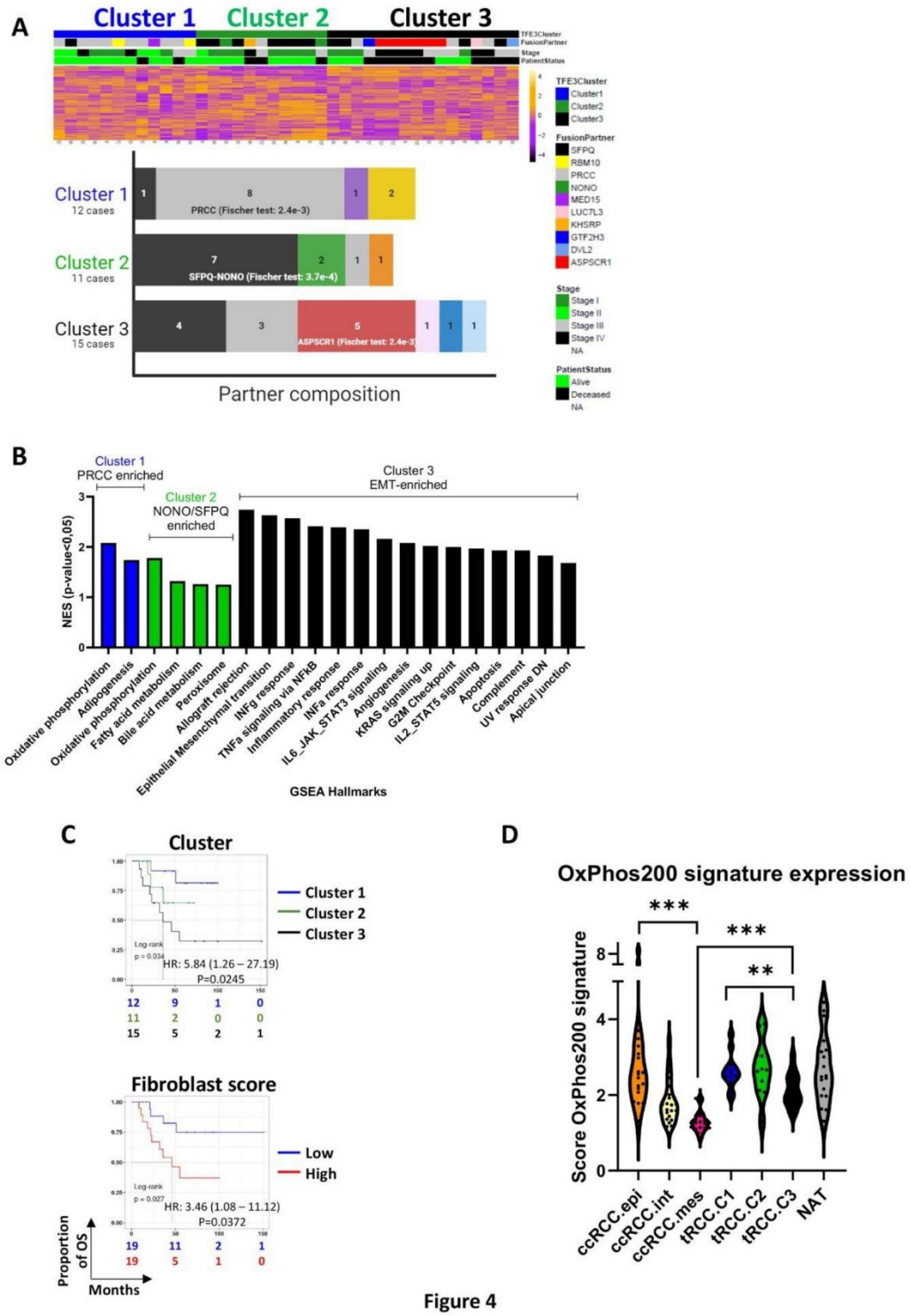


Figure 4

Figure 4. EMT and OxPhos in human tRCC. A. Heatmap of the unsupervised clustering of gene expression with the indicated clinical and molecular parameters, showing the distribution

of tRCC in 3 clusters. Distribution of tRCC samples according to fusion partner type within the 3 major clusters; p-values were indicated for most frequently represented partner in each cluster. **B.** GSEA ontology analyses of differentially expressed genes between the different clusters indicating the enriched terms (p.value<0.05, FDR<0.25) for each cluster (with only the top15 for Cluster 3). **C**. Kaplan-Meier curves for overall survival in tRCC patients according to classification in the three clusters or fibroblast score, calculated using MCP counter, according to the optimal cut point method with the associated log-rank p-value. Hazard ratio (HR) was calculated using univariate Cox proportional-hazards model. **D.** Expression score of the 200-signature genes of GSEA-OxPhos Hallmark in NAT, each tRCC cluster and the indicated ccRCC tumour subtypes from the TCGA KIRC collection. Overall scores were compared by Wilcoxon test (**: p.val<0.01, ***: p.val<0.001).

Article 3: TFE3 fusion proteins drive oxidative metabolism, ferroptosis resistance and general RNA synthesis in translocation renal cell carcinoma

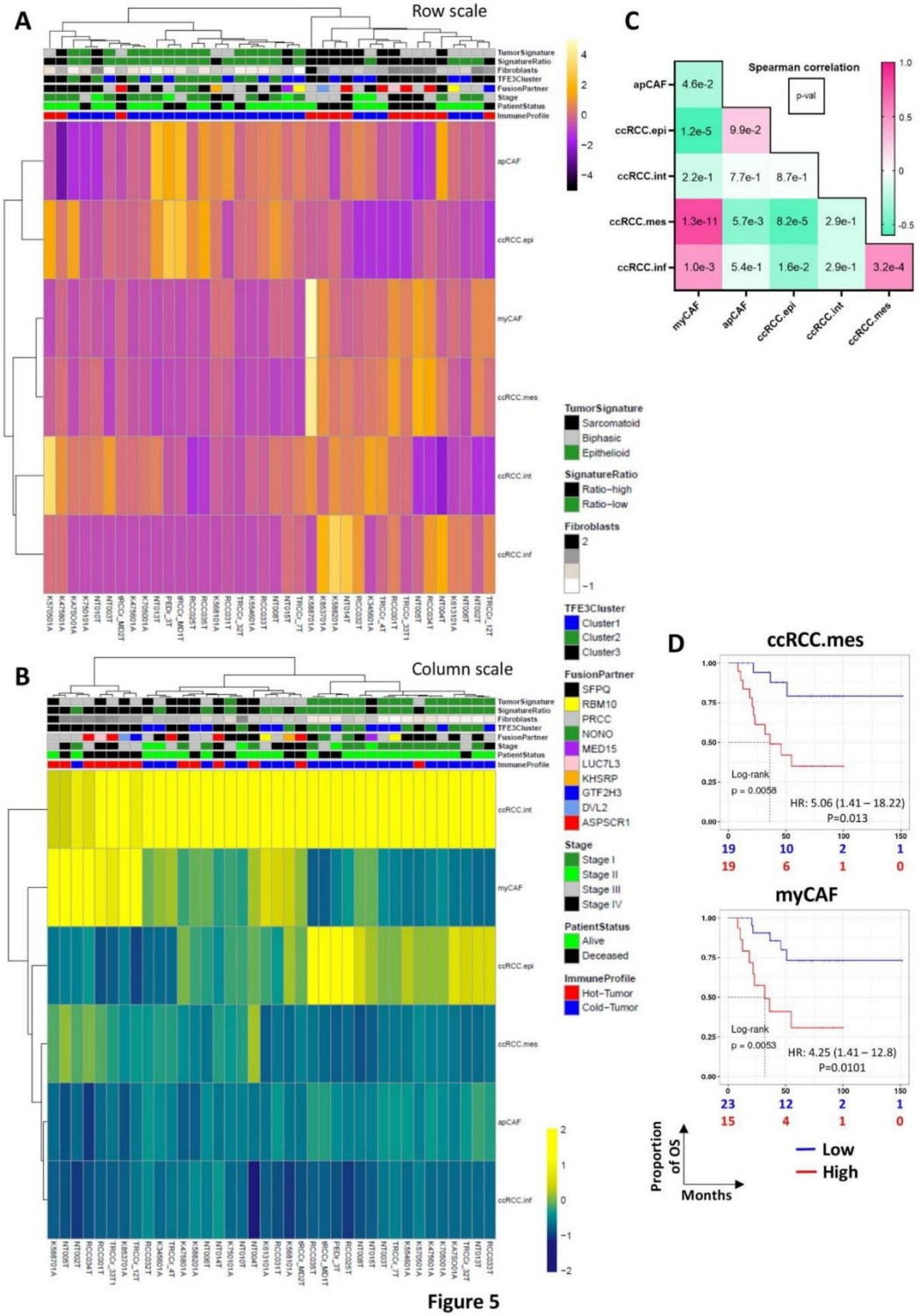


Figure 5

Figure 5. Deconvolution of tRCC tumours using ccRCC tumour cell and CAF signatures. **A-B.** Heatmaps showing deconvolution using the ccRCC tumour cell and CAF signatures inferred by CIBERSORTx and displayed as row-scaled (A) or column-scaled (B) absolute scores on bulk RNA-seq data from the tRCC tumour samples. **C.** Spearman correlation coefficient (coloured box) and associated p-value (number in box) between the indicated populations. **D.** Kaplan-Meier curves for overall survival according to the ccRCC.mes and myCAF proportions using optimal cut-off value method with the associated log-rank p-value and the HR from the univariate Cox proportional-hazards model.

Article 3: TFE3 fusion proteins drive oxidative metabolism, ferroptosis resistance and general RNA synthesis in translocation renal cell carcinoma

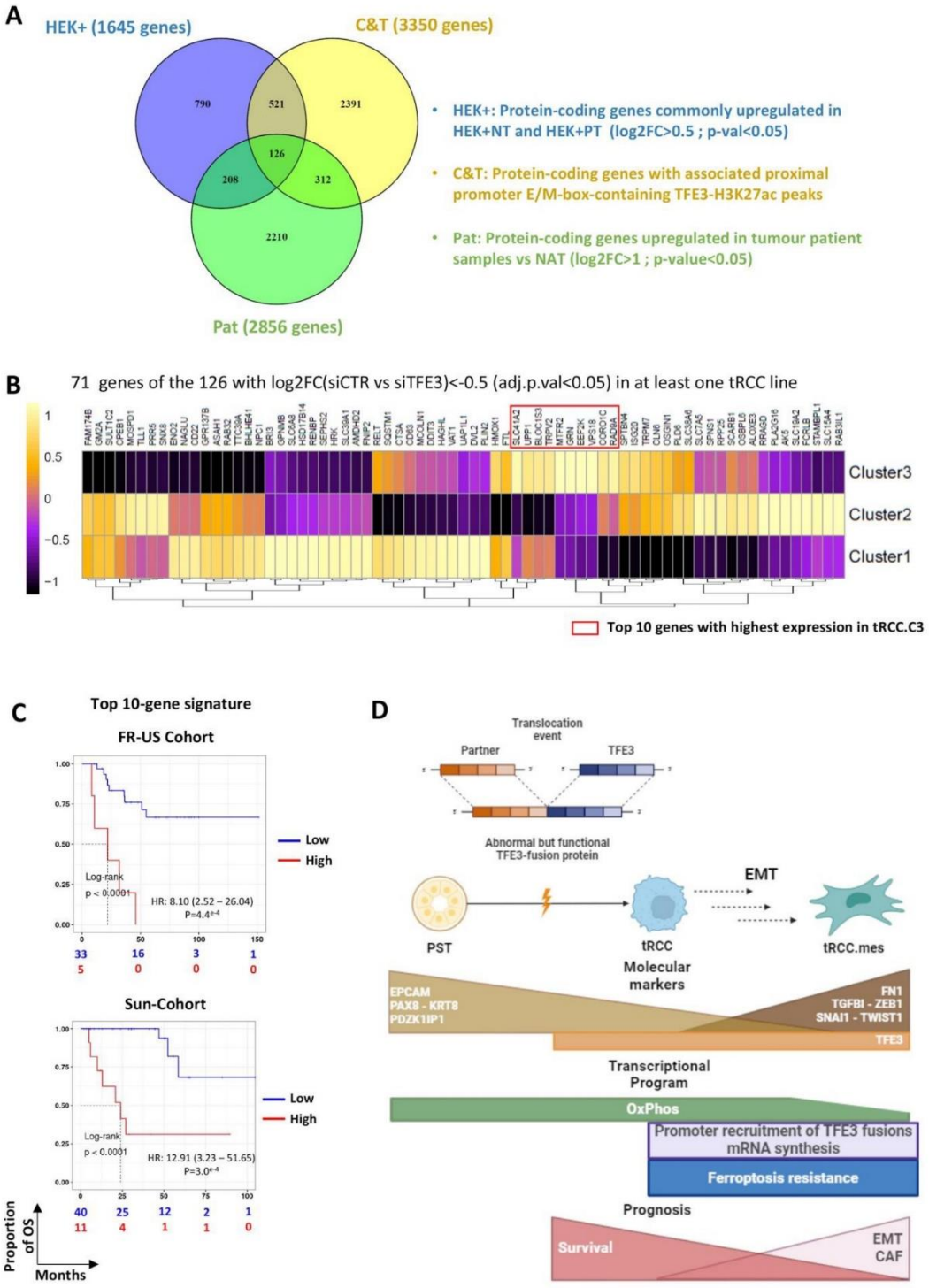
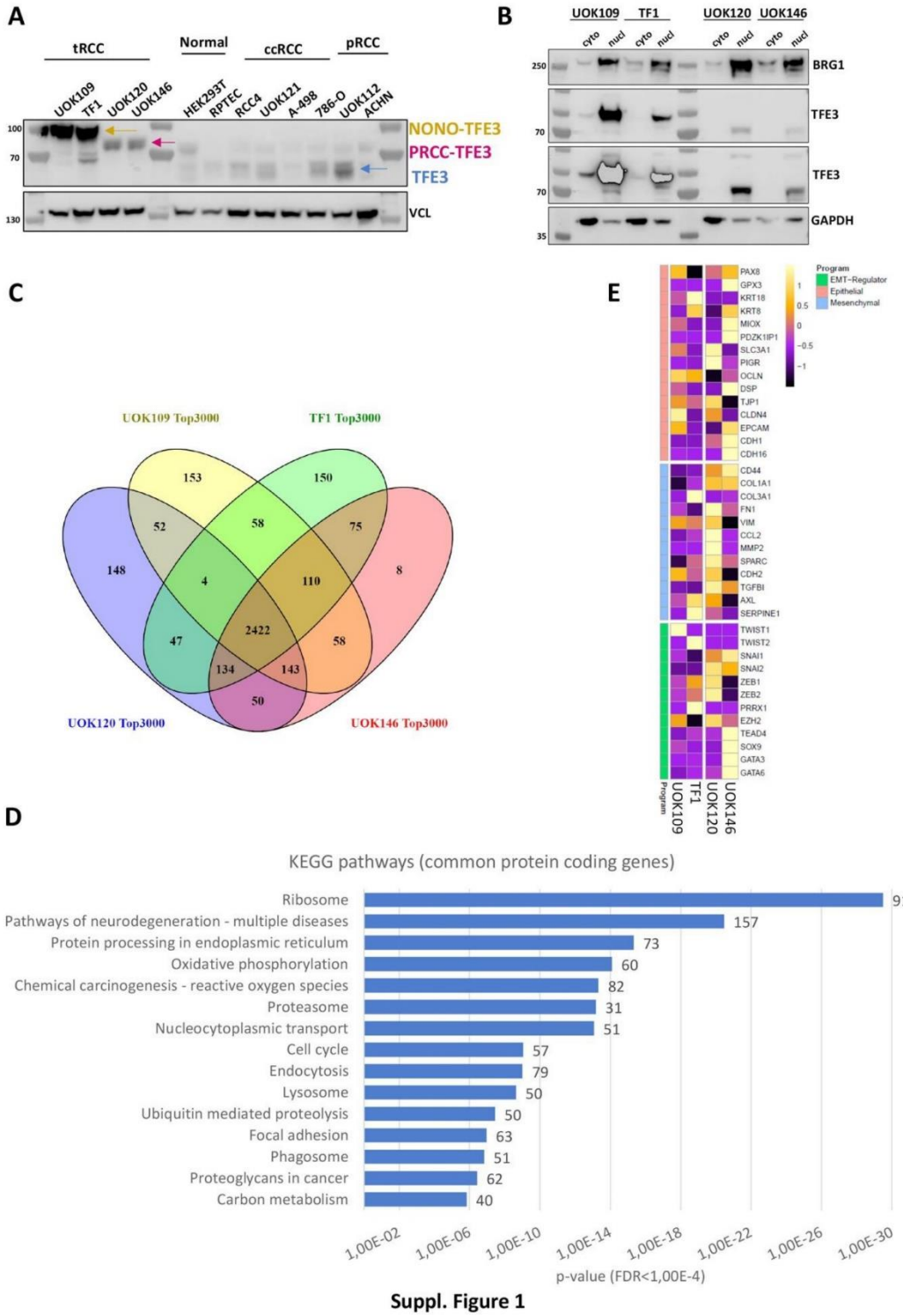


Figure 6

Figure 6. Association of gene expression signatures with patient outcome. A. Venn diagram showing the intersection of: 1645 protein-coding genes up-regulated in HEK293T lines with ectopic expression of either NONO-TFE3 or PRCC-TFE3 (log2FC>0.5, adj.pval<0.05), 3350 protein-coding genes associated with concomitant M/E-box-containing TFE3 and H3K27ac peaks in each tRCC line, and 2856 protein-coding genes up-regulated in tumor samples compared to NAT (log2FC>1, adj.pval<0.05). **B.** The 126 common genes defined in panel A were further filtered to identify those down-regulated following siTFE3 (log2FC<-0.5, adj.p.val<0.05) in at least one of the 4 tRCC cell lines to generate a final collection of 71 genes. Heatmaps of the expression of these 71 genes in each cluster of the FR-US tRCC patient cohort. The red square represents the top 10-genes with strongest expression in tRCC-Cluster 3. **C.** Kaplan-Meier curve for overall survival in patients according to the expression score calculated for the 10 gene signature using the optimal cut-point method with the associated log-rank p-value. HR was calculated using the univariate Cox proportional-hazard model. **D.** Schematic model based on the analyses in this study showing Proximal Straight Tubule (PST) cell transformation in tRCC after the translocation event involving TFE3, the EMT and its associated markers. Upon transformation, tRCC acquire ferroptosis resistance and maintain elevated OxPhos that tends to decrease with the EMT. High nuclear accumulation of TFE3 fusions leads to their promiscuous promoter recruitment stimulating of mRNA synthesis. Finally, EMT and the presence of myCAFs correlate with poor outcome.

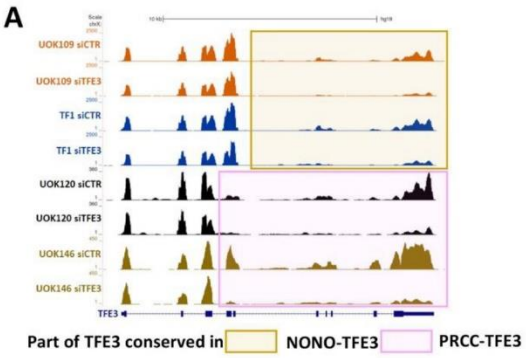
Article 3: TFE3 fusion proteins drive oxidative metabolism, ferroptosis resistance and general RNA synthesis in translocation renal cell carcinoma



Suppl. Figure 1

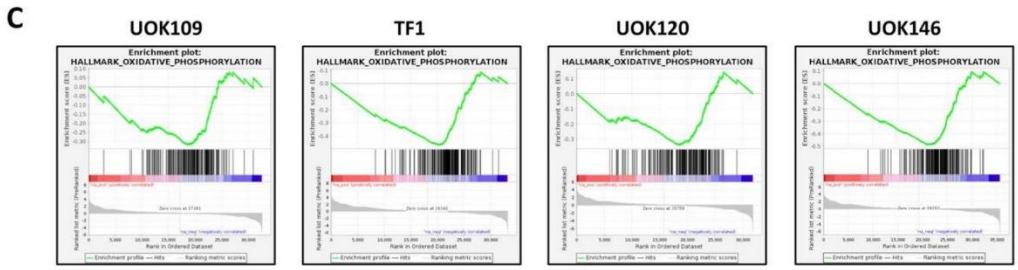
Figure S1. Expression of TFE3 and its fusion proteins in cell lines. A. Western blot analysis of total protein extracts from the indicated cell lines: tRCC (UOK109, TF1, UOK120, UOK146), 'normal' kidney (HEK293T, RPTEC), ccRCC (RCC4, UOK121, A-498, 786-O) and papillary RCC (UOK112, ACHN). Vinculin (VCL) was used as loading control. Blue arrow indicates native TFE3 protein, orange arrow NONO-TFE3 and pink arrow PRCC-TFE3 fusion protein. **B.** Western blot analysis of cytoplasmic (cyto) and nuclear (nucl) fractions from the tRCC cell lines. BRG1 was used as positive control for the nuclear extract and GAPDH for cytoplasmic extract. The same blot was exposed 1 second for NONO-TFE3 and 30 seconds for PRCC-TFE3. **C.** Venn diagram indicating overlap between the 3000 most expressed protein-coding genes in each tRCC line. **D.** DAVID gene ontology analysis of the common 2422 highest expressed protein-coding genes using KEGG. Each indicated KEGG pathway presented an FDR < 10^{-4} and number of genes found in each pathway were noted to the right of the bar. **E.** Heatmap showing the expression of selected epithelial or mesenchymal marker genes in each line.

Article 3: TFE3 fusion proteins drive oxidative metabolism, ferroptosis resistance and general RNA synthesis in translocation renal cell carcinoma



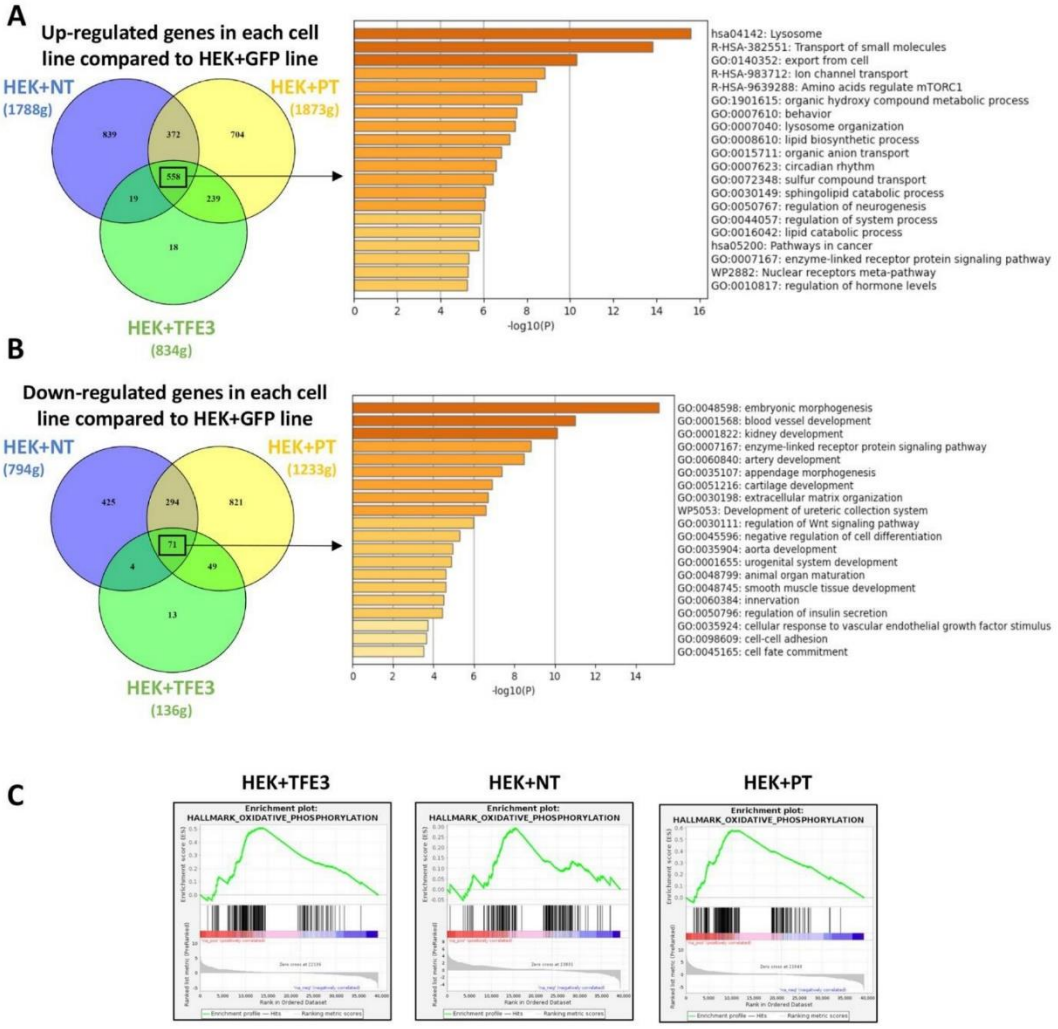
B

	Read count at fusion junction		Number of deregulated genes	
	siCTR	siTFE3	Up	Down
UOK109	72	22	186	299
TF1	119	35	598	400
UOK120	14	0	277	329
UOK146	151	11	288	356



Suppl. Figure 2

Figure S2. Genes de-regulated by siTFE3 silencing. **A.** UCSC view of the RNA-seq reads at the *TFE3* gene locus. Orange box represents the TFE3 region present in NONO-TFE3 fusion, while pink box encompasses the region conserved in PRCC-TFE3 fusion. **B.** Table indicating the read count at junction between TFE3 and the fusion partners and the number of deregulated genes in siTFE3 samples compared to siCTR (log2FC > ±1; adj.p.val<0.05). **C.** Enrichment plots of the Oxidative Phosphorylation (OxPhos) pathway after GSEA analysis between siCTR and siTFE3 conditions in each line.

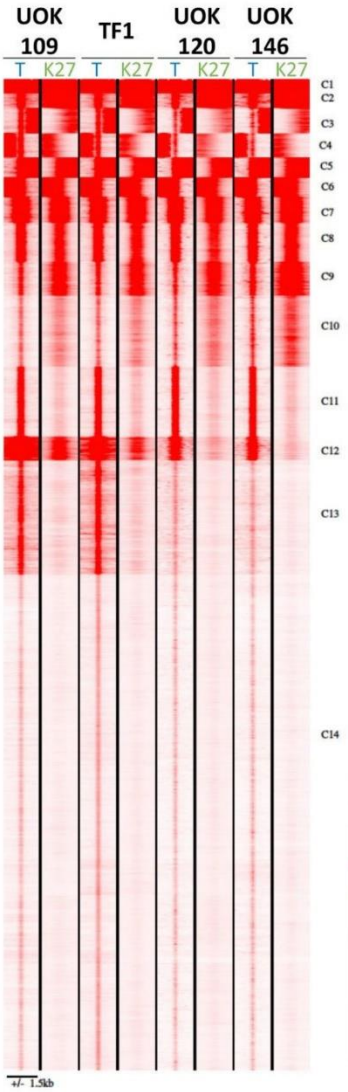


Suppl. Figure 3

Figure S3. Genes regulated by ectopically expressed TFE3 fusion proteins. A-B. Venn diagrams showing the number of deregulated genes in each line and the associated ontology analysis of the commonly regulated genes using MetaScape software. **C**. Enrichment plots of the Oxidative Phosphorylation (OxPhos) pathway according to GSEA analysis after ectopic expression of native TFE3 or NT or PT fusion proteins in HEKT cells over GFP condition.

Article 3: TFE3 fusion proteins drive oxidative metabolism, ferroptosis resistance and general RNA synthesis in translocation renal cell carcinoma

A

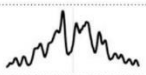
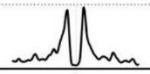
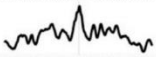
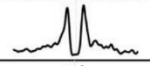
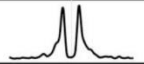


Non-redundant TFE3 peaks
(169 896 sites)

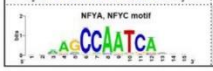
B

Cell line	Cut&Tag TFE3	ChIPseq H3K27ac	TFE3 with H3K27ac and E-box
UOK109	65901	44600	18466 (28%)
TF1	78598	56957	19919 (25%)
UOK120	33819	64545	11106 (33%)
UOK146	42760	51008	12714 (30%)

C

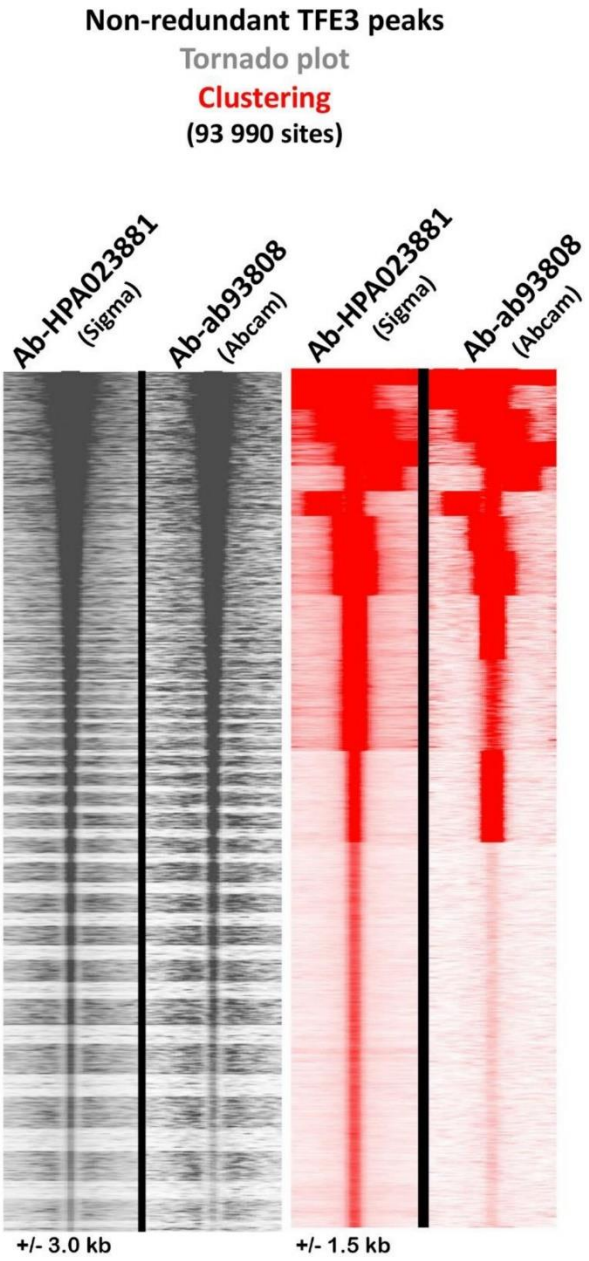
Cell line	Motif		e-value	Distribution
UOK109	SP2 (MA0516.3) KLF5 (MA0599.1) SP4 -MA0685.2)		3.4e-126	
	FOSL2 (MA0478.1) FOSB::JUNB (MA1135.1) FOS::JUN (MA0099.3)		2.0e-120	
	MITF (MA0620.3) TFEB (MA0692.1) TFEB_full		1.7e-53	
TF1	FOSL2 (MA0478.1) FOSL2::JUNB (MA1138.1)		4.0e-164	
	KLF12 (MA0742.2) KLF10 (MA1511.2) SP2 (MA0516.3)		8.1e-84	
	MITF (MA0620.3) USF1_DBD TFEB (MA0692.1)		3.2e-46	
UOK120	TFEB (MA0692.1) TFEB_full TFE3_DBD		3.0e-618	
UOK146	TFEB (MA0692.1) TFEB_full TFE3_DBD		8.6e-587	

D

Top1000 Peaks – RSAT analysis	UOK109		TF1		UOK120		UOK146	
Motif	Rank	e-value	Rank	e-value	Rank	e-value	Rank	e-value
	2	2.9e-61	6	3.4e-28	1	1.6e-66	2	6.9e-56
	1	1.5e-79	1	5.6e-73	2	3.0e-54	1	3.6e-58
	3	5.8e-29	3	7.6e-30	/	/	/	/

Suppl. Figure 4

Figure S4. Profiling of TFE3 fusion protein genomic occupancy. A. Read density maps for TFE3 fusion protein occupancy (T) and H3K27ac (K27) at all non-redundant sites. **B.** Total numbers of peaks with the indicated characteristics in each cell line after removal of low-occupied sites (C14 of the density map in panel A). **C.** Results of MEME-ChIP analyses of the DNA motifs at the top 1000 TFE3 bound sites in each cell line, with associated e-value and motif distribution over the peak. **D**. Results of RAST analyses of transcription factor binding motifs at the top 1000 TFE3 bound sites in each cell line showing associated rank and e-value.



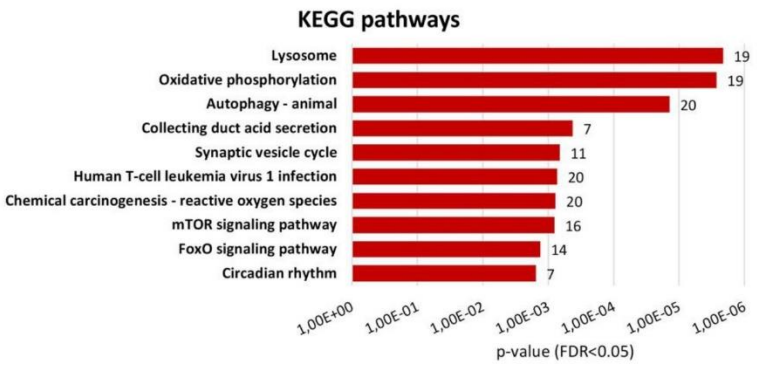
Suppl. Figure 5

Figure S5. Replicate TFE3 Cut&Tag using independent antibodies in UOK109 cells. Tornado (left panel) and read density plots (right panel) of the replicate TFE3 Cut&Tag experiments with the antibodies shown above each track.

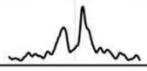
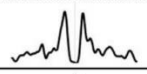
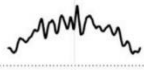
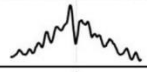
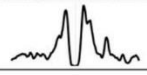
A

HEK293T Cell line	Cut&Tag TFE3 peaks	TFE3 peaks at TSS (500nt)
+ GFP	1058	615 (58%)
+ TFE3	17250	9716 (56%)
+ NONO-TFE3	28815	14690 (51%)
+ PRCC-TFE3	20255	9707 (48%)

B



C

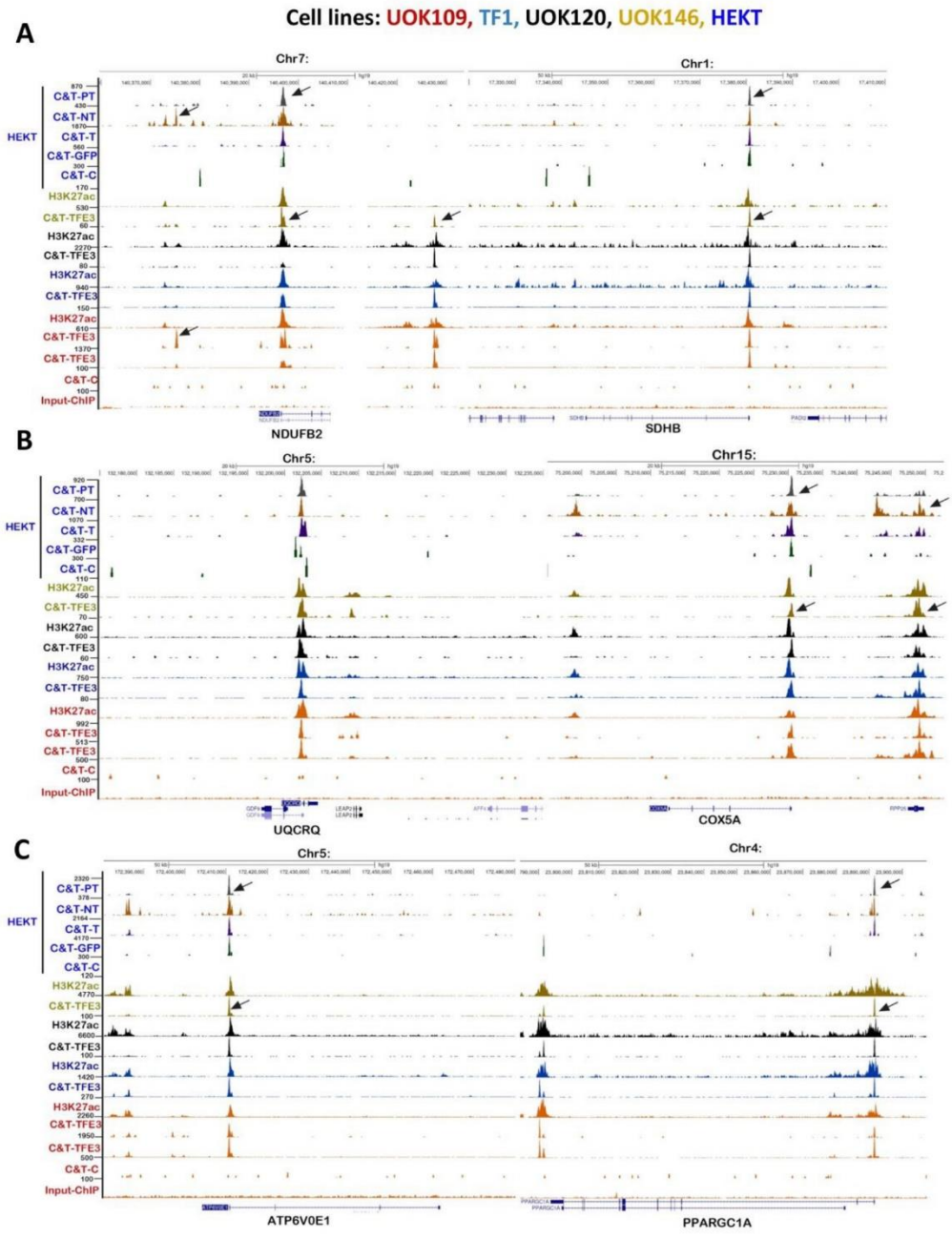
Cell line HEK293T +	Motif		e-value	Distribution
GFP	TFE3_DBD TFEB (MA0692.1) TFEB_full		1.0e-184	
TFE3	KLF15 (MA1513.1) KLF12 (MA0742.2) KLF10 (MA1511.2)		2.5e-203	Not centrally enriched
	TFEC (MA0871.2) USF2 (MA0526.4) Bhlhb2_primary (UP00050_1)		1.2e-54	
NONO-TFE3	KLF15 (MA1513.1) SP2 (MA0516.3) KLF12 (MA0742.2)		3.6e-428	
	NFYA (MA0060.3) NFYC (MA1644.1) NFYB (MA0502.2)		3.2e-330	
PRCC-TFE3	SP2 (MA0516.3) KLF12 (MA0742.2) KLF10 (MA1511.2)		1.6e-196	
	TFE3_DBD TFEB (MA0692.1) TFEB_full		1.1e-91	

D

Number of OxPhos genes common to the 4 tRCC lines With TFE3+H3K27ac-Ebox associated peak	
Among the 235 OxPhos-signature genes	145
Among the downregulated genes in siTFE3 (54)	33
Among the upregulated genes in HEK+ (59)	44
Among both downregulated genes in siTFE3 + upregulated genes in HEK+ (23)	19

Suppl. Figure 6

Figure S6. TFE3 binding in HEKT cells. A. Total numbers of peaks and those at the proximal promoters (500nt from the TSS) in the indicated HEKT cell line. **B.** KEGG ontology analyses of genes with binding of endogenous TFE3 in the proximal promoter in the HEK-GFP line. Each indicated KEGG pathway presented an FDR < 0.05 and number of genes found in each pathway were noted to the right of the bar. **C.** Results of MEME-ChIP analyses of the DNA motifs at the top 1000 TFE3 bound sites in each cell line, with associated e-value and motif distribution over the peak. **D.** Number of OxPhos genes in each category associated with TFE3 occupied M/E box containing a H3K27ac-marked regulatory elements common to each cell line.

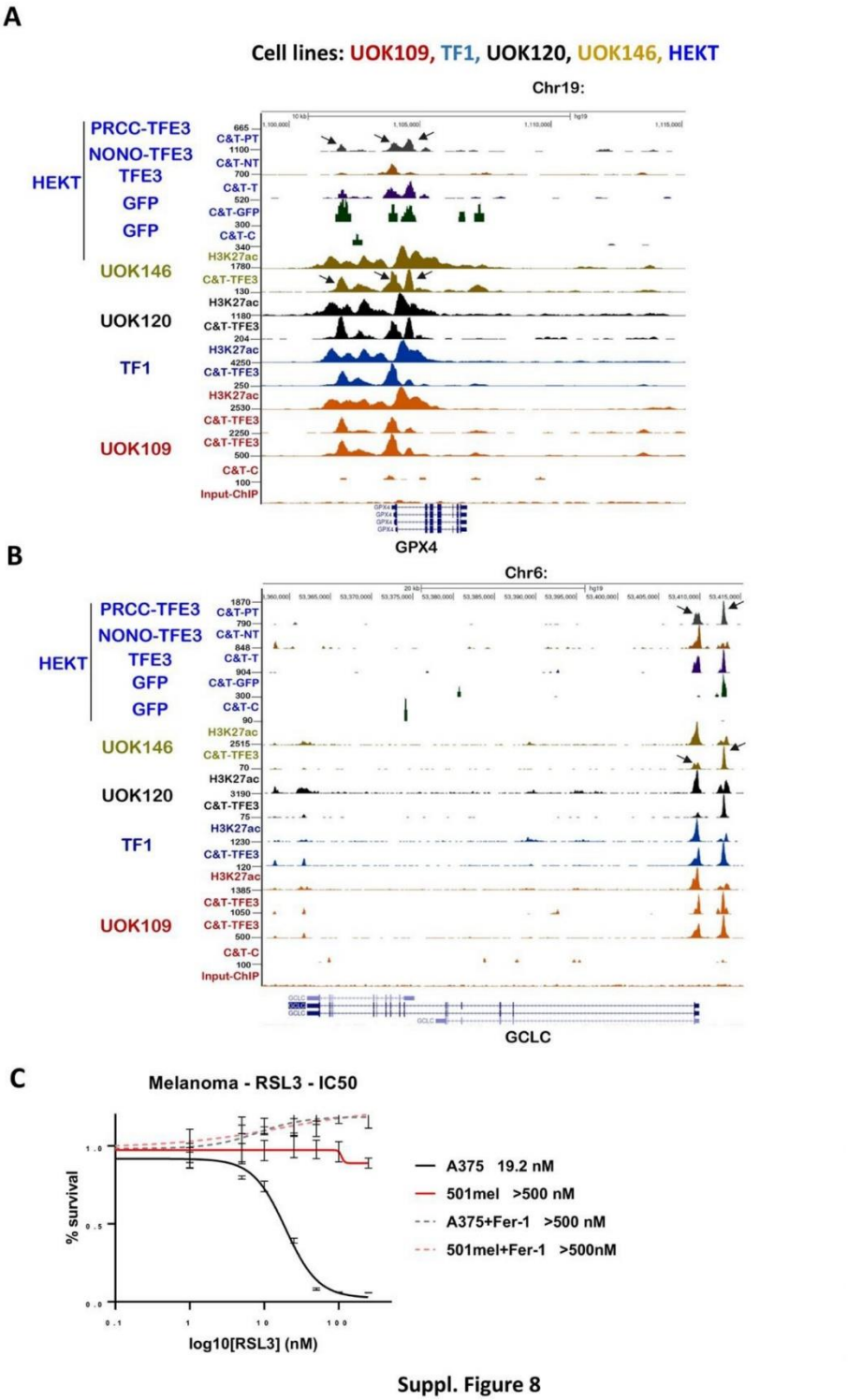


Suppl. Figure 7

Figure S7. Representative examples of TFE3 fusion protein binding at genes involved in OxPhos. A-C. UCSC genome browser views of the TFE3 C&T, and H3K27ac ChIP-seq in

each cell line, as indicated by the colour code, at the indicated loci: *NDUFB2* complex I (A, left panel); *SDHB* complex II (A, right panel); *UQCRQ* complex III (B, left panel), *COX5A* complex IV (B, right panel), the *ATPV6V0E1* (C, left panel) and *PPARGC1A* (C, right panel) gene loci. Arrows indicate TFE3 binding to H3K27ac-maked sites comprising an M-box motif.

Article 3: TFE3 fusion proteins drive oxidative metabolism, ferroptosis resistance and general RNA synthesis in translocation renal cell carcinoma



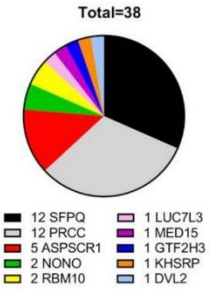
Suppl. Figure 8

Figure S8. TFE3 fusion regulation of ferroptosis. A-B. UCSC genome browser views of the TFE3 C&T, ATAC-seq and H3K27ac ChIP-seq in each tRCC and HEKT cell line, as indicated by the colour code, at the GPX4 (A) and GCLC (B) loci. **C**. RSL3 IC50 values of control melanoma lines in presence or absence of 1 mM ferristatin-1 (Fer-1) as indicated.

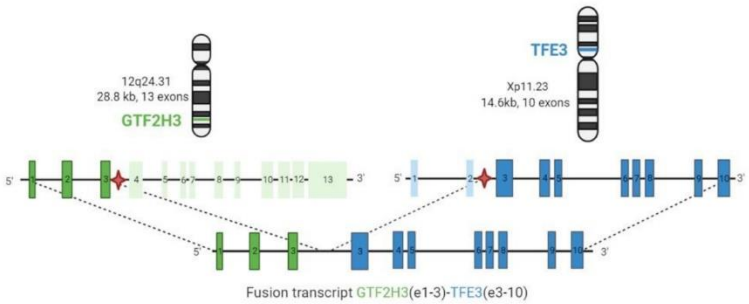
A

Clinicopathologic characteristics	Total (n=38)	M stage, n (%)	
Age, median (range)	33.5 (3-65)	M0	29 (76.31%)
Gender, n (%)		M1	6 (15.79%)
Male	13 (34.21%)	M2	1 (2.63%)
Female	25 (65.79%)	Mx	2 (5.26%)
T stage, n (%)		Stage, n (%)	
≤T2	23 (60.53%)	I-II	20 (52.63%)
≥T3	15 (39.47%)	III-IV	18 (47.37%)
N stage, n (%)		Alive status, n (%)	
N0	24 (63.16%)	Alive	24 (63.16%)
N1	6 (15.79%)	Dead	14 (36.84%)
N2	4 (10.53%)	OS, median (range) (months)	47 (8.33-231.17)
Nx	4 (10.52%)		

B

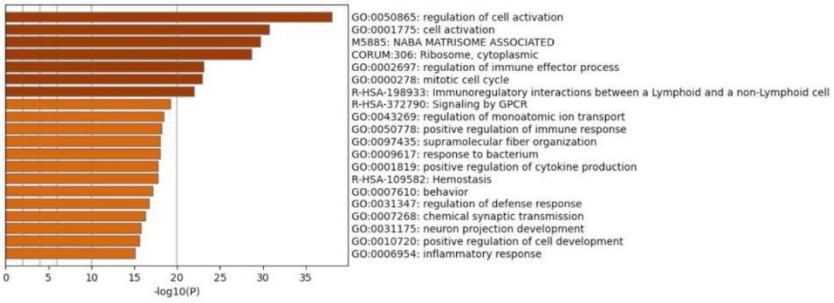


C



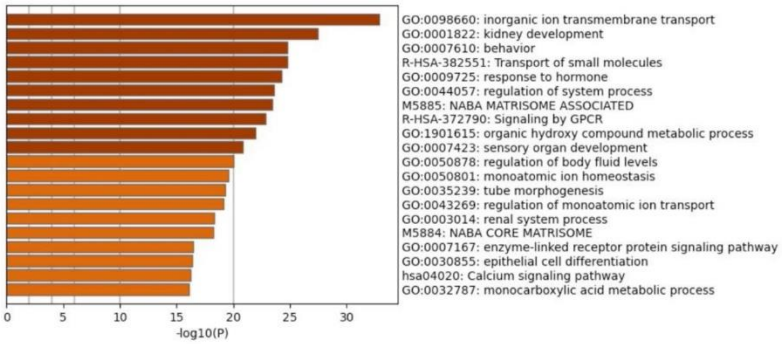
D

Gene ontology on protein coding genes upregulated in Tumor samples vs NAT (2856g)



E

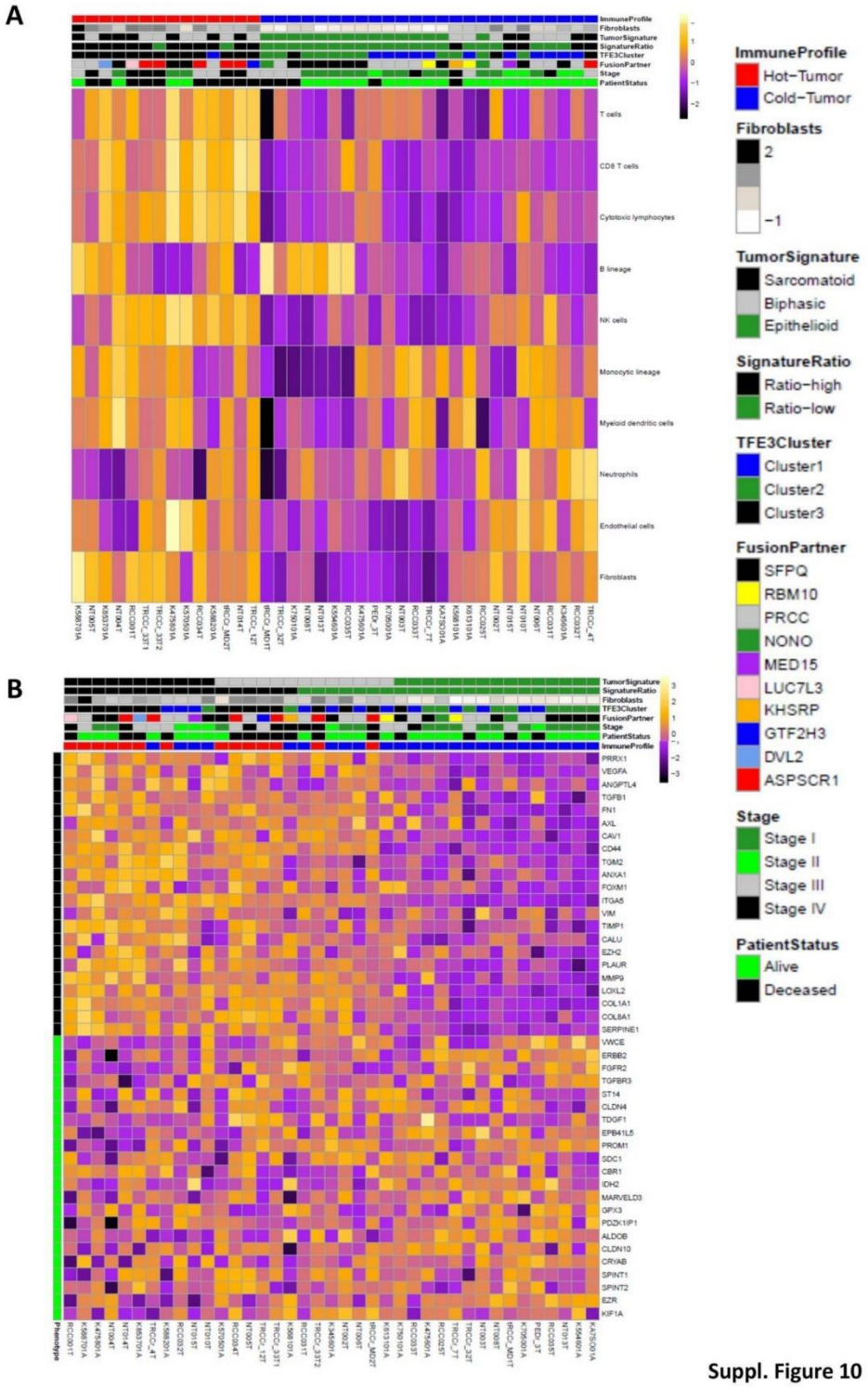
Gene ontology on protein coding genes upregulated in NAT vs Tumor samples (2146g)



Suppl. Figure 9

Figure S9. Clinical characteristics of the tRCC cohort. A. Table presenting the clinical and pathology characteristics of the tRCC patient tumor cohort. **B**. Pie chart showing the contribution of the indicated fusion partners to the cohort. **C.** Schematic representation of a novel *TFE3* fusion with *GTF2H3* encoding the p34 subunit of general transcription factor TFIIH. **D-E**. Ontology of genes up- (D) or down- (E) regulated in tRCC tumours compared with NAT.

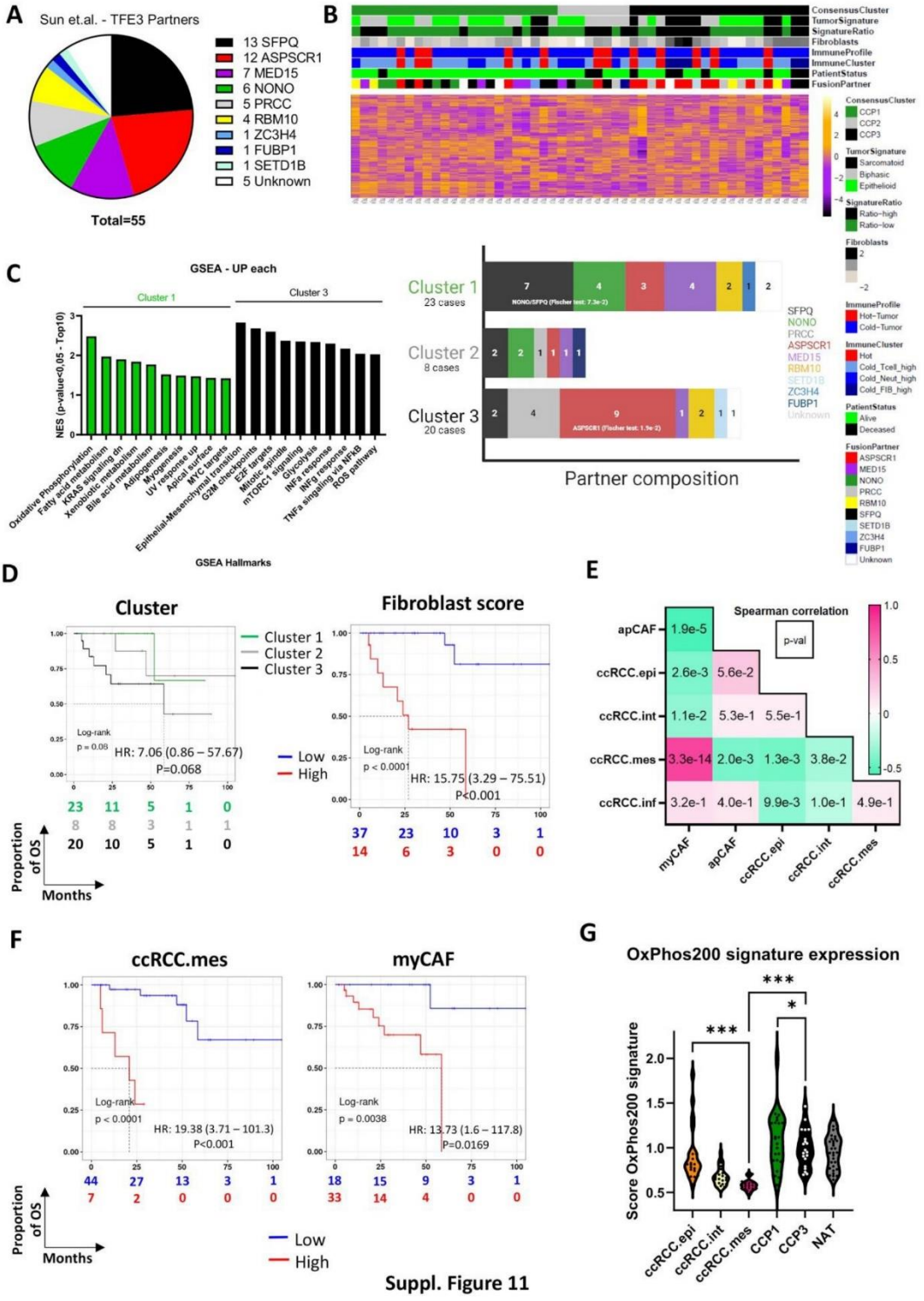
Article 3: TFE3 fusion proteins drive oxidative metabolism, ferroptosis resistance and general RNA synthesis in translocation renal cell carcinoma



Suppl. Figure 10

Figure S10. Immune infiltration and EMT in tRCC. A. MCP counter analyses of the tRCC cohort. **B.** Heatmap of the differential expression of a collection of epithelial and EMT related makers in the tRCC cohort displayed as a Z score.

Article 3: TFE3 fusion proteins drive oxidative metabolism, ferroptosis resistance and general RNA synthesis in translocation renal cell carcinoma



Suppl. Figure 11

Figure S11. Clinical and molecular characteristics of an independent tRCC cohort. A. Pie chart showing the contribution of the indicated fusion partners to the cohort. **B.** Heatmap of the unsupervised clustering with the indicated clinical and molecular parameters showing the division of samples in 3 clusters (upper panel). Distribution of samples by fusion partner within the 3 major clusters (lower panel). **C.** GSEA analyses of genes differentially expressed between clusters 1 and 3 indicating the Top10 enriched terms (p-value<0.05, FDR<0.05). **D**. Kaplan-Meier curves for overall survival in patients according to the cluster or the fibroblast score, calculated from MCP-counter, using the optimal cut point method with the associated log-rank p-value and Hazard ratio (HR) from the univariate Cox proportional-hazard model. **E.** Spearman correlation coefficient (coloured box) and associated p-value (number in box) between the indicated populations after deconvolution using the ccRCC tumour cell and CAF signatures inferred by CIBERSORTx on bulk RNA-seq data from the tRCC tumour samples. **F.** Kaplan-Meier curves for overall survival in patients according to ccRCC.mes or myCAF scores using the optimal cut point method with the associated log-rank p-value and Hazard ratio (HR) from the univariate Cox proportional-hazard model. **G**. Expression score of the OxPhos genes in NAT, each tRCC cluster and the indicated ccRCC tumour types from the TCGA.KIRC collection. Overall scores were compared by Wilcoxon test (*: p.val<0.05, **: p.val<0.01, ***: p.val<0.001).

Supplemental Dataset 1. Genes up and down regulated by siTFE3 compared to siCTR in each tRCC cell line. On each spreadsheet are indicated Gene name, Ensemble gene ID, Description, normalized reads in each condition, Log2 fold change and adjusted p-value. Note for up-regulated genes log2 fold change > + 1, and adjusted p-value <0.05, but for down-regulated genes log2 fold change <- 0,5, and adjusted p-value <0.05.

Supplemental Dataset 2. OxPhos Genes. Page 1. Legend, **Page 2**. List of OxPhos genes from the GSEA signature, **Page 3.** non-redundant list of OxPhos genes from the GSEA and KEGG signatures signature.

Supplemental Dataset 3. Genes up and down regulated by ectopic TFE3 and fusion protein expression in HEKT cells. On each spreadsheet are indicated. Ensemble gene ID, Gene name, Description, gene biotype (only coding genes are listed) GO, normalized reads in each condition, Log2 fold change and adjusted p-value.

Supplemental Dataset 4. Page 1. List of E/M-box containing peaks in open chromatin comment to all 4 tRCC cell lines. Indicated are peak ID, Peak coordinates, Chromosomal locations of peaks, locations relative to gene, distance to TSS, nearest Ensemble ID of nearest promoter, gene name and description. **Page 2.** Core program of protein coding genes with a TFE3 peak +/- 500 nt from the TSS. **Page 3.** KEGG ontology analyses of the core program protein coding genes.

Article 3: TFE3 fusion proteins drive oxidative metabolism, ferroptosis resistance and general RNA synthesis in translocation renal cell carcinoma

Article 4: The chromatin remodeler BRG1 is required for melanoma cell plasticity

This article has not been submitted yet, it is in the final phases of redaction. It describes the role of BRG1 (encoded by *SMARCA4*), a component of the SWI/SNF complex. BRG1 is essential for 3D growth of mesenchymal cells, in association of PRRX1, a transcription factor known to promote cell invasion.

I helped in the characterization of the interaction between PRRX1 and BRG1 by co-immunoprecipitation. This PRRX1/BRG1 complex promotes the expression of genes implicated in cell invasion and EMT. I also assessed the impact of the depletion of BRG1 or BRM on other proteins of the SWI/SNF complex and helped the main author to perform spheroid formation assays to characterize the effect of BRG1 depletion in these conditions.

The chromatin remodeler BRG1 is required for melanoma cell plasticity.

Bujamin Vokshi [1], Guillaume Davidson [2], Antonin Lallement [2], Alexandre Haller [2], and Irwin Davidson [2,3]#.

[1] Current address. Toska lab, Dept. of Oncology, Sidney Kimmel Cancer Center Johns Hopkins University School of Medicine 1651 Jefferson St., Baltimore MD 21287

[2] Institut de Génétique et de Biologie Moléculaire et Cellulaire. 1 rue Laurent Fries, 67404 Illkirch Cedex, Strasbourg, France.

[3] 'Equipe Labellisée' Ligue National contre le Cancer. Paris, France.

To whom correspondence should be addressed

E mail: irwin@igbmc.fr

Fax: +33 3 88 65 32 01. Tel: +33 3 88 65 34 45

Running Title: BRG1 in dedifferentiated melanoma cells

Key words: chromatin remodelling, SMARCA4, PRRX1, plasticity

The authors declare no potential conflicts of interest.

Abstract.

Cell heterogeneity is a key feature of melanoma tumours. The melanocyte lineage identity factor Microphthalmia-associated transcription factor (MITF) regulates gene expression in more differentiated melanocytic-type melanoma cells using the SWI/SNF chromatin-remodeling complex as a cofactor. Here, we address the role of SWI/SNF catalytic subunit BRG1 in undifferentiated mesenchymal (MES) melanoma cells that do not express MITF. BRG1 silencing regulated motility and invasion, but had minimal effects on proliferation of MES cells cultured in 2D conditions. In contrast, BRG1 was required for the formation of 3D melanospheres that involved transition towards a more neural-crest stem cell-like (NC) state. BRG1 binds a set of MES-specific enhancers together with PRRX1, a key transcription factor promoting melanoma invasion, that is upregulated in NC/MES intermediate state melanospheres along with many of its target genes. BRG1 and PRRX1 bound sites overlap with those of AP1 and we show that PRRX1 interacts with both AP1 and SWI/SNF. Our data provide novel insights into how interplay between BRG1, AP1 and PRRX1 regulate gene expression in MES melanoma cells and their transition to a MES/NC intermediate state. This NC/MES intermediary state, highly expressing PRXX1 and other key EMT transcription factors, more closely mimics the characteristics of metastatic melanoma than fully mesenchymal cells.

Introduction.

Melanoma is the most aggressive skin cancer arising from the oncogenic transformation of melanocytes by several recurrent somatic mutations, the most frequent of which are *BRAF*^{V600E} an activating mutation found in 50-60% of tumours, *NRAS* (≈20% of tumours) and loss of function mutations in *NF1* (≈10%) [1]. Despite progress in patient outcome through the use of targeted MAP kinase inhibitor and immune checkpoint therapies, a large number of patients acquire resistance or are considered as non-responders. One of the major factors leading to resistance is cellular plasticity and heterogeneity [2]. Many studies defined cell populations with different epigenetic profiles and transcriptional signatures in cell lines, PDX tumours and mouse and human melanoma [3 4 5 6] comprising melanocytic (MEL), neural crest stem cell (NC) type, mesenchymal (MES) and various intermediate [7 8] cell states whose identities are determined by the expression and activity of key transcription factors [9 10].

We previously characterized how MITF and SOX10 drive the MEL program. Integration of ChIP-seq and RNA-seq after their siRNA-mediated silencing identified a large set of direct target coding and non-coding genes that drive melanoma proliferation and survival [11 12]. Proteomics also revealed that MITF and SOX10 interacted with the SWI/SNF remodelling complex and we showed that the catalytic subunit BRG1 is essential for proliferation of MEL cells, melanocyte development and mouse melanoma *in vivo* [12 13]. MITF and SOX10 actively recruit BRG1 to chromatin to establish the epigenetic landscape of MEL cells and promote the melanocytic gene expression program.

Here, we investigated the role of BRG1 in MES melanoma cells that do not express MITF or SOX10. Our findings demonstrate that BRG1 silencing in MES cells had minimal impact on cell proliferation and gene expression under 2D culture conditions. However, it had a significant regulatory effect on genes essential for the formation of 3D melanospheres that

involved a transition towards a more NC state characterized by up-regulation of expression of PRRX1 and other key transcription factors involved in epithelial to mesenchymal transition (EMT). Chromatin immunoprecipitation sequencing (ChIP-seq) revealed localization of BRG1 to distinct sets of regulatory elements in MES compared to MEL cells. In MES cells, BRG1 localizes to AP1-TEAD-bound enhancers a subset of which are additionally bound by transcription factor PRRX1. Our data provide evidence that BRG1 plays an important role melanoma cell plasticity and the transition between the MES and NC states. MES cells grown as 3D spheroids adopt a NC/MES intermediate state, highly expressing PRRX1 and other key EMT transcription factors, that better mimics the characteristics expected of metastatic melanoma than fully MES cells.

Results.

BRG1 is required for optimal growth of mesenchymal melanoma cells.

To address the function of BRG1 in MES cells, we used 3 previously characterised primary cultures from patients MM099, MM047 and MM029 [5 14] each of which display MES gene expression signatures and epigenetic profiles, but different oncogenic mutations, $BRAF^{V600E}$ in MM099, $BRAF^{V600K}$ in MM029 and $NRAS^{Q61*}$ in MM047. To better characterise the gene expression profiles of these and other melanoma lines, we compared their gene expression signatures with the previously described signatures for MES, NC, MEL and IFNg/antigen presenting (IFNg) as well as additional intermediate state signatures [7]. As previously reported [5], the MM099 signature (MM099 cells grown in 2D conditions) was characteristic of the MES phenotype similar to that of the MM029, and MM074R [15] (Fig. S1A). MM047 on the other hand, displayed a signature closer to an NC/MES intermediate than a fully MES signature. This is in agreement with the fact that the MM047 cells were driven to a more MES-like state upon resistance to the CDK7 inhibitor THZ1 [15]. In contrast, 501Mel, IGR37,

Mel888 displayed the expected MEL signature, whereas MM074 had a more INT/MEL phenotype consistent with its ability to adopt a more MEL (MM047VR) or MES (MM074R) state upon drug challenge [15]. These signatures were also reflected in the differential expression of a set of marker genes with MITF and SOX10 highly expressed in the MEL state, NES, NGFR and PRRX1 most expressed in the NC state, and SOX9 in the MES state (Fig. S1B). Extending these analyses to the public data on the MM0-lines from Wouters et al [14] consolidated the differential expression of key marker genes in the MEL-INT-NC-MES gradient (Fig. S1C-D). Immunoblot experiments confirmed the lack of detectable MITF and SOX10 expression, but expression of SOX9 in the MM029, MM047 and MM099 lines, whereas the opposite was seen in 501Mel cells (Fig. S2A). In contrast, similar levels of expression of SWI/SNF subunits including BRG1, BRM, BAF170, BAF155 and BAF53 were seen in both cell types.

To address the role of BRG1, we performed siRNA-mediated silencing of BRG1, BRM or control siRNA (siC). Each siRNA selectively targeted either BRG1 or BRM with no change in the RNA level of BRM upon BRG1 silencing and *vice versa* (Fig. S2B). BRG1 silencing did however lead to BRM protein accumulation without affecting mRNA level and *vice versa* suggesting a competition between the catalytic subunits for SWI/SNF complex formation (Fig. 1A). BRG1 silencing elicited only a minor reduction in colony forming capacity of MM099 and MM029 cells, but a stronger effect in MM047 cells (Fig. 1B). BRG1 silencing also had only minor impact on proliferation of MM099 and MM029 cells with a mild increase in slow proliferating cells, but a more important effect in MM047 cells (Fig. 1C). BRM silencing had little effect in all lines. Similar to the above, BRG1 silencing did not induce apoptosis and induced senescence only in MM047 cells, while BRM silencing had little effect in all cell types (Figs. 1D and S2C).

We performed RNA-seq on siBRG1- and siBRM-silenced MM099 and MM047 cells. Consistent with the minimal effects on cell proliferation, only 156 genes were deregulated in

MM099 and 137 in MM047 cells using a standard cut-off (log2 fold change ± 1 and adjusted p-value ≤0.05: Fig. 1E and Supplemental Dataset 1). Nevertheless, close to 1000 de-regulated genes were identified using a less stringent cut-off (log2 fold change ± 0.5 and adjusted p-value ≤0.01) in each cell type (Fig. S3A). Gene set enrichment analyses (GSEA) revealed an increase in terms associated with inflammation and cytokine signalling in both lines and a reduction in apical junction and cell proliferation that was more pronounced in MM047 consistent with the slower growth seen above (Fig 1F). In agreement with the GSEA, further ontology analyses confirmed up-regulation of inflammation and cytokine signalling as well as cell migration (Fig. 1G). Cellular compartment (CC-FAT) ontology analyses showed that both up and down regulated genes were predominantly associated with the membrane and the extracellular matrix. Comparison between the lines revealed 104 up and 94 common down-regulated genes that showed similar ontology signatures to those of each line individually, with cell migration and cytokine signalling as represented pathways (Fig. S3B-C). In contrast, BRM silencing regulated only a small number of genes that were not further analysed.

Together the above results showed that BRG1 silencing had little impact on cell physiology and/or gene expression in the most MES MM099 and MM029 lines, but a more pronounced effect in the MM047 cells that displayed a more MES/NC transition state phenotype. These effects remained mild compared to the potent proliferation arrest, senescence and strongly de-regulated gene expression previously reported when the same siRNAs were used to silence BRG1 in 501 Mel cells[12].

BRG1 regulates invasion and migration of mesenchymal melanoma cells.

The above ontology analyses suggested a potential role of BRG1 in cell adhesion and migration. In a Boyden chamber assay with a matrigel barrier, siBRG1 silencing promoted increased invasion of all 3 cell lines, whereas silencing of BRM had no visible effect (Fig. S4A). Similarly, shBRG1 silencing also strongly increased invasion that was not noticeably

increased by further siBRM silencing (Fig. S4B). Furthermore, medium from the siBRG1, but not the siBRM or siC silenced cells increased invasion of untreated MM099 cells indicating that secreted cytokines were responsible at least in part for the increased invasive capacity, consistent with the RNA-seq results described above (Fig. S4C). In a wound-healing assay, siBRG1 silencing increased cell motility allowing a more rapid wound closure compared to siC or siBRM (Fig. S4D). Thus, paradoxically, BRG1 acts to negatively regulate migration and invasion of these cells that have otherwise been characterized by their high motile and invasive characteristics [14].

Differential BRG1 genomic occupancy in mesenchymal and melanocytic cells.

To better understand the role of BRG1 in the MM099 cells, we performed native BRG1 ChIP-seq as previously described in 501Mel melanocytic cells[12] and ChIP-seq for H3K27ac. More than 110 000 BRG1 peaks were detected of which around 48% co-localized with H3K27ac-marked nucleosomes (Fig. 2A). Ontology analyses of the genes associated with the BRG1-bound sites in each cluster revealed a wide range of cellular functions in agreement with the large number of bound promoters (Fig. 2B).

It has been previously shown that AP1 and TEAD factors are key determinants of the MES gene expression program [5 16 14]. We integrated ChIP-seq data for FOSL2 and TEAD4 from MES-type Sk-Mel-147 cells [17] with the BRG1 and H3K27ac ChIP-seq data. Alignment with more than 78 000 FOSL2 bound sites revealed several clusters, where AP1 and TEAD4 were located between BRG1-bound and H3K27ac-marked nucleosomes (C1), where AP1 and TEAD4 were bound between H3K27ac-marked nucleosomes with reduced BRG1 (C2) and where AP1 and TEAD4 located between BRG1-bound nucleosomes in absence of H3K27ac (C3, Fig. 2C-D). Cluster C1 therefore defines a set of potentially transcriptionally active regulatory promoter and enhancer elements.

We next compared BRG1 genome distribution in MM099 and 501Mel cells. More BRG1-bound sites were seen in MM099 compared to 501Mel, although the proportions in each genomic compartment were comparable (Fig. S5A). The majority of sites were cell-specific, but with around 14 000 sites commonly occupied in both cell types (Figs. S5A and B). Ontology analyses of the nearest genes in each cluster showed that the commonly occupied sites were associated with diverse fundamental cellular functions such as transcription, translation and intracellular transport (Fig. S5C). The MM099-specific sites were associated with inflammatory response or angiogenesis, whereas 501Mel-specific sites were enriched in genes involved in cell cycle, mitosis and cell division consistent with our previous observations that MITF and SOX10 recruited BRG1 to regulatory elements associated with these genes associated with these functions [12]. Moreover, the term cell cycle appears in both the common and 501Mel-specific clusters defining distinct gene sets regulated in both cell types or specifically in 501Mel cells. For example, the Citron kinase (CIT) is a key Rho effector that functions to maintain proper the mid-body structure during cell mitosis [18]. In 501Mel cells, the *CIT* locus comprises multiple MITF binding sites, a SOX10 binding site with H3K27ac and prominent BRG1 at the promoter and intergenic regions (Fig. S5D). In MM099 cells, BRG1 occupancy was restricted to sites at the 3' end of the gene. CIT is regulated by MITF and BRG1 in 501Mel cells, but not in MM099 cells, and is therefore an example of a mitotic gene whose expression was regulated under the control of MITF and SOX10 specifically in MEL cells [12]. Similarly, BRG1 and H3K27ac were widely distributed across the *MITF* locus in 501Mel cells, whereas H3K27ac was lost in MM099 cells and BRG1 was restricted to the promoter region of the A and B isoforms (Fig. S5D). The opposite was seen at the MES expressed *SOX9* locus, with selective H3K27ac and BRG1 occupancy in MM099 cells (Fig. S5E).

Minnoye et al, [16] used ATAC-seq to identify putative melanocytic- or mesenchymal-specific enhancer elements based on differential chromatin accessibility in melanoma from

multiple species and identified conserved melanoma enhancers showing cross-species accessibility. Around 40% of the identified MES regulatory elements showed strong BRG1 occupancy at the flanking nucleosomes in MM099 cells and around 30% showed strong concomitant H3K27ac signal (Fig. 2E). In contrast, only a small subset of sites occupied in 501Mel cells corresponded to MES enhancers. Conversely, more than 50% of MEL enhancers showed BRG1 occupancy and H3K27ac in 501Mel cells, with only a subset showing signal in MM099 cells (Fig. 2E). These data showed that BRG1 selectively occupied MES or MEL regulatory elements in a cell-specific manner. Moreover, for each cell type, sites showing highest signal for BRG1 and H3K27ac were those that showed the highest percentage of cross-mammalian species conservation (Fig. 2F).

BRG1 promotes a mesenchymal to a neural crest-like state transition upon growth as 3D spheroids.

The presence of BRG1 at MES enhancers contrasted with the observation that BRG1 did not appear to be a major regulator of MES cell proliferation under standard 2D growth conditions. To better address the role of BRG1, we asked whether it may play a more important role in regulating 3D growth. MES cells were transfected with siRNAs and seeded for growth as 3D melanospheres. In all 3 lines, siBRG1 silencing efficiently inhibited melanosphere formation, while siBRM had no measurable effect (Fig. 3A). Similarly, use of a shRNA targeting BRG1 also potently inhibited 3D growth of MM099 cells (Fig. 3B). We analysed the effect of BRG1 silencing in ultra-low attachment round-bottom plates where individual spheres can be visualized and the number of viable cells subsequently determined by an ATP-cell quantification assay. As seen with free growing spheres, BRG1 silencing led to a potent reduction in single spheroid growth of MM099 cells (Fig. 3C). Thus, while siBRG1 silencing had limited effect on cell proliferation in 2D, it had a more potent effect on 3D spheroid formation.

Given this more potent effect, we performed RNA-seq after siBRG1 or siBRM silencing of cells grown under 3D conditions. We first made a comparison of the siC control in the 2D and 3D conditions to identify genes de-regulated upon the transition from monolayer to spheroid growth. An important re-programming of gene expression was observed with more than 2600 and 3000 genes up or down-regulated in 3D conditions, respectively (Fig. 4A and Supplemental Dataset 2). In 3D conditions, ontology and GSEA analyses revealed a strong enrichment in genes associated with the extracellular matrix, hypoxia, TNF and inflammatory signalling, whereas genes involved in cell proliferation and DNA replication were down-regulated (Fig. 4B-C). The 2D to 3D transition is thus associated with reduced proliferation, increased hypoxia and a major reorganisation of the cell membrane and extracellular space.

To better understand the altered gene expression program, we compared the MM099 2D and 3D gene expression signatures in the context of those described above in Fig. S1. In comparison to the MM099 2D signature, the MM099 3D shifted towards a MES/NC transition signature with similarity to the A375 and SK-Mel-24 NC-type cells (Fig. S1A-B). The shift from a MES to NC-MES transition state upon 3D growth spheres up-regulated NES and a set of transcription factors including PRRX1, but down-regulated AXL and EGFR (Fig. S1E). Nevertheless, expression of SOX10 or NGFR, 2 key markers of the full NC state were only weakly expressed. Hence, while BRG1 had minor effects on MM099 in a 2D MES setting, it was required for the shift towards the NC state associated with 3D sphere formation. None of the lines showed a strong IFNg signature apart from IFNg-treated MM074 cells, but the MM099 3D cells displayed a mild IFNg signature consistent with the increased interferon and inflammatory signals seen in the ontology analyses (Fig. S1A and Fig. 4B-C).

BRG1 silencing in spheroids led to reduced expression of more than 430 genes and up-regulation of 193 genes (Fig. 4D and Supplemental Dataset 2). GSEA and ontology analyses indicated up-regulation of genes associated with the interferon and inflammatory pathways and

cell proliferation, whereas down-regulated genes were associated with the extracellular matrix and hypoxia signatures in agreement with the reduced 3D growth (Fig. 4E-F). Thus, many ontology terms associated with genes up-regulated upon 3D growth (hypoxia, angiogenesis) were downregulated upon BRG1 silencing (Fig 4F). Strikingly, a set of more than 243 genes that were strongly induced upon 3D growth were down-regulated by siBRG1 silencing, whereas a smaller set of genes normally repressed upon 3D growth were up-regulated by 3D silencing (Fig. 4G, Supplemental Dataset 3). Comparison with previous signatures showed that many genes associated with the locomotion and ECM signature described by Verfaillie *et al.*[17] as TEAD4 targets were either induced or repressed upon 3D growth and this regulation was upset by BRG1 silencing (Fig. 4H, Supplemental Dataset 4). RNA-seq of siBRM knockdown spheres where no effect on growth was observed revealed minimal effects on gene expression with little effect on genes induced or repressed upon the 2D to 3D transition (Fig. 4G). These data identified a set of genes that required BRG1 for their up-regulation upon transition to 3D growth.

PRRX1 promotes MES 3D spheroid growth.

To identify transcription factors potentially involved in upregulating genes upon the 2D to 3D transition and the expression of genes down-regulated by BRG1 silencing, we analysed the RNA-seq data using the EnrichR package and the ARCHS4 data base. Analysis of the 1000 most upregulated genes upon 2D to 3D transition identified TWIST1, NFATC4, the homeodomain proteins PRRX1 and PRRX2 and OSR1/2 as the top associated factors (Fig. S6A). Similarly, analysis of the 433 genes down-regulated by BRG1 silencing identified TWIST1, OSR1, PRRX1 and PRRX2 as associated factors (Fig. S6B). Examination of the RNA-seq data showed that OSR1 and OSR2 were only weakly expressed in melanoma cells and not regulated upon the 2D-3D transition, while PRRX2 was not expressed. In contrast,

TWIST1, NFATC4 (and the related NFATC1/2) and PRRX1 expression was up-regulated in 3D conditions and TWIST1 and PRRX1 were down-regulated upon BRG1 silencing (Fig. S1E and Fig. S6C). However, TWIST1 expression was not specific to MES cells and was also expressed in 501Mel cells (Fig. S6D). Two other EMT regulators, ZEB1 and the recently described TCF4 [19], were also strong upregulated upon the transition to 3D (Fig. S6C).

Of these transcription factors, PRRX1 has recently emerged as a key factor driving melanoma invasion [8]. PRRX1 expression was activated during the transition from the MEL to MES states, up-regulated in 3D conditions and down-regulated upon BRG1 silencing (Fig. 4A and D, Fig. S1E). Examination of the *PRRX1* locus showed more extensive BRG1 occupancy in MM099, mainly of intronic regions, than seen in 501Mel cells (Fig. S5E). While PRRX1 expression was clearly associated with undifferentiated states, a closer examination of its expression in melanoma cell lines showed higher expression in the transition from the NC to MES states exemplified by A375, Sk-Mel-24 and the MM099 3D cells than in the MES state exemplified by MM099(2D) and MM029 (Fig. S1B and E). Similar observations were made in the MM-collection of lines where PRRX1 was highest expressed in the NC/MES intermediate (Fig. S1D).

Currently available antibodies that we tested worked poorly in detection of endogenous PRRX1. To overcome this and map PRRX1 genomic binding, we generated a MM099 line with ectopic doxycycline (Dox) inducible HA-tagged PRRX1 expression. We used these cells to perform Cut&TAG with HA-antibody mapping 16 600 PRRX1 binding sites. Alignment of these sites with BRG1 and H3K27ac ChIP-seq indicated a group of sites associated with both BRG1 and H3K27ac (C1), sites associated with either of the two (C2, C3) or sites devoid of both (C4, C5, Fig. 5A). We then used RSAT to analyse the DNA motifs present at the 1000 most bound PRRX1 sites. Surprisingly, the most enriched motif was FOS-JUN (AP1) followed by KLF motifs, with potential Homeodomain recognized motifs being much less numerous and

of low enrichment (Fig. 5B). In agreement with this, alignment of the PRRX1 bound sites with the FOSL2 and TEAD4 profiles from the MES-type Sk-Mel-147 cells identified a set of sites bound by all 3 factors (C1-C3, Fig. 5A). PRRX1 was also found at FOSL2 and TEAD4 sites at MES enhancers in particular those showing the highest BRG1 and H3K27ac signal (Fig. 5C). Overlap of the binding sites for each factor indicated a large number of shared sites (Fig. 5D). More than 10 000 PRRX1 binding sites were shared with either FOSL2 or TEAD4. Moreover, when we analysed the DNA-binding motifs at the remaining 6000 PRRX1 sites, JUN-FOS was again the most enriched motif with no significant enrichment of Homeodomain binding motifs (data not shown). This suggests that although these sites are not bound by FOSL2 in SK-Mel-47 cells they may be bound in MM099, or are bound by other JUN-FOS heterodimers.

These observations suggested that PRRX1 may recognize the genome via interaction with AP1, rather than by direct interaction with Homeodomain recognition motifs on the DNA. To assess this, we performed HA-immunoprecipitation of tagged PRRX1. Two independent cell extracts were precipitated with HA antibody that efficiently precipitated HA-PRXX1 (Fig. 5E). Re-incubation of this blot revealed co-precipitation with BRG1 after HA IP, but not the IgG control. Each set of IPs was then examined for the presence of other proteins. In addition to BRG1, the PBRM and ARID2 subunits of SWI/SNF were specifically co-precipitated with HA-PRRX1 (Fig. 5F). Similarly, FOSL2 and JUN also specifically coprecipitated with HA-PRRX1 (Fig. 5G). These data therefore supported the idea that PRRX1 was recruited to the genome through interactions with AP1 and interacted with the SWI/SNF complex.

We next examined the requirement of PRRX1 and its regulation of target genes in 3D-spheroid growth. SiRNA-mediated PRRX1 silencing reduced proliferation of MM099 cells but did not induce senescence or apoptosis (Fig. 6A-B) Similarly, in agreement with PRRX1 up-regulation upon 3D growth, siPRXX1 silencing reduced MM099 and MM047 spheroid growth (Fig. 6C). To ask if PRRX1 target genes were regulated during the 2D-3D transition, we

selected 200 genes with strongest PRRX1 bound sites in their proximal promoters for example *LNP1*, *RAB5B* and *CBX5* (Fig. S7A) and examined their expression. The expression of a majority of these genes was up-regulated upon 3D growth, while smaller subsets were down-regulated or showed little change (Fig. 6A; note that *SMARCA4* was included in the list as control although it is not a PRRX1 target gene using the criteria described above). We also examined a set of genes defined as the PRRX1-regulon [8]. Again, a majority were up-regulated upon 3D growth, with smaller numbers showing the opposite regulation or regulation by siBRG1 silencing in 3D conditions (Fig. 6B). We performed the converse analyses asking which of the 627 genes up or downregulated by siBRG1 in 3D conditions comprised a PRRX1 binding site associated with H3K27ac (consistent with an active enhancer) and BRG1 within 30kb of their TSS. Of these, 140 of the 627 deregulated genes, encompassing both up and downregulated gene sets, (22%) displayed prominent PRRX1 binding, both at promoter proximal and/or distal sites. For example, *FAT4*, downregulated upon siBRG1 in 3D, displayed prominent PRRX1 binding associated with FOSL2, BRG1 and H3K27ac downstream of the TSS (Fig S7B). At the *SERPINB2* locus upregulated by siBRG1, PRRX1 bound to both promoter proximal and upstream sites associated with H3K27ac, BRG1 and AP1/TEAD4 (Fig. S7B). Ontology analyses of these 140 genes showed strong enrichment in membrane and extracellular proteins involved in cell motility and receptor signalling (Fig. 6C). Together these data identify a set of PRRX1 target genes most of which were upregulated upon the 2D-3D transition along with PRRX1 itself.

Discussion.

Here we investigated the role of BRG1 in MES melanoma cells. BRG1 silencing had little effect on cell proliferation in the most fully MES MM099 or MM029 cells grown under standard 2D conditions, but impacted the transition from a MES to an NC/MES state when

grown as 3D spheroids. In contrast, BRG1 silencing had more impact in MM047 cells that displayed a NC/MES intermediate state even under 2D conditions, consistent with our previous observation that they adopted a more MES state upon acquisition of THZ1 resistance [15]. BRG1-mediated chromatin remodelling therefore seemed to have greater regulatory function in this intermediate state and upon state transitions.

Previously, Wouters [14] showed that MM099, MM047 and MM029 cells all displayed high motility. Surprisingly, BRG1 silencing further increased their increased their motility and invasive capacity. Increased MM099 invasion was also seen using medium from the siBRG1 silenced MM099 cells showing that it was mediated at least in part through increased expression of secreted cytokines. Indeed, expression of multiple cytokines was upregulated by BRG1 silencing including CSF1, implicated in immune checkpoint therapy resistance [20], CXCL1 [21] and VCAM1 shown to promote immune infiltration and metastases in breast cancer and melanoma [22]. In particular, its function in cross-talk with endothelial cells may be particularly pertinent give their role in promoting melanoma tumour growth [8]. We note that the genes encoding these cytokines were also potently up-regulated in 3D spheroids. Thus, BRG1 acted to repress migration through modulating expression of genes encoding cytokines and other membrane and extracellular proteins.

MES MM099 cells grown as spheroids transit from a MES to a MES/NC intermediate state characterized by up-regulated expression of several NC markers such as *AQP1*, *NOTCH3*, and *NES*, but reduced expression of MES markers such as *AXL*, *AMIGO2*, and *ABCG2*. Karras et al [8] described a pre-EMT NC stem cell population, characterized by expression of the above markers, that drives melanoma tumour growth. In contrast, transition to the MES/NC state was also characterized by upregulation of multiple EMT transcription factors, notably PRRX1 and TCF4, both of which have been shown to play a role in EMT and/or melanoma invasion [8 19], as well as many of their target genes. The MM099 spheroids even in the absence of other cells of

the TME therefore more closely mimic *in vivo* metastatic melanoma cells than cells grown in 2D.

We unexpectedly observed that PRRX1 genome occupancy overlapped with a subset of AP1/TEAD bound MES-specific enhancers. Almost all PRRX1 sites overlapped those of AP1, whereas a large set of AP1 sites did not show PRRX1 binding. We did not observe enrichment of cognate PRRX1 binding motifs at these sites, but as PRRX1 coprecipitated with AP1 and SWI/SNF subunits, this rather suggested that PRRX1 is recruited to the genome indirectly via interactions with AP1 and/or SWI/SNF. Furthermore, AP1 acts as a pioneer factor interacting with SWI/SNF and directing it to remodel and open chromatin at enhancers [23]. Thus, in melanoma cells, a network of interactions between AP1, PRRX1 and SWI/SNF may activate MES enhancers.

AP1 was previously reported to act as a platform for recruitment of other transcription factor such as ZEB1 and YAP1 in breast cancer cells [24], or the glucocorticoid receptor (GR)[25]. GR binds a subset of the sites lacking its cognate binding motifs via interactions with AP1 that as mentioned above acts as a pioneer factor opening chromatin for GR binding[25]. AP1 further interacts with and recruits TCF21 to the genome in smooth muscle cells [26] and with BMAL1 in senescent cells [27]. AP1 is proposed to be the key regulator of MES enhancers [28]. This is not inconsistent with the observations reported here that suggest AP1 may act autonomously, but also as a pioneer factor and platform to integrate other transcription factors such as TEAD4, PRRX1 and possibly others yet to be identified, recruiting them to the genome and activating MES enhancers. This contrasts with MEL melanoma cells, where MITF and SOX10 cooperate to recruit SWI/SNF to activate MEL enhancers [12].

Together our data provide novel insights into regulation of gene expression in MES melanoma cells. Our experiments show that BRG1 binds to MES-specific enhancers and is required for transition from the MES to a MES/NC intermediary state. This transition involves

up-regulated expression of PRRX1 that occupies and activates a subset of MES enhancers via interactions with AP1 and SWI/SNF leading to regulation of genes involved in melanoma metastasis.

Methods.

Cell culture.

Melanoma cell lines 501Mel were grown in RPMI 1640 medium supplemented with 10% foetal calf serum (FCS). MM029, MM047 and MM099 were grown in HAM-F10 medium supplemented with 10% FCS, 5.2 mM GLUTAMAX and 25 mM HEPES. To generate HA-PRRX1 cells, cDNA for the longest protein-coding human PRRX1 transcript (NM_022716.4) was purchased from Genscript and Gateway-cloned into a modified lentiviral vector pInducer20-HA(Cter)-puro. $5x10^5$ low-passage MM099 cells were infected and selected with 3ug/ul puromycin for 3 days before passaging and subsequent downstream analyses. pINDUCER20-mCHERRY-HA was used as a control.

Si/shRNA silencing

SiRNA knockdown experiments were performed with the corresponding ON-TARGET-plus SMARTpools purchased from Dharmacon Inc. (Chicago, Il., USA). SiRNAs were transfected using a 1:2 ratio of Lipofectamine RNAiMax (Invitrogen, La Jolla, CA, USA) for 6 hours. Transfected cells were harvested after 72 hours for subsequent analyses. Lentiviral shRNA vectors were obtained from Sigma (Mission shRNA series) in the PLK0-puro vector. We used shSMARCA4 construct (TRCN0000015549) and lentiviral empty vector as a control and infected $5x10^5$ MM099 cells. Infected cells were maintained under 3ug/ml puromycin selection for 3 days before passaging and subsequent analyses.

Proliferation, viability and senescence analyses by flow cytometry

To assess proliferation after siRNA treatment, cells were stained with Cell Trace Violet (Invitrogen) on the day of transfection and harvested 72 hours later for flow cytometry analysis. To assess cell viability, cells were harvested 72 hours after siRNA transfection and co-stained with Annexin-V and propidium iodide (Biolegend Apoptosis kit, #640914) following manufacturer instructions for flow cytometry analysis. To assess senescence, cells were treated with 100nM bafilomycin A1 for 1hr followed by 2mM C12FDG (Invitrogen) for 2hr before being washed and harvested for flow cytometry analyses. Cells were analysed on a FACS Celesta or Fortessa (BD Biosciences) and data were analysed using Flowjo v6.8.

3D sphere formation and cell quantification

SiRNA-treated cells were harvested, counted and seeded at a density of 1×10^4 in ultra-low attachment, round-bottomed 96-well plates (Corning Costar) for spheroid formation in KO-DMEM medium. For free-floating melanospheres, harvested cells were counted and seeded in KO-DMEM at a density of 2×10^5 on 24-well suspension-culture plates (Greiner Bio-One). At day 7 post-siRNA, 4X magnification images of wells were taken with a phase-contrast microscope. Melanosphere viability was assessed using the CellTiter-Glo Luminescent Cell Viability Assay (Promega, Madison, WI, USA). After addition of 100 µl of CellTiter Glo reagent to each well for 10 min with orbital rotation, luminescence was measured on a BioTek Luminescence microplate reader (using Gen5 software).

RNA preparation, quantitative PCR and RNA-seq analysis

RNA isolation was performed according to standard procedure (Qiagen kit). qRT-PCR was carried out with SYBR Green I (Roche) and SuperScript IV Reverse Transcriptase (Invitrogen) and monitored using a LightCycler 480 (Roche). The mean of *ACTB*, *TBP*, *RPL13A* and *GAPDH* gene expression was used to normalize the results. Primer sequences for each cDNA were designed using Primer3 Software and are available upon request.

RNA-seq was performed essentially as previously described [29] After sequencing raw reads were pre-processed in order to remove adapter and low-quality sequences (Phred quality score below 20) using cutadapt version 1.10 [30] and reads shorter than 40 bases were discarded. Reads were mapped to rRNA sequences using bowtie version 2.2.8 [31] and matching reads removed. Remaining reads were then mapped onto the hg19 assembly of Homo sapiens genome using STAR version 2.5.3a [32]. Gene expression quantification was performed from uniquely aligned reads using htseq-count version 0.6.1p1 [33] , with annotations from Ensembl version 75 and "union" mode. Only non-ambiguously assigned reads were retained for further analyses. Read counts were normalized across samples with the median-of-ratios method. Comparisons of interest were performed using the Wald test for differential expression and implemented in the Bioconductor package DESeq2 version 1.16.1 [34]. Genes with high Cook's distance were filtered out and independent filtering based on the mean of normalized counts was performed. P-values were adjusted for multiple testing using the Benjamini and Hochberg method. Gene set enrichment analyses were done with the GSEA software v3.0 using the Hallmark gene sets of Molecular Signature Database v6.2 [35]. Gene Ontology analysis was performed using the Database for Annotation, Visualization and Integrated Discovery [36]. Heatmap representations of RNA-seq gene expression were generated with the pheatmap R package (v1.0.12).

Raw fastq data for MELO, IGR-39, IGR-37, SK-MEL-24, SK-MEL-5, CJM and MOXIMVI were retrieved from the Cancer Cell Line Encyclopedia (https://sites.broadinstitute.org/ccle/) and processed as described above. To compare the transcriptional signatures of the indicated melanoma cells with signatures described in Tsoi et al., [7], geometric mean expression values for each signature were calculated in the different cell lines. Heatmaps of row-scaled signature values and key genes expression were generated using Prism or by pheatmap R package (v1.0.12).

Gene ontology analyses were performed with the Gene Set Enrichment Analysis software GSEA v3.0 using the hallmark gene sets of the Molecular Signatures Database v6.2 and the functional annotation clustering function of DAVID.

ARCHS4 TF enrichment prediction

To identify the transcription factors potentially driving the gene expression changes upon the 2D-2D transition and in BRG1-silenced melanospheres, the up- and down-regulated genes with criteria log2FC ± 0.5 and p-value ≤ 0.05 were analyzed on the EnrichR platform (Ma'ayan lab, https://maayanlab.cloud/Enrichr/) using the ARCHS4 package (ref: https://doi.org/10.1038/s41467-018-03751-6). The expression of the top 10 TFs were assess in our RNA-seq data.

Protein extraction and Western blotting

Whole cell extracts were prepared by the freeze-thaw technique using LSDB 500 buffer (500 mM KCl, 25 mM Tris at pH 7.9, 10% glycerol (v/v), 0.05% NP-40 (v/v), 16mM DTT, and protease inhibitor cocktail). Cell lysates were subjected to SDS–polyacrylamide gel electrophoresis (SDS-PAGE) and proteins were transferred onto a nitrocellulose membrane. Membranes were incubated with primary antibodies in 5% dry fat milk and 0.01% Tween-20 overnight at 4 °C. The membrane was then incubated with HRP-conjugated secondary antibody (Jackson ImmunoResearch) for 1h at room temperature, and visualized using the ECL detection system (GE Healthcare). Antibodies: MITF (MS-771-P, Interchim), SOX10 (ab155279, Abcam), SOX9 (82630, Cell signaling), BRG1 (ab110641, Abcam), BRM (11966, Cell signaling), BAF170 (A301-038A, Bethyl laboratories), BAF155 (sc-10756, Santa Cruz), BAF53A (ab131272; abcam), ACTB (2D7, IGBMC), BAF47 (91735, Cell signaling), PRRX1 (PA518831, ThermoFisher), VCL (V4505, Sigma-Aldrich).

Wound healing assays

500,000 cells were seeded in 6-well collagen-coated plates before being transfected with siRNAs. 48hrs later (or when maximum confluency was obtained), a total number of 4 scratches were made in the plates using sterile 2ul tips. Pictures were captured on a phase contrast microscope at T0 and each day until cells from one condition achieved total wound healing. Quantification of cell migration was done using ImageJ with the wound healing macro.

Boyden chamber assays

SiRNA-treated cells were harvested, counted and 200,000 single cells were seeded in Boyden chambers (24-well 8um inserts, Corning) in corresponding media without serum. For invasion assays, 100ul of diluted Matrigel (1:20, 356234, Corning) was added in each insert and left to dry for 2hrs at 37°C before being washed twice with PBS and seeded with the 200,000 cells. 24hrs later, migrated cells were fixed using PFA 4% for 10 min before being stained using Crystal violet for 10 min. Excess stain was washed 3 times in PBS before images were captured on a phase contrast microscope. Quantification of migrated cells was done by resuspension of staining using 100mM acetic acid for 15min before luminescence was measured on a BioTek Luminescence microplate reader (using Gen5 software).

Immunoprecipitation

Whole cell extracts were prepared by the freeze-thaw technique using LSDB 500 buffer (500 mM KCl, 25 mM Tris at pH 7.9, 10% glycerol (v/v), 0.05% NP-40 (v/v), 16mM DTT, and protease inhibitor cocktail). Up to 1mg of whole cell extracts were diluted in LSDB without KCl to a final concentration of 150mM KCl and incubated overnight with 5ug of specific antibodies. The next day, 50ul of washed magnetic protein-A/G beads (Dynabeads, Invitrogen) were added to the extracts for 2hr. Beads were washed 3 times in LSDB 300mM KCl, twice in LSDB 150mM and immunoprecipitates were eluted in 100ul of 0.1M glycine pH 2.8 at room temperature for 15min, before addition of 10ul of Tris-HCl pH 8. For SDS-PAGE analysis, 10 to 15ul of eluted proteins were boiled in equal amount of Laemmli buffer before being loaded

on the gels. Antibodies: BRG1 (ab110641, Abcam), PRRX1 (PA518831, ThermoFisher), FOSL2 (sc-447 Santa Cruz), JUN (#9165 Cell Signalling Technology) PBRM1 (#ABE70, Merck Millipore), ARID2 (#82342 Cell Signalling Technology); HA (#H6908, Sigma Aldrich) Control IgG (ab171870, Abcam).

Chromatin immunoprecipitation and sequencing

BRG1 ChIP experiments were performed on native MNase-digested chromatin. 10×10^7 to 20×10^8 freshly harvested MM099 cells were resuspended in 1.5 ml ice-cold hypotonic buffer (0.3M Sucrose, 60 mM KCl, 15 mM NaCl, 5 mM MgCl2, 0.1 mM EDTA, 15 mM Tris–HCl pH 7.5, 0.5 mM DTT, 0.1 mM PMSF, PIC) and cytoplasmic fraction was released by incubation with 1.5 ml of lysis-buffer (0.3M sucrose, 60 mM KCl, 15 mM NaCl, 5 mM MgCl2, 0.1 mM EDTA, 15 mM Tris–HCl pH 7.5, 0.5 mM DTT, 0.1 mM PMSF, PIC, 0.5% (vol/vol) IGEPAL CA-630) for 10 min on ice. The suspension was layered onto a sucrose cushion (1.2 M sucrose, 60 mM KCl, 15 mM NaCl, 5 mM MgCl2, 0.1 mM EDTA, 15 mM Tris–HCl [pH 7.5], 0.5 mM DTT, 0.1 mM PMSF, PIC) and centrifuged for 30 min 4°C at 4700 rpm in a swing rotor. The nuclear pellet was resuspended in digestion buffer (0.32Msucrose, 50 mM Tris–HCl [pH 7.5], 4 mM MgCl2, 1 mM CaCl2, 0.1 mM PMSF) and incubated with 10ul of Micrococcal Nuclease (NEB) for 7 min at 37˚C. The reaction was stopped by addition of 20ul of EDTA 0,5M and suspension chilled on ice for 10 min. The suspension was cleared by centrifugation at 10,000 rpm (4˚C) for 10 min and supernatant (chromatin) was used for further purposes. Chromatin was digested to around 80% of mono-nucleosomes as judged by extraction of the DNA and agarose gel electrophoresis. H3K27ac ChIP experiments were performed on 0.4% PFA-fixed chromatin isolated from MM099 cells according to standard protocols. ChIP-seq libraries were prepared using MicroPlex Library Preparation kit v2 and sequenced on the Illumina Hi-seq 4000 as single-end 50-base reads.

Sequenced reads were mapped to the Homo sapiens genome assembly hg19 using

Bowtie with the following arguments: -m 1 --strata --best -y -S -l 40 -p 2. Peak detection was performed using the MACS software v1.4.3 for transcription factors and v2.1.1 in "broad" mode for chromatin marks [37]. Peaks were annotated with the 'annotatePeaks' command from HOMER v4.11 using the GTF annotation file from ENSEMBL v75. Global clustering analyses with quantitative comparisons and representations of read density heatmaps were performed using seqMINER [38]. Motif discovery from ChIP-seq peak sequences was performed using the RSAT peak-motifs algorithm [39] using JASPAR core non-redundant vertebrates (2020) [40] (http://rsat.sb-roscoff.fr/peak-motifs_form.cgi) for comparing discovered motifs. Visualization of ChIP-seq signal at specific gene loci of interest was achieved using the UCSC genome browser (https://genome.ucsc.edu/). Identification of sequences containing specific binding motifs was done using FiMo [41]. Cut&Tag was performed using the Active Motif CUT&Tag-IT kit and anti-HA 12CA5 antibody following the manufacturer's instructions.

Immunostaining

Cells were seeded at a density of 5 x 10^5 on 4-well chamber slides (Lab-Tek, ThermoFisher) and transfected with siRNAs. 72hr post-siRNA treatment, cells were first fixed in 4% PFA for 10min and then permeabilised with 3 × 5 min 0.1% Triton in PBS, blocked for 1 hr in 5% skim milk in PBS, and incubated overnight in 5% milk with primary antibodies. The following antibodies were used: BRG1 (ab110641, Abcam), BRM (11966, Cell signaling) and ACTB (2D7, IGBMC). Then, cells were washed 3 × 5 min 0.1% Triton in PBS, and incubated with secondary antibodies, Cy3 mouse-anti-rabbit, and Cy5 rabbit-anti-mouse (Invitrogen) for 2 hrs. Cells were subsequently incubated with 1/2000 DAPI nuclear stain for 10 min, washed 3 × 5 min in PBS, dried and mounted with Vectashild. Visualization was done using inverted confocal microscope SP8 UV.

Acknowledgements

We thank Drs JC. Marine and G. Ghanem for the MM029, MM047 and MM099 primary melanoma cells, L Minnoye and S. Aerts for proving the coordinates of MES and MEL enhancers and for the statistics concerning the cross-species conserved elements, all the staff of the IGBMC common facilities in particular cell culture, flow cytometry, confocal microscopy and biomol. This work was supported by institutional grants from the Centre National de la Recherche Scientifique, the Institut National de la Santé et de la Recherche Médicale, the Université de Strasbourg, the Association pour la Recherche contre le Cancer (CR, contract number PJA 20181208268), the Ligue Nationale contre le Cancer, the Institut National du Cancer, the ANR-10-LABX-0030-INRT French state fund through the Agence Nationale de la Recherche under the frame programme Investissements d'Avenir labelled ANR-10-IDEX-0002-02. The IGBMC high throughput sequencing facility is a member of the "France Génomique" consortium (ANR10-INBS-09-08). ID is an 'équipe labellisée' of the Ligue Nationale contre le Cancer. BV was supported by fellowships from the ANR and the Ligue Nationale contre le Cancer.

Author Contributions

BV, AL, AH, performed the experiments; GD, BV performed the bioinformatics analyses; BV, GD and ID conceived the experiments, analysed the data and wrote the paper.

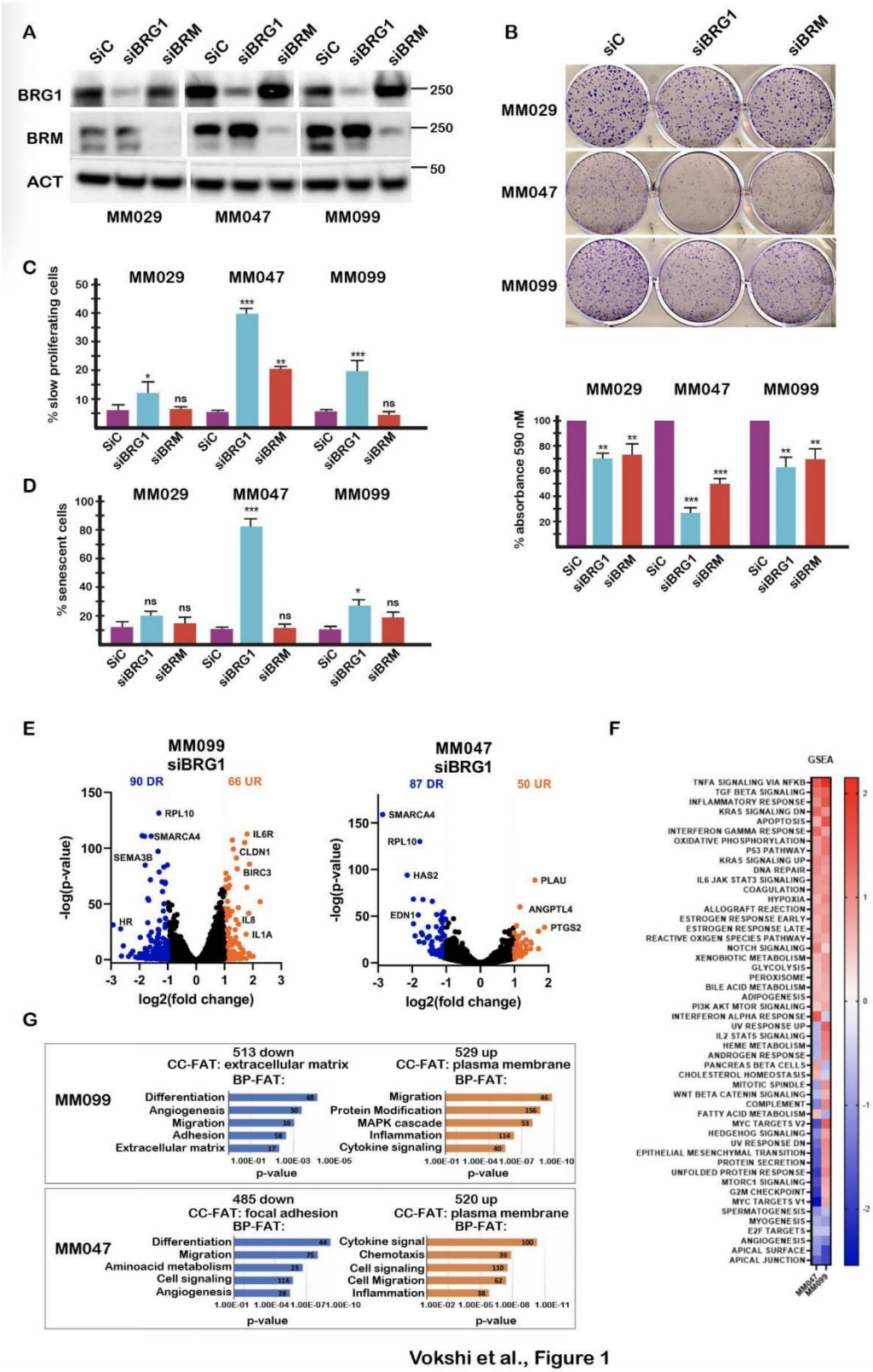
References.

1. Heinz, S. *et al.* Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell* **38**, 576–589 (2010).
2. Arozarena, I. & Wellbrock, C. Phenotype plasticity as enabler of melanoma progression and therapy resistance. *Nat Rev Cancer* **19**, 377–391 (2019).
3. Ennen, M. *et al.* MITF-High and MITF-Low Cells and a Novel Subpopulation Expressing Genes of Both Cell States Contribute to Intra- and Intertumoral Heterogeneity of Primary Melanoma. *Clin Cancer Res* **23**, 7097–7107 (2017).
4. Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
5. Verfaillie, A. *et al.* Decoding the regulatory landscape of melanoma reveals TEADS

as regulators of the invasive cell state. *Nat Commun* **6**, 6683 (2015).
6. Rambow, F. *et al.* Toward Minimal Residual Disease-Directed Therapy in Melanoma. *Cell* **174**, 843-855 e19 (2018).
7. Tsoi, J. *et al.* Multi-stage Differentiation Defines Melanoma Subtypes with Differential Vulnerability to Drug-Induced Iron-Dependent Oxidative Stress. *Cancer Cell* **33**, 890-904 e5 (2018).
8. Karras, P. *et al.* A cellular hierarchy in melanoma uncouples growth and metastasis. *Nature* **610**, 190–198 (2022).
9. Goodall, J. *et al.* Brn-2 represses microphthalmia-associated transcription factor expression and marks a distinct subpopulation of microphthalmia-associated transcription factor-negative melanoma cells. *Cancer Res* **68**, 7788–94 (2008).
10. Rambow, F., Marine, J. C. & Goding, C. R. Melanoma plasticity and phenotypic diversity: therapeutic barriers and opportunities. *Genes Dev* **33**, 1295–1318 (2019).
11. Strub, T. *et al.* Essential role of microphthalmia transcription factor for DNA replication, mitosis and genomic stability in melanoma. *Oncogene* **30**, 2319–32 (2011).
12. Laurette, P. *et al.* Transcription factor MITF and remodeller BRG1 define chromatin organisation at regulatory elements in melanoma cells. *Elife* **10.7554/eLife.06857**, (2015).
13. Laurette, P. *et al.* Chromatin remodellers Brg1 and Bptf are required for normal gene expression and progression of oncogenic Braf-driven mouse melanoma. *Cell Death Differ* **27**, 29–43 (2020).
14. Wouters, J. *et al.* Robust gene expression programs underlie recurrent cell states and phenotype switching in melanoma. *Nat Cell Biol* **22**, 986–998 (2020).
15. Berico, P. *et al.* CDK7 and MITF repress a transcription program involved in survival and drug tolerance in melanoma. *EMBO Rep* e51683 (2021) doi:10.15252/embr.202051683.
16. Minnoye, L. *et al.* Cross-species analysis of enhancer logic using deep learning. *Genome Res* **30**, 1815–1834 (2020).
17. Fontanals-Cirera, B. *et al.* Harnessing BET Inhibitor Sensitivity Reveals AMIGO2 as a Melanoma Survival Gene. *Mol Cell* **68**, 731-744 e9 (2017).
18. Bartual, S. G. *et al.* The citron homology domain as a scaffold for Rho1 signaling. *Proc Natl Acad Sci U S A* **118**, e2110298118 (2021).
19. Pozniak, J. *et al.* A TCF4-dependent gene regulatory network confers resistance to immunotherapy in melanoma. *Cell* **187**, 166-183.e25 (2024).
20. Neubert, N. J. *et al.* T cell–induced CSF1 promotes melanoma resistance to PD1 blockade. *Sci. Transl. Med.* **10**, eaan3311 (2018).
21. Dhawan, P. & Richmond, A. Role of CXCL1 in tumorigenesis of melanoma. *J Leukoc Biol* **72**, 9–18 (2002).
22. Wieland, E. *et al.* Endothelial Notch1 Activity Facilitates Metastasis. *Cancer Cell* **31**, 355–367 (2017).
23. Wolf, B. K. *et al.* Cooperation of chromatin remodeling SWI/SNF complex and pioneer factor AP-1 shapes 3D enhancer landscapes. *Nat Struct Mol Biol* **30**, 10–21 (2023).
24. Feldker, N. *et al.* Genome-wide cooperation of EMT transcription factor ZEB 1 with YAP and AP -1 in breast cancer. *The EMBO Journal* **39**, e103209 (2020).
25. Biddie, S. C. *et al.* Transcription Factor AP1 Potentiates Chromatin Accessibility and Glucocorticoid Receptor Binding. *Molecular Cell* **43**, 145–155 (2011).
26. Zhao, Q. *et al.* TCF21 and AP-1 interact through epigenetic modifications to regulate coronary artery disease gene expression. *Genome Med* **11**, 23 (2019).
27. Jachim, S. K. *et al.* BMAL1 modulates senescence programming via AP-1. *Aging (Albany NY)* **15**, 9984–10009 (2023).
28. Mauduit, D. *et al.* Analysis of long and short enhancers in melanoma cell states. *Elife* **10**, e71735 (2021).

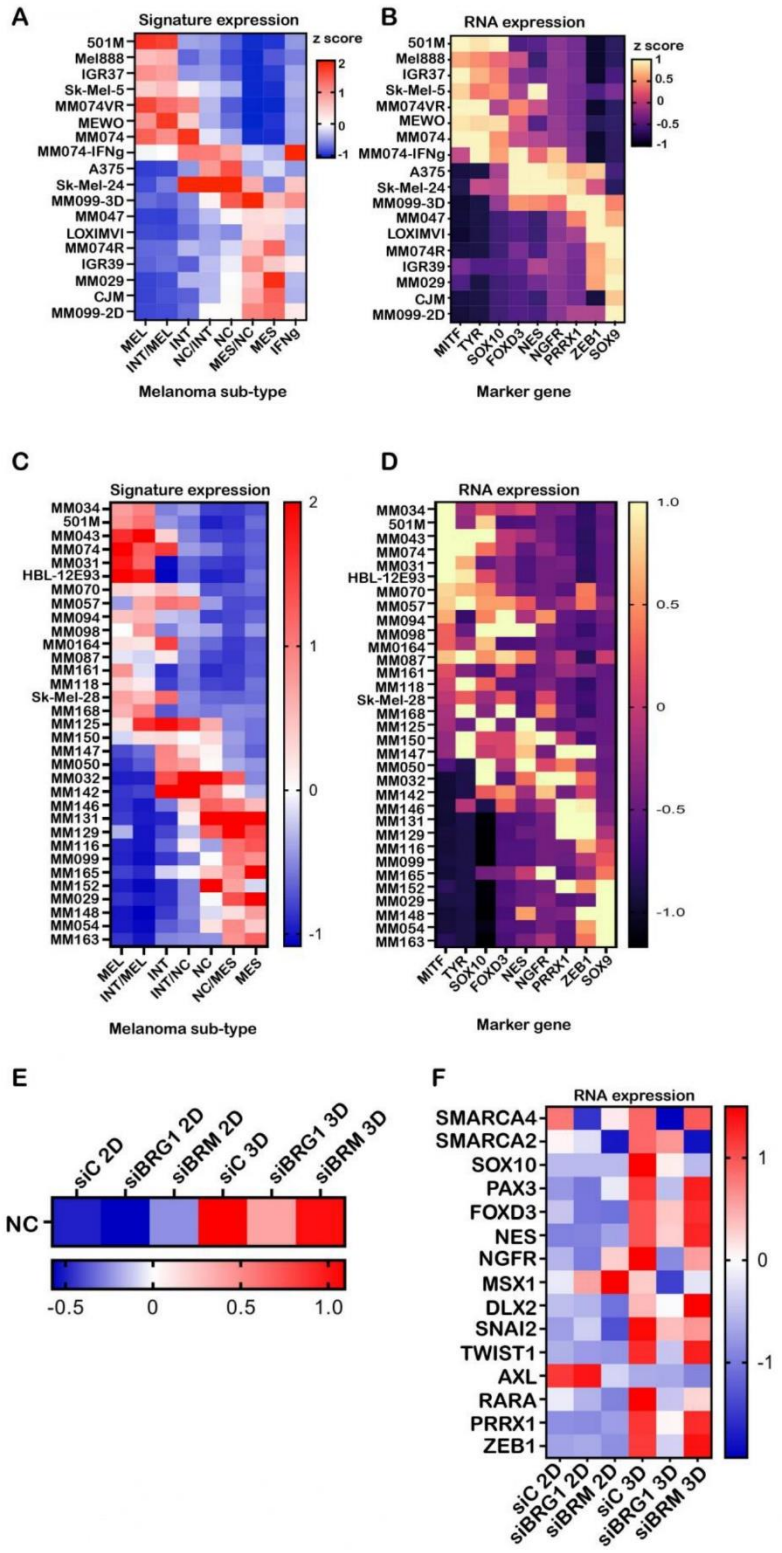
29. Vokshi, B. H. *et al.* SMARCB1 regulates a TFCP2L1-MYC transcriptional switch promoting renal medullary carcinoma transformation and ferroptosis resistance. *Nat Commun* **14**, 3034 (2023).
30. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet j.* **17**, 10 (2011).
31. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25 (2009).
32. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
33. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–9 (2015).
34. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol* **11**, R106 (2010).
35. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550 (2005).
36. Sherman, B. T. *et al.* DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Research* **50**, W216–W221 (2022).
37. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).
38. Ye, T. *et al.* seqMINER: an integrated ChIP-seq data interpretation platform. *Nucleic Acids Res* (2010).
39. Santana-Garcia, W. *et al.* RSAT 2022: regulatory sequence analysis tools. *Nucleic Acids Research* **50**, W670–W676 (2022).
40. Castro-Mondragon, J. A. *et al.* JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Research* **50**, D165–D173 (2022).
41. Bailey, T. L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37**, W202-8 (2009).

Article 4: The chromatin remodeler BRG1 is required for melanoma cell plasticity



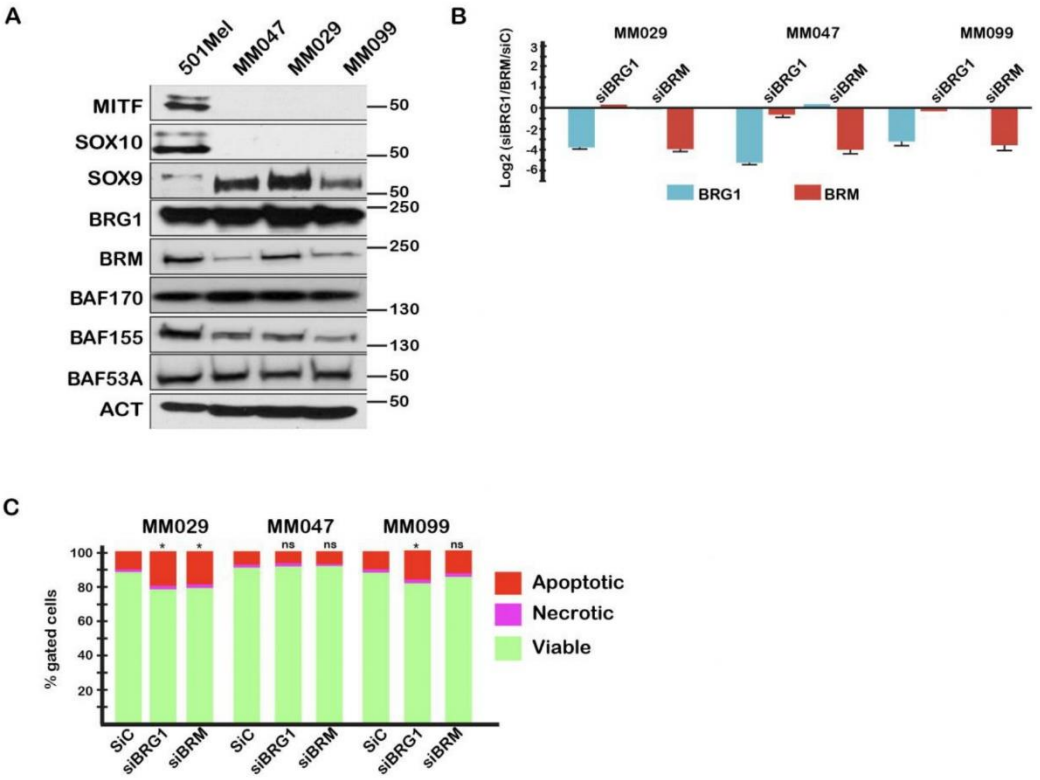
Vokshi et al., Figure 1

Figure 1 BRG1 and BRM are dispensable in fully mesenchymal melanoma cells. **A.** Immunoblots on cell extracts from the indicated cell lines for the indicated proteins after siRNA silencing of BRG1 or BRM. **B**. Colony forming ability of the indicated cell lines after siBRG1 or siBRM silencing. Cells were stained 7 days days after transfection with 0.5% Crystal Violet. Quantitation of the surviving stained cells is shown in the lower panel. **C-D**. The number of slow proliferating (C) or senescent cells (D) were determined by flow cytometry using cell trace violet assay or after labelling with C_{12}FDG respectively. **E**. Impact of BRG1 and BRM silencing on gene expression. Volcano plots showing de-regulated gene expression after siBRG1 silencing in the indicated cell lines. **F**. GSEA analyses BRG1 regulated genes in MM099 and MM047 cells. **G.** David CC-FAT and BP-FAT categories of the up and down regulated genes in each cell type after BRG1 silencing showing the number of genes and the p-values. In all experiments n=3 and unpaired t-test analyses were performed by Prism 5. P-values: *= p<0,05; **= p<0,01; ***= *p<0,001;* data are mean ± SEM.



Vokshi et al., Fig. S1

Figure S1. Expression of cell state signatures and key markers in melanoma cell lines. A. Heatmap comparisons of the gene expression signatures of the indicated cell lines/primary cultures with those defined by Tsoi et al., [7] **B.** Heatmap representation of the expression of the indicated marker genes in each cell line. **C-D**. Heatmap comparisons of the gene expression signatures of the collection of cell cultures analysed by Wouters et al [14] and expression of the indicated marker genes in each culture. **E**. Heatmap representation of the expression of the indicated marker genes in MM099 cells under the different conditions.



Vokshi et al., Fig. S2

Figure S2. Impact of BRG1 and BRM silencing in mesenchymal melanoma cells. A. Immunoblot showing the expression of the indicated proteins in each cell line/primary culture.

Molecular mass markers are shown on the right in kDa. **B.** RT-qPCR quantification of BRG1 or BRM expression in each cell lines under the indicated conditions expressed as Log2-Fold change relative to the siRNA control. **C.** Apoptosis was quantified by flow cytometry using Annexin-V 72 hours after siBRG1 or siBRM silencing. In all experiments n=3 and unpaired t-test analyses were performed by Prism 5. P-values: *= p<0.05; **= p<0.01; ***= *p<0.001*; data are mean ± SEM.

A

Log2 FC +/- 1 p<0.05	siBRG1		siBRM	
	Up	Down	Up	Down
MM047	50	87	0	8
MM099	66	90	6	20
Log2 FC +/- 0.5 p<0.01				
MM047	520	485	50	114
MM099	529	513	39	94

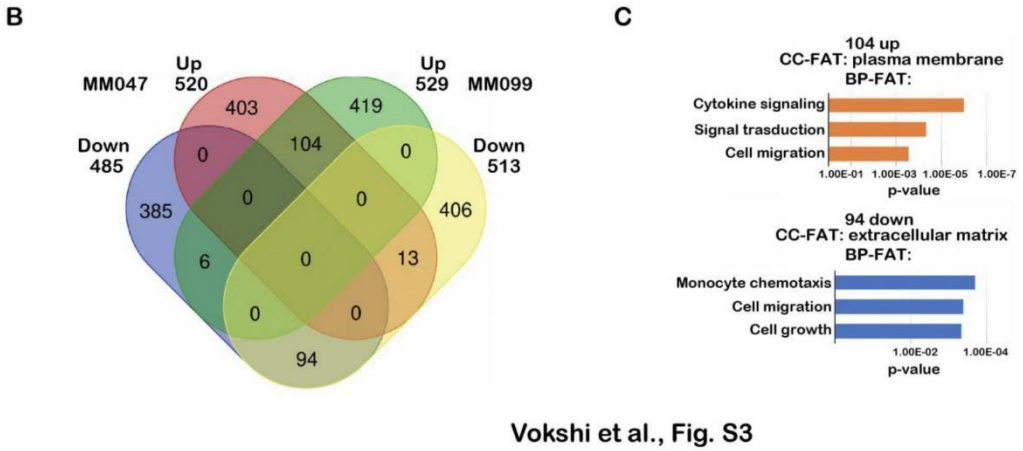


Figure S3. BRG1 is a transcriptional modulator of cytokine signalling in mesenchymal melanoma cells. **A**. Table showing the number of deregulated genes using the indicated

criteria. **B**. Venn diagram illustrating the overlap between gene expression changes in each cell type. **C**. DAVID ontology analyses of the common up- or downregulated genes.

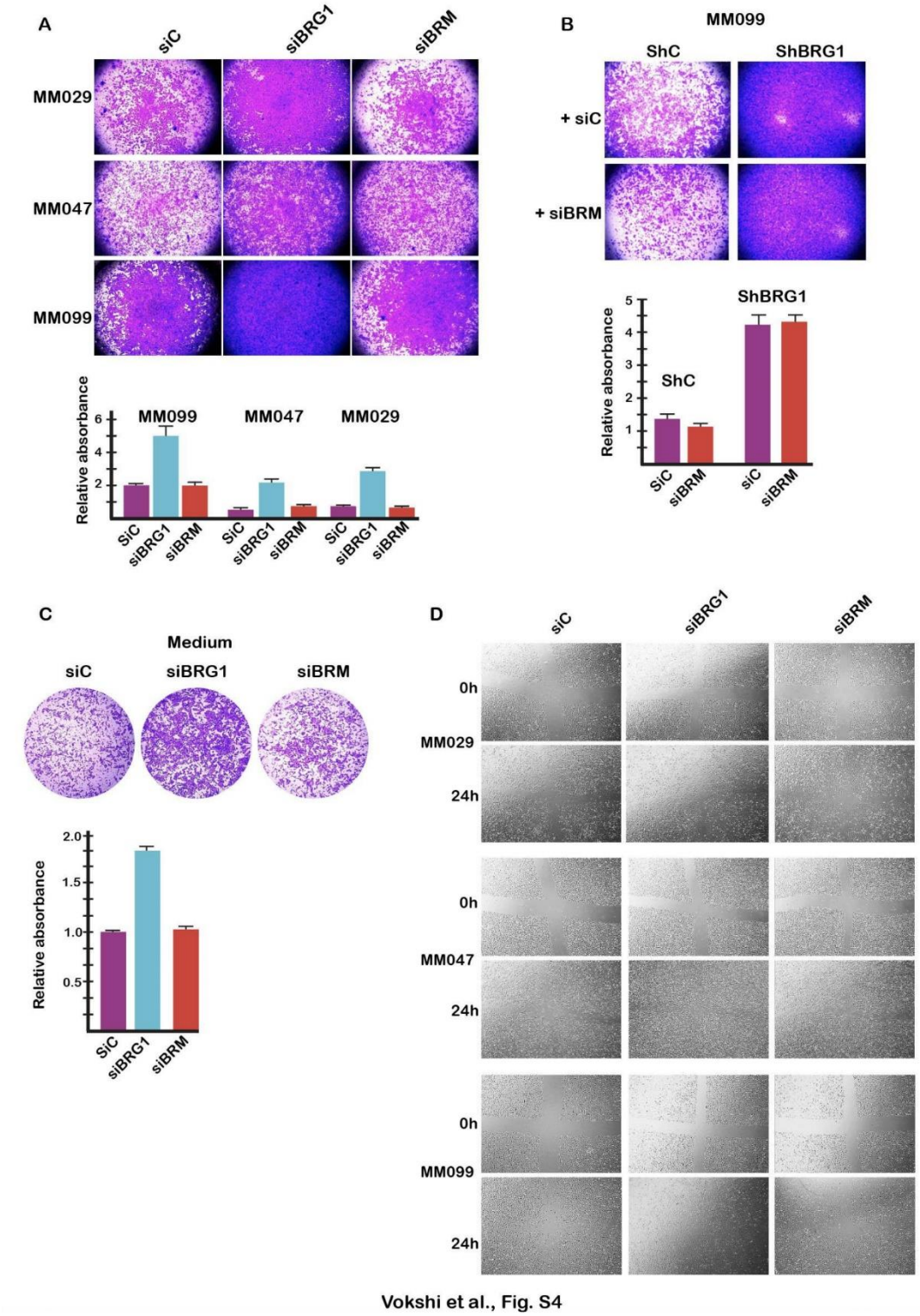
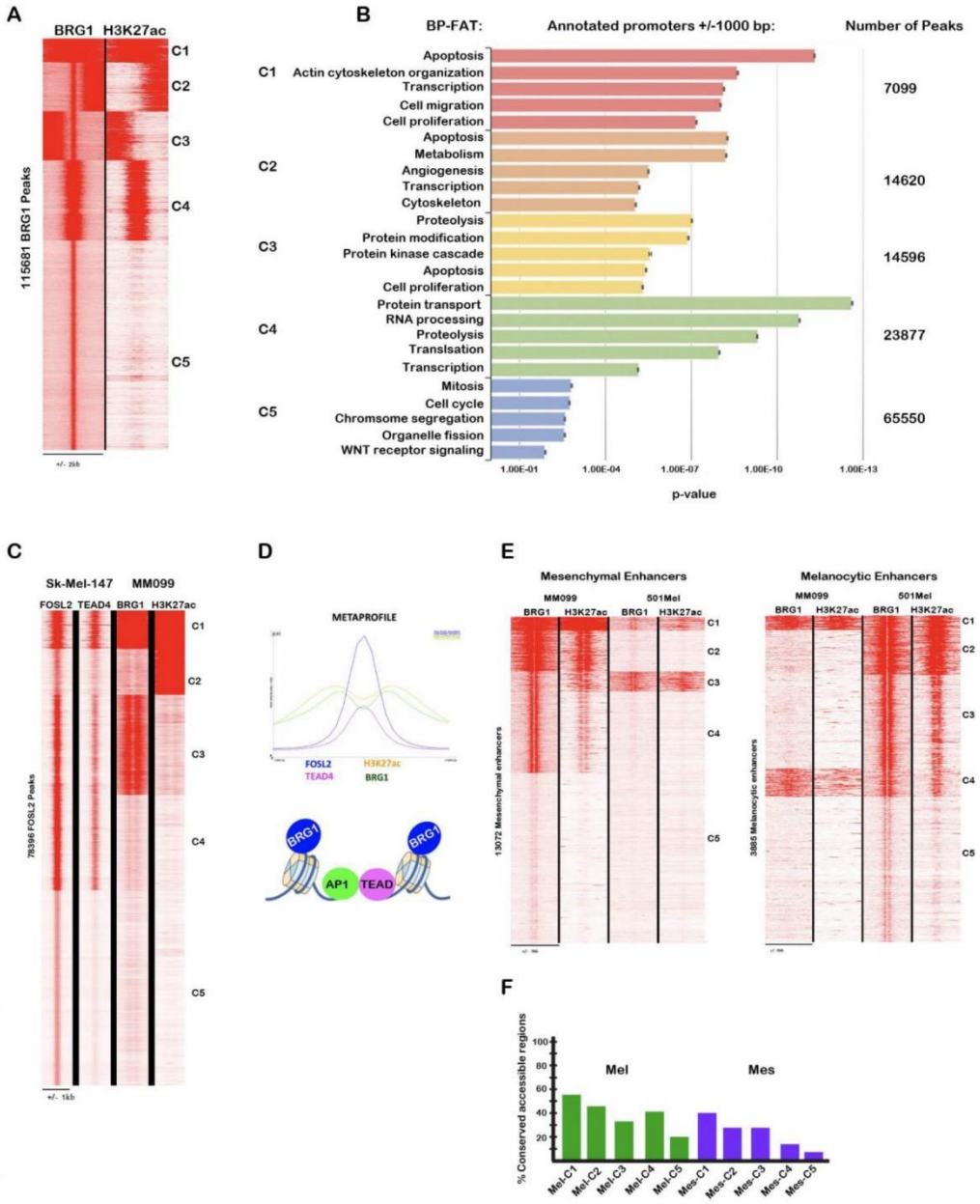


Figure S4. BRG1 regulates melanoma cell motility and invasion. **A**. Boyden chamber assay showing cells that crossed the matrigel barrier after siBRG1 or siBRM silencing compared to

control siRNA. **B**. Boyden chamber assay showing cells that crossed the matrigel barrier after shBRG1 silencing compared to control shRNA in the presence or absence of concomitant siBRM silencing. **C.** Boyden chamber assay showing cells that crossed the matrigel barrier after incubation with conditioned medium from cells transfected with the indicated siRNAs. **D**. Wound healing assay illustrating increased motility of siBRG1 silenced cells. In all experiments n=3 and unpaired t-test analyses were performed by Prism 5. P-values: *= $p<0.05$; **= $p<0.01$; ***= $p<0.001$; data are mean ± SEM.



Vokshi et al., Fig. 2

Figure 2. BRG1 genome occupancy in MM099 cells. **A**. Read density map showing comparison of BRG1 bound nucleosomes and H3K27ac modified nucleosomes in MM099 cells. **B**. For each category, the BRG1 bound regions corresponding to proximal promoters were identified and their ontology determined as DAVID BP-FAT. **C**. Read density map

showing comparison of FOSL2 and TEAD4 bound sites in Sk-Mel-147 cells with BRG1 and H3K27ac in MM099 cells. **D.** The meta-profile of cluster C1 shows that AP1 and TEAD factors bind to regulatory elements flanked by BRG1 bound and H3K27ac-marked nucleosomes. **E**. Read density maps showing BRG1 occupancy of nucleosomes surrounding the MES and MEL specific enhancers defined by Minnoye et al., [16] in MM099 and 501Mel cells illustrating the selective occupancy of these sites in each cell type. **F**. The % of cross-species conserved accessible regions present in each of the clusters shown in the heatmaps of panels D and E.

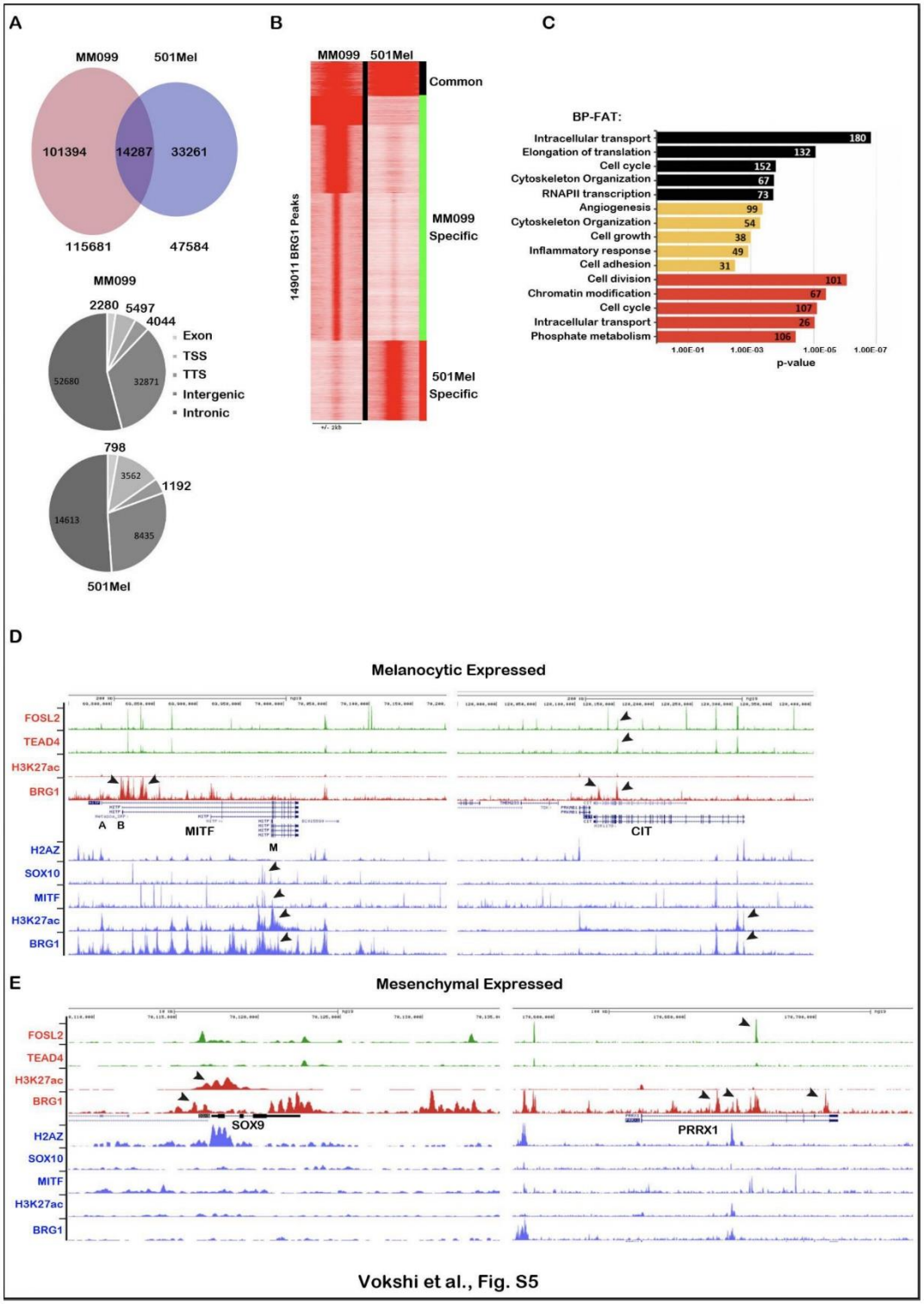
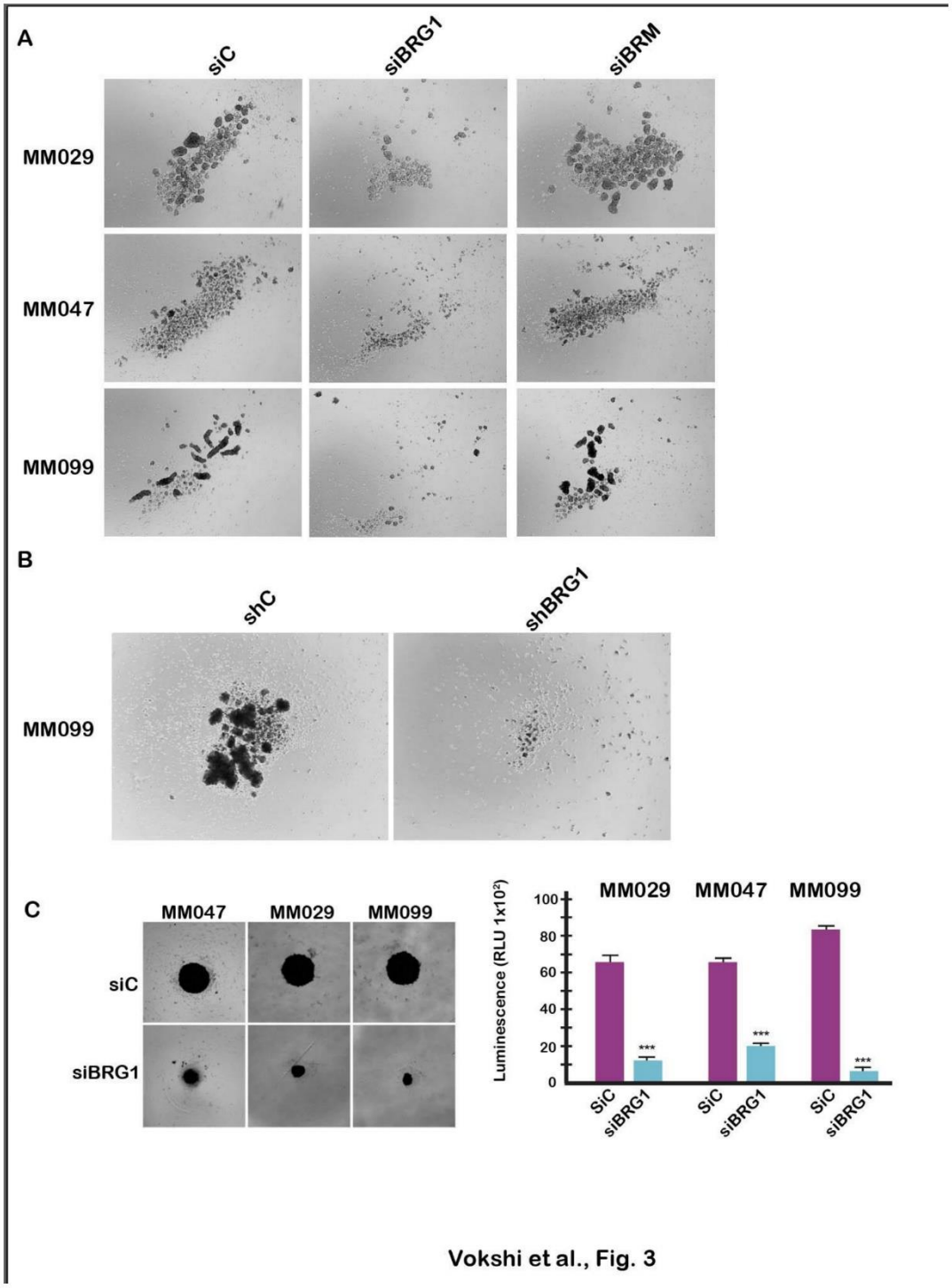

A
MM099
501Mel
101394
14287
33261
115681
47584
MM099
2280
5497
4044
Exon
TSS
TTS
Intergenic
Intronic
52680
32871
798
3562
1192
14613
8435
501Mel
B
MM099
501Mel
Common
MM099
Specific
501Mel
Specific
149011 BRG1 Peaks
C
BP-FAT:
Intracellular transport
Elongation of translation
Cell cycle
Cytoskeleton Organization
RNAPII transcription
Angiogenesis
Cytoskeleton Organization
Cell growth
Inflammatory response
Cell adhesion
Cell division
Chromatin modification
Cell cycle
Intracellular transport
Phosphate metabolism
180
132
152
67
73
99
54
38
49
31
101
67
107
26
106
1.00E-01
1.00E-03
1.00E-05
1.00E-07
p-value
D
Melanocytic Expressed
FOSL2
TEAD4
H3K27ac
BRG1
A B
MITF
M
CIT
H2AZ
SOX10
MITF
H3K27ac
BRG1
E
Mesenchymal Expressed
FOSL2
TEAD4
H3K27ac
BRG1
SOX9
PRRX1
H2AZ
SOX10
MITF
H3K27ac
BRG1
Vokshi et al., Fig. S5

Figure S5. Comparison of BRG1 genome occupancy in MM099 and 501 Mel cells. **A**. Venn diagrams showing overlap between BRG1-bound sites in 501Mel and MM099 cells (top panel) as well as pie charts illustrating the distribution of sites with respect to defined genomic regions (middle and bottom panels). **B**. Read density map showing comparison of BRG1 bound nucleosomes in MM099 and 501Mel cells. **C**. DAVID BP-FAT ontology analyses of the genes closest to the sites in each category. UCSC screenshots of specific loci showing the indicated ChIP-seq tracks in 501Mel cells in blue and MM099 cells in red. **D.** Gene loci (*MITF* and *CIT*) showing preferential BRG1 occupancy in 501Mel cells. The transcriptional start sites of the A, B and M MITF isoforms are indicated. Arrow heads show occupied sites around the M isoform start site in 501Mel cells absent in MM099 cells and occupied sites around the A-B isoform start sites in MM099 cells absent in 501Mel cells. At the *CIT* locus, arrow heads show occupied sites around the start site in 501Mel cells absent in MM099 cells and intronic occupied sites in MM099 cells absent in 501Mel cells. **E.** Gene loci (*SOX9, PRRX1*) showing preferential BRG1 occupancy in MM099 cells. Arrow heads show occupied sites in MM099 absent in 501Mel.



Vokshi et al., Fig. 3

Figure 3. BRG1 is required for MES cell spheroid growth. **A.** Bright field images of free growing spheroids 7 days after silencing of BRG1 or BRM compared to control siRNA. **B**. Bright field images of free growing spheroids of shBRG1 or shC cells after 7 days. **C.** Bright field images of single spheroids 7 days after silencing of BRG1 compared to control siRNA. Right panel shows quantification of ATP levels as a measure of cell number in the spheroids in the different conditions. n=3 with 4 technical replicates for each N. Unpaired t-test analysis were performed by Prism 5. P-values: *= p<0,05; **= p<0,01; ***= p<0,001; data are mean ± SEM.

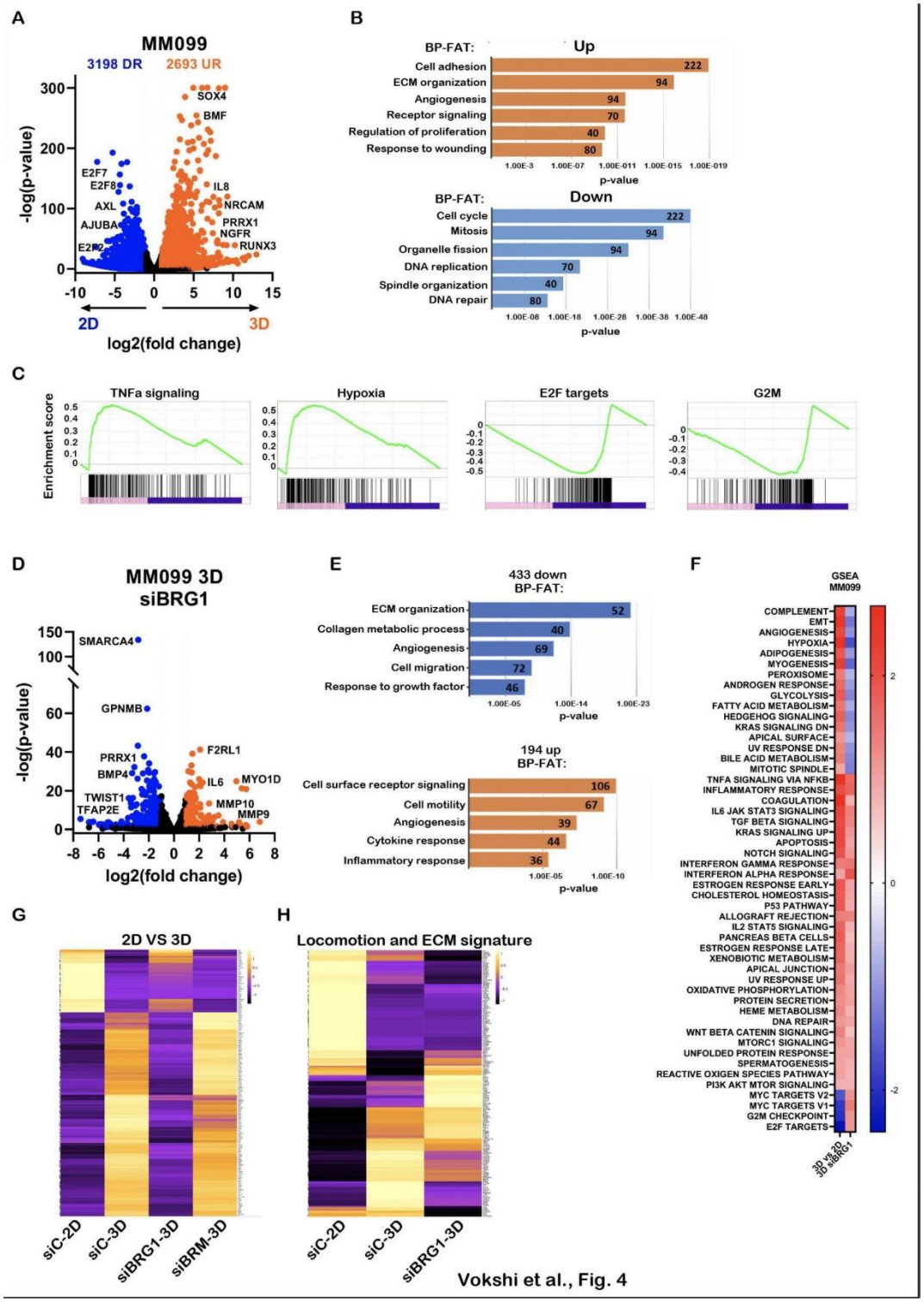
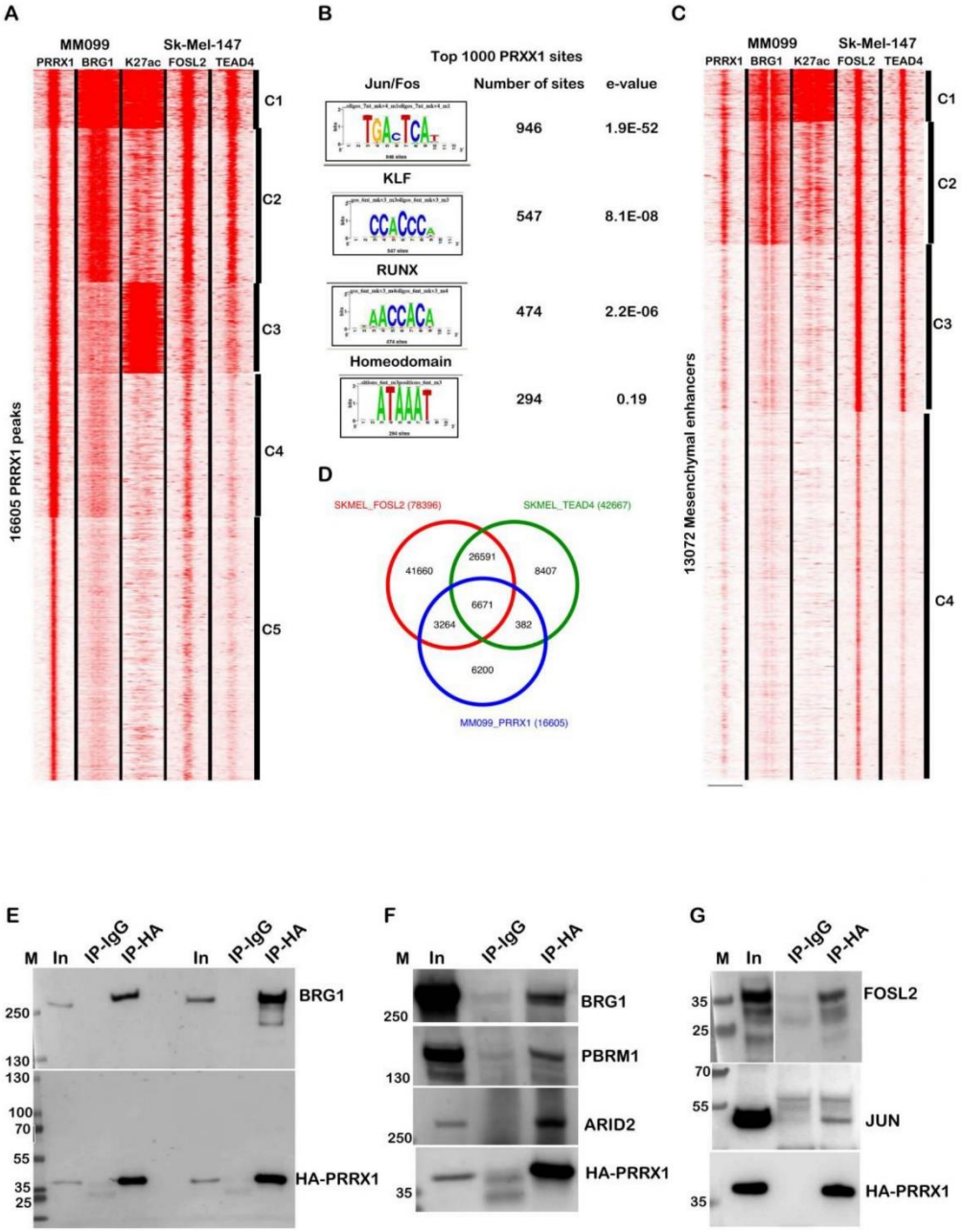
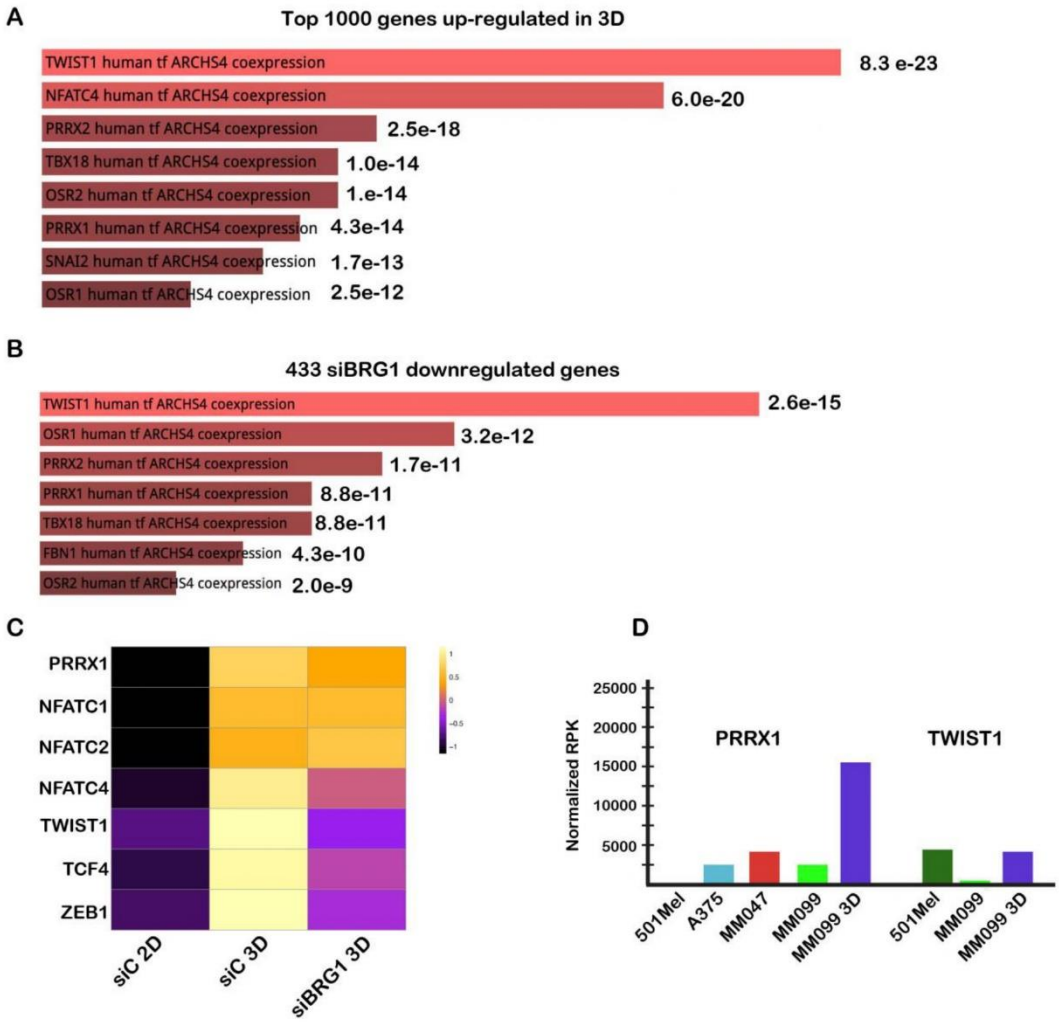

A
MM099
3198 DR
2693 UR
-log(p-value)
log2(fold change)
2D
3D
SOX4
BMF
IL8
NRCAM
PRRX1
NGFR
RUNX3
E2F7
E2F8
AXL
AJUBA
B
BP-FAT:
Up
Cell adhesion
ECM organization
Angiogenesis
Receptor signaling
Regulation of proliferation
Response to wounding
p-value
Down
Cell cycle
Mitosis
Organelle fission
DNA replication
Spindle organization
DNA repair
C
TNFa signaling
Hypoxia
E2F targets
G2M
Enrichment score
D
MM099 3D
siBRG1
SMARCA4
GPNMB
PRRX1
BMP4
TWIST1
TFAP2E
F2RL1
IL6
MYO1D
MMP10
MMP9
E
433 down
BP-FAT:
ECM organization
Collagen metabolic process
Angiogenesis
Cell migration
Response to growth factor
194 up
BP-FAT:
Cell surface receptor signaling
Cell motility
Angiogenesis
Cytokine response
Inflammatory response
F
GSEA
MM099
COMPLEMENT
EMT
ANGIOGENESIS
HYPOXIA
ADIPOGENESIS
MYOGENESIS
PEROXISOME
ANDROGEN RESPONSE
GLYCOLYSIS
FATTY ACID METABOLISM
HEDGEHOG SIGNALING
KRAS SIGNALING DN
APICAL SURFACE
UV RESPONSE DN
BILE ACID METABOLISM
MITOTIC SPINDLE
TNFA SIGNALING VIA NFKB
INFLAMMATORY RESPONSE
COAGULATION
IL6 JAK STAT3 SIGNALING
TGF BETA SIGNALING
KRAS SIGNALING UP
APOPTOSIS
NOTCH SIGNALING
INTERFERON GAMMA RESPONSE
INTERFERON ALPHA RESPONSE
ESTROGEN RESPONSE EARLY
CHOLESTEROL HOMEOSTASIS
P53 PATHWAY
ALLOGRAFT REJECTION
IL2 STAT5 SIGNALING
PANCREAS BETA CELLS
ESTROGEN RESPONSE LATE
XENOBIOTIC METABOLISM
APICAL JUNCTION
UV RESPONSE UP
OXIDATIVE PHOSPHORYLATION
PROTEIN SECRETION
HEME METABOLISM
DNA REPAIR
WNT BETA CATENIN SIGNALING
MTORC1 SIGNALING
UNFOLDED PROTEIN RESPONSE
SPERMATOGENESIS
REACTIVE OXIGEN SPECIES PATHWAY
PI3K AKT MTOR SIGNALING
MYC TARGETS V2
MYC TARGETS V1
G2M CHECKPOINT
E2F TARGETS
3D vs 2D
3D siBRG1
G
2D VS 3D
siC-2D
siC-3D
siBRG1-3D
siBRM-3D
H
Locomotion and ECM signature
siC-2D
siC-3D
siBRG1-3D
Vokshi et al., Fig. 4

Figure 4. Gene expression changes associated with transition from 2D to 3D growth. **A**. Volcano plot showing de-regulated gene expression of control silenced cells in 2D and 3D conditions. **B.** David BP-FAT categories showing the number of genes and the p-values. **C.** GSEA analyses of deregulated genes in 3D growth. **D**. Volcano plot showing de-regulated gene expression in BRG1 silenced spheroids. **E.** David BP-FAT categories of BRG1-regulated genes showing the number of genes and the p-values. **F.** GSEA analyses of BRG1-regulated genes. **G.** Heatmap showing the normal and de-regulated expression of the genes identified in the Venn diagram upon the transition from 2D to 3D growth with siBRM shown as additional control. **H**. Heatmap showing the expression of the Locomotion and ECM signature genes of Verfaillie et al., [5] upon the transition from 2D to 3D growth with or without BRG1 silencing.



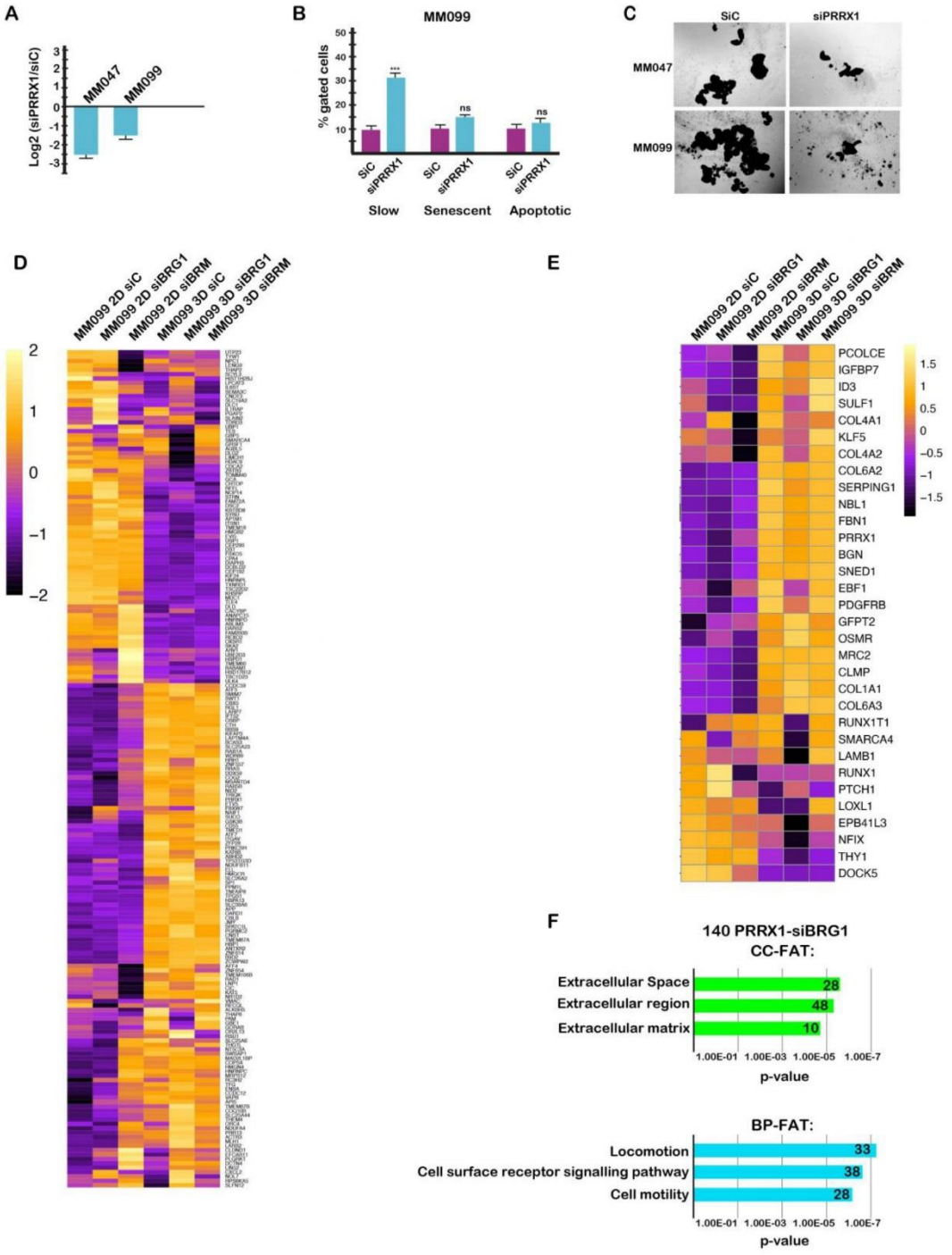
Vokshi et al., Fig. 5

Figure 5. Profiling of PRRX1 genomic occupancy. **A**. Read density heatmap of PRRX1 Cut&Tag in MM099 cells integrated with H3K27ac and FOSL2 and TEAD4 in Sk-Mel-147 cells. RSAT analyses of the top 1000 PRRX1-bound sites showing enrichment in AP1 motifs, but no significant enrichment of Homeodomain motifs. **C.** Read density heatmap of PRRX1 H3K27ac, FOSL2 and TEAD4 at the collection of MES enhancers. **D**. Intersect overlap of PRRX1 bound sites with FOSL2 and TEAD4. **E-G**. Immunoblots showing immunoprecipitation of HA-tagged PRRX1 from MM099 cells. Duplicate HA IPs and corresponding IgG controls were revealed with antibody against PRXX1 (bottom panel E- or BRG1 (upper panel E). Each IP was then rerun on additional gels and revealed with the indicated antibodies in panels F and G (bottom panel G was revealed with HA antibody).



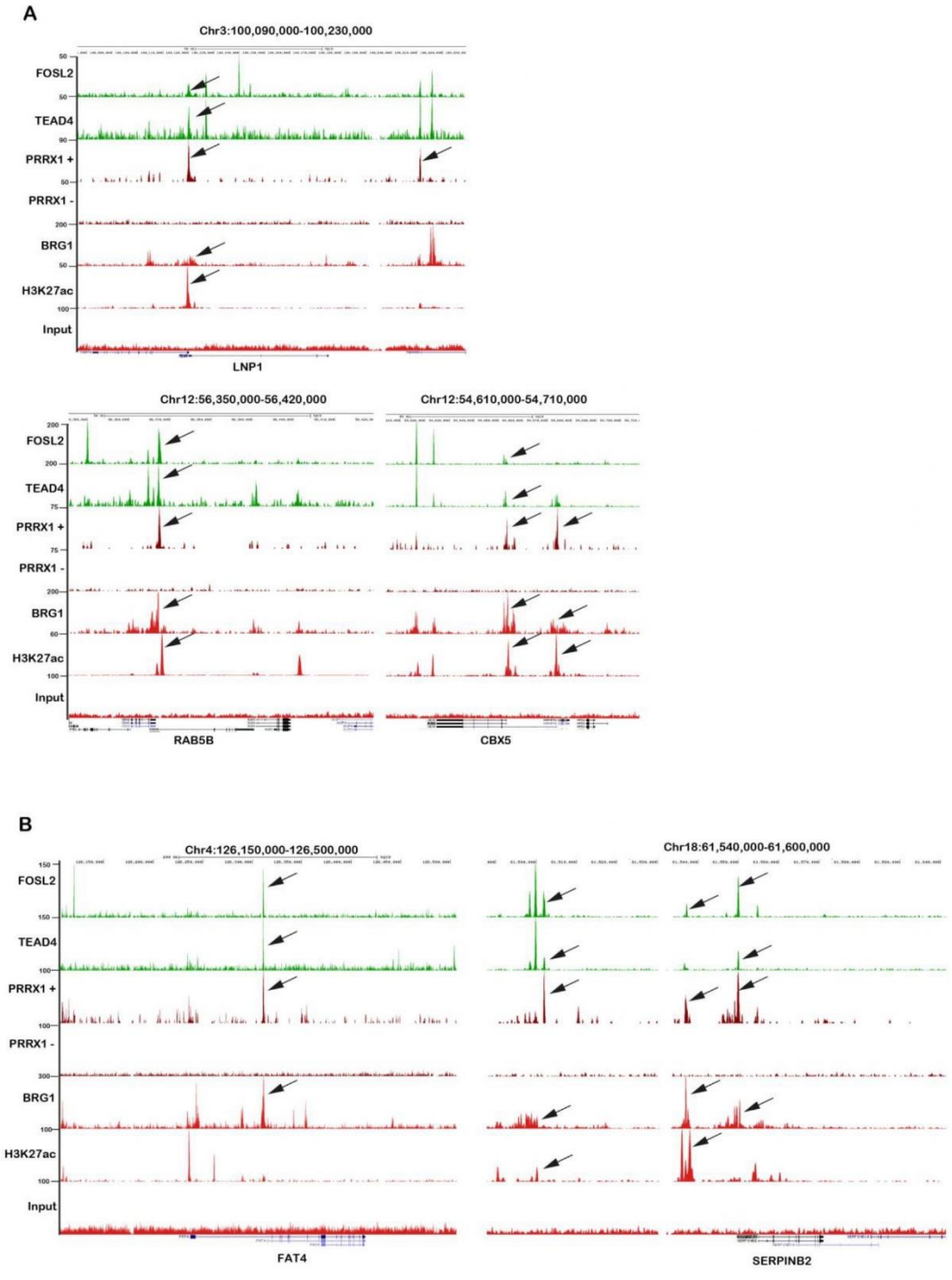
Vokshi et al., Fig. S6

Figure S6. Transcription factors involved in spheroid growth. **A-B**. The 1000 most upregulated genes upon the 2D-3D transition or those down-regulated by BRG1 silencing in 3D spheroids were analysed using EnrichR. The ARCHS4 TF co-expression category identified several transcription factors potentially involved in their regulation. **C**. Heatmap showing the relative expression of selected transcription factors in 2D vs 3D and following siBRG1 as a Z score. **D**. PRRX1 and TWIST1 expression assessed by RNA-seq in the indicated cell lines.



Vokshi et al., Fig. 6

Figure 6. PRXX1 is required for gene regulation involved in MES 3D spheroid growth. **A**. RT-qPCR illustrating PRRX1 knockdown in each cell type. **B** Proportion of slow growing senescent and apoptotic MM0099 cells following PRRX1 silencing assessed as described in Fig. 1. **C**. Bright field images of free growing spheroids following PRRX1 silencing. In all experiments n=3 and unpaired t-test analyses were performed by Prism 5. P-values: *= $p<0.05$; **= $p<0.01$; ***= $p<0.001$; data are mean ± SEM. **D**. Heatmap showing changes in expression of the top 200 genes with promoter proximal PRRX1 binding in MM0099 cells under the indicated conditions. **E.** Heatmap showing changes in expression of the genes defined as the PRRX1 regulon in Karras et al., [8] in MM0099 cells under the indicated conditions. **F**. DAVID gene ontology analyses of the 140 siBRG1 regulated genes with PRRX1 binding sites.



Vokshi et al., Fig. S7

Figure S7. Binding of PRRX1 at selected target genes. **A.** UCSC screenshots of representative genes up-regulated upon 3D growth with PRRX1 binding sites in their proximal promoter. The ChIP-seq data for the other indicated factors is also shown illustrating PRRX1 overlap with FOSL2, TEAD4, BRG1 and H3K27ac. **B.** Genome tracks of the *FAT4* loci, a gene down-regulated by siBRG1 and the *SERPINB2* loci, a gene upregulated upon siBRG1, in 3D conditions. PRRX1 frequently overlaps with FOSL2, TEAD4, BRG1 and H3K27ac both at promoter proximal and promoter distal sites. In panels A and B, arrows indicate representative peaks where these overlaps are observed.

Supplementary Dataset 1. Summary of RNA-seq results following BRG1 silencing in MM047 or MM099 cells. Shown are gene names, description, fold change, p-value and adjusted p-value. As indicated, other pages on the spreadsheet show the ontology analyses of each gene set.

Supplementary Dataset 2. Summary of RNA-seq results following BRG1 silencing in 3D-grown spheres of MM099 cells. Shown are gene names, description, fold change, p-value and adjusted p-value. As indicated, other pages on the spreadsheet show the ontology analyses of each gene set.

Supplementary Dataset 3. List of genes identified as essential to 3D growth and regulated by BRG1 in MM099 cells. Shown are gene names, description, fold change, p-value and adjusted p-value.

Supplementary Dataset 4. List of genes identified as essential to locomotion and ECM organization and regulated by TEAD factors in MM047 cells (described in Verfaillie *et al.*, 2015). Shown are gene names, description, fold change, p-value and adjusted p-value.

Alexandre HALLER

Characterization of long non-coding RNA LENT, a potential therapeutic target in cutaneous melanoma

Résumé

Le mélanome est le plus agressif des cancers de la peau, représentant 1 % des cancers cutanés mais responsable de la plupart des décès liés à ces cancers. Les mélanomes métastatiques sont traités par immunothérapie ou par une inhibition ciblée de MAPK. Néanmoins, la résistance primaire ou acquise met les chercheurs au défi de trouver de nouvelles thérapies. Dans ce contexte, le laboratoire d'accueil a identifié une série de longs ARNs non-codants (LncRNA) spécifiques du mélanome.

Mon projet concerne le lncRNA LENT (LncRNA Enhancer of Translation) fortement exprimé dans les mélanomes par rapport aux autres cancers ou tissus. LENT est régulé par le facteur de transcription MITF et exprimé dans les cellules de mélanome mélanocytiques. Le silencing de LENT inhibe la prolifération des mélanomes et induit l'apoptose. La purification de LENT couplée à la spectrométrie de masse a révélé une interaction sélective avec la résolvase de G quadruplex DHX36 qui régule la traduction d'ARNm et se localise aux mitochondries dans les cellules de mélanome. La délétion de LENT module l'association de nombreux ARNm avec DHX36 et les polysomes, modulant leur traduction. Cette délétion promeut la mitophagie et réduit la capacité de réponse au stress des cellules de mélanome, entrainant leur mort. LENT représente ainsi une nouvelle cible thérapeutique pour traiter le mélanome cutané.

Mots-clés : Mélanome, Long ARNs non-codants, Mitochondries, Traduction, G-quadruplex

Résumé en anglais

Melanoma is the most aggressive form of skin cancers, accounting for 1% of all cutaneous cancers, but is responsible for the majority of deaths from these cancers. Metastatic melanoma is treated with immunotherapy or targeted MAPK inhibition. However, primary or acquired resistance is challenging researchers to find new therapies. In this context, the host laboratory has identified a series of melanoma-specific long non-coding RNAs (lncRNAs).

My project concerns the lncRNA LENT (LncRNA Enhancer of Translation), which is highly expressed in melanoma compared with other cancers or normal tissues. LENT is regulated by the transcription factor MITF and expressed in melanocytic melanoma cells. Silencing of LENT inhibits melanoma proliferation and induces apoptosis. Purification of LENT coupled to mass spectrometry revealed a selective interaction with the G quadruplex resolvase DHX36 which regulates mRNA translation and localizes to mitochondria in melanoma cells. LENT deletion modulates the association of many mRNAs with DHX36 and polysomes, modulating their translation. This deletion promotes mitophagy and reduces the stress response capacity of melanoma cells, leading to their death. LENT therefore represents a new therapeutic target for treating cutaneous melanoma.

Keywords: Melanoma, Long noncoding RNAs, Mitochondria, Translation, G-quadruplex