



Functional genomics characterization of genetic predisposing factors in uveal melanoma

Anne-Céline Derrien

► To cite this version:

Anne-Céline Derrien. Functional genomics characterization of genetic predisposing factors in uveal melanoma. Biotechnology. Université Paris sciences et lettres, 2021. English. NNT : 2021UPSLS083 . tel-03561357

HAL Id: tel-03561357

<https://pastel.hal.science/tel-03561357>

Submitted on 8 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à l'Institut Curie

Functional genomics characterization of genetic predisposing factors in uveal melanoma

Génomique fonctionnelle des prédispositions génétiques du mélanome uvéal

Soutenue par

Anne-Céline DERRIEN

Le 10 novembre 2021

Ecole doctorale n° 561

**Hématologie, oncogénèse
et biothérapies**

Spécialité

Génétique



Composition du jury :

Robert BALLOTTI Directeur de recherche DRE, Inserm U1065	<i>Président</i>
Anne BOWCOCK Professor, Mount Sinai School of Medicine	<i>Rapporteur</i>
Laufey AMUNDADOTTIR Senior Investigator, Head of Lab, NCI, NIH	<i>Rapporteur</i>
Sergio ROMAN-ROMAN Head of Translational Research Department, Institut Curie	<i>Examineur</i>
Marc-Henri STERN Directeur de recherche, Inserm U830	<i>Directeur de thèse</i>

La résilience, c'est l'art de naviguer dans les torrents.

Boris Cyrulnik

C'est l'audace de l'être humain, cette si petite chose orgueilleuse, c'est le désir de savoir, la soif d'apprendre, le besoin de découvrir. C'est l'amour de l'inconnu, la perspective de trembler, de rire, d'être découragé, rassuré, de tout perdre ou de tout gagner, de chercher, sans fin, la réponse aux questions que l'on se pose. De vivre.

Alexis Michalik

Le succès, c'est d'aller d'échec en échec sans perdre son enthousiasme.

Winston Churchill

REMERCIEMENTS

Les travaux ayant abouti à la rédaction de cette thèse sont le fruit de quatre années qui m'auront vu évoluer tant professionnellement que personnellement, ce qui n'aurait pu être possible sans le soutien, la participation et l'amitié de nombreuses personnes.

Je remercie tout d'abord l'U830, son directeur Olivier Delattre ainsi que tous ses membres, pour leur accueil chaleureux au sein de cette unité de recherche dont j'ai été fière de faire partie. Merci également à Carole Drique et Keila Risal, pour leur rôle crucial dans la gestion de nos projets de recherche, ainsi que pour leur gentillesse sans faille.

Je remercie tous les membres de mes comités de thèse, qui auront à travers leurs idées et nos discussions participé au façonnement de ce projet, notamment le Dr Josh Waterfall, le Dr Josselin Noirel, le Dr Raphael Margueron et le Dr Olivier Delattre pour tous leurs conseils et leur expertise. De la même manière, je tiens à remercier chaleureusement ceux qui ont accepté de faire partie de mon jury de thèse, une reconnaissance toute particulière allant à la Professeure Anne Bowcock et au Dr Laufey Amundadottir pour avoir accepté le rôle de rapporteurs de cette thèse, ainsi qu'au Dr Robert Ballotti, Président du jury, et au Dr Sergio Roman-Roman, qui en sont les examinateurs.

Merci à la Cellule Enseignement de l'Institut Curie, au programme Européen H2020-MSCA IC-3i-PhD Cofund et à la Ligue Contre le Cancer, pour leur recrutement et leur soutien pendant 4 ans.

Pour avoir été mon mentor, pour avoir conçu ce projet, l'avoir suivi de près avec vif intérêt, et pour m'avoir donné toute sa confiance, mes remerciements les plus sincères se dirigent à Marc-Henri. Ton soutien sans faille depuis mes débuts au laboratoire, ton expertise, ta curiosité et ton enthousiasme scientifiques auront été essentiels à la réalisation de mes travaux. Merci d'avoir à la fois suivi de près ce projet, l'agréant de conseils techniques et de mises en perspective enrichissantes, mais aussi de m'avoir accordé liberté et confiance dans mes travaux au quotidien et dans les résultats émanant de ceux-ci, car cette liberté m'a apporté autonomie, confiance et indépendance.

A tous les membres de l'équipe DRUM, plus que des collègues, vous êtes des amis. Je ne vous cache pas l'émotion qui me submerge à l'écriture de ces mots. Lenha, merci de m'avoir tenue par la main lors de mes débuts dans l'équipe, merci pour ta bienveillance, ta patience et tout ce que tu m'as appris, car c'est bien toi qui m'as intégrée à cette équipe plus que quiconque. Dorine, tu as été un soutien essentiel quand rien n'allait, la meilleure des voisines de bureau, et par-dessus tout la force de caractère que tu dégages m'a inspirée à en prendre l'exemple. Stéphane, en plus de nos travaux communs, de nos expériences partagées que j'ai tant appréciées, tu m'as marquée par ton altruisme, ta volonté et ta joie de vivre. Merci pour avoir été à bien des moments mon confident quand rien ne semblait aller au labo, et pour tout ton soutien. Je pense également à Manon, qui a elle aussi participé aux moments de joie dans le labo, m'a tant fait rire et m'a marquée par sa gentillesse. A tous ceux qui ont participé activement aux travaux de ce projet, je vous remercie tant pour votre intelligence et votre travail que j'admire, que pour votre amitié et tous nos moments de joie qui resteront gravés à tous jamais : Manuel, Olivier, Alexandre H., Alexandre E., c'est grâce à vous que cette thèse est ce qu'elle est, et je mesure la chance d'avoir travaillé à vos côtés. Avec André, Thomas, Thibault, Tatiana, Anaïs, Agathe, et tous les autres que je n'oublie pas, c'est grâce à vous tous qu'il a été si facile d'aller au travail chaque jour, pour échanger avec vous et donner le meilleur de moi-même. Merci grand Alex, pour ta gentillesse et ton altruisme, petit Alex, meilleur co-doctorant/co-galérien qui soit, pour ton écoute et ta sincérité, Manuel pour toutes tes anecdotes aux déjeuners (malgré le kimchi) et pour avoir été le meilleur voisin de bureau post-Covid, Thomas tes encouragements et nos discussions en

cette fin de thèse... Olivier, pour ton altruisme, ta bienveillance, ta gentillesse, tes blagues incessantes et tes histoires folles, tes conseils et toutes nos discussions, je te remercie sincèrement, tu ne mesures peut-être pas tout ce que tu m'as apporté.

Navigating the PhD storm on the same boat, my warmest thoughts go to my second family, the IC3i team. Through our endless discussions, trips, dinners, parties, we have shared so many moments of joy and of doubts, and your support was essential. Tommaso, Danny, Ram, Maciej, Theb, thank you for all the moments shared together, it has been a pleasure being around you, seeing you grow and become doctors. Above all, merci aux Filles de Curie, Silvia, Sandra, Darine, Raquel, Samyuktha, Ozgë, as their friendship is without a doubt one of my main achievements at Curie. Ozgi, thank you for your strong personality that inspires me, for being so genuine; Sammy, for your incessant smile, all the moments of laughter shared together; Raquel, for your energy, your determination and your contagious passion for life; Darine, for your altruism, your tolerance, your kindness; Sandri, for having shared so many moments of laughter, for your friendship, your pillar role in all IC3i events; and Silvi, for your generosity, your listening, your support when we were the only two stuck with our manuscripts, your contagious laugh, your selflessness; last but not least Deep, thank you for I have found a true friend in you, and I always appreciate our conversations about anything. You have all been essential to my time here at Curie, my gratitude for having met you all is only topped by the hope that we stay close now that we embark on new adventures.

Il serait difficile de clore ces remerciements, visibles en premier, écrits en dernier et pensés toujours, sans rappeler la place essentielle de mes plus chers amis et de ma famille.

A Valentine (mon Valenton), qui m'aura tant écoutée, tant apporté de soutien et de joie pendant cette période, à elle et à Arthur (Turde), car cette thèse a grandi en parallèle de notre amitié croissante qui m'est essentielle. A Pierre, qui a suivi mes travaux et écouté mes états d'âmes sur la dernière ligne droite ; à Chloé, et tous les autres, avec qui je suis fière de partager ce parcours.

A vous quatre, papa, maman, Diane, Loïc. Vous savez tout l'amour, la force et le courage que je trouve en vous, et toute la reconnaissance que je vous dois pour votre soutien et votre écoute, ces quatre années et depuis toujours. Vous qui me permettez chaque jour de viser un peu plus haut, d'aimer et de me surpasser, vous qui donnez un sens à ma vie et à mes actions, et m'avez appris ce qu'est la joie.

A Jacques, Claudine, Jean, Martine, Marc, Déborah, Hugo, Alexandre, Bruno. Vous êtes la famille sans laquelle le travail n'a pas de sens.

Enfin, à Maxime, mon pilier dans la vie et mon confident dans les émotions qu'elle entraîne. A toi, avec qui partager tout cela a été si facile, toi qui effaces mes doutes, m'entends, et m'animes d'un feu d'ambition et de courage. A toi, que j'ai envie de rendre fier autant que je le suis de toi.

Et à ceux qui ne sont plus, mais sont en moi chaque jour pour me donner force, courage, et le devoir de vivre ma meilleure vie. L., H., A.

Laurence, ma marraine, merci d'avoir veillé sur moi de là-haut pendant tout ce chemin. Cet accomplissement, je sais que tu en serais très fière. Cet accomplissement, je te le dédie.

TABLE OF CONTENTS

REMERCIEMENTS.....	3
TABLE OF CONTENTS	6
TABLE OF FIGURES	9
ABSTRACT.....	10
ABBREVIATIONS	12
INTRODUCTION.....	13
GENETIC BASIS OF CANCER	13
CANCER GENETICS	13
ACTIVATING AND LOSS-OF-FUNCTION EVENTS	13
Oncogenes	14
Tumor suppressor genes	14
GENETIC ALTERATIONS	16
DNA mutations	17
Mutational signatures in cancer	18
ACQUIRED VS. INHERITED MUTATIONS IN CANCER.....	20
Somatic mutations.....	20
Inherited germline mutations and familial forms of cancer	22
GENETIC SUSCEPTIBILITY TO CANCER	25
STRONG AND WEAK PREDISPOSITION	26
APPROACHES TO IDENTIFY SUSCEPTIBILITY GENES	27
Rare, high-penetrance genes.....	27
Moderate penetrance - candidate gene approach.....	31
Weak penetrance - case-control association studies	32
GENOME-WIDE ASSOCIATION STUDIES	33
GWAS study design.....	33
Methodologies associated with GWAS studies.....	36
Post-GWAS functional studies: challenges and strategies	38
Susceptibility loci in cancer	41
MELANOCYTES AND MELANOGENESIS.....	43
MELANOCYTES	43
EMBRYOLOGY AND MELANOCYTOGENESIS.....	44
MELANOGENESIS	45
FUNCTION OF CUTANEOUS MELANOCYTES	48
UVEAL MELANOCYTES	49
DETERMINANTS OF EYE PIGMENTATION	51
POPULATION GENETICS AND PIGMENTATION	53
UVEAL MELANOMA	54
CLINICAL ASPECTS.....	54
GENOMIC LANDSCAPE OF UVEAL MELANOMA	55
SOMATIC MUTATIONS	56
Mutations of the $G\alpha_q$ pathway.....	57
BSE events (<i>BAP1</i> , <i>SF3B1</i> and <i>EIF1AX</i>)	62

CHROMOSOMAL IMBALANCES	68
Monosomy 3.....	68
8q gain	69
INTEGRATIVE GENOMIC LANDSCAPE OF UM.....	69
Epigenetic oncogenic mechanisms	72
PROGNOSTIC FACTORS	72
CLINICAL AND HISTOPATHOLOGICAL FEATURES.....	73
CYTOGENETIC CLASSIFICATION	73
GENE EXPRESSION PROFILING	75
METASTATIC PROGRESSION AND TUMOR EVOLUTION	76
MALIGNANT PROGRESSION MODELS	76
TUMOR EVOLUTION.....	78
Genomic evolution.....	78
Immune microenvironment.....	79
TREATMENT	80
PRIMARY TUMOR	80
METASTASIS	81
Conventional chemotherapy	81
Hepatic therapies.....	82
Targeted therapies.....	82
Immunotherapy	83
EPIDEMIOLOGY	84
ENVIRONMENTAL RISK FACTORS	85
IRIS MELANOMA AND EXPOSURE TO ULTRA-VIOLET RADIATION	86
GENETIC SUSCEPTIBILITY	86
FAMILIAL UM	86
<i>BAP1</i>	88
Other rare germline mutations.....	89
MODERATE TO LOW-RISK COMMON VARIANTS.....	91
<i>TERT/CLPTM1L</i> locus on chromosome 5	91
Pigmentation polymorphisms.....	94
MBD4.....	96
BASE EXCISION REPAIR PATHWAY	96
DNA GLYCOSYLASES	98
TDG and MBD4 substrate specificities.....	99
MBD4 GENE AND PROTEIN STRUCTURE	100
BIOLOGICAL FUNCTIONS OF MBD4	101
SUPPRESSION OF MUTAGENESIS AND DNA REPAIR	101
APOPTOSIS	103
TRANSCRIPTIONAL REPRESSION AND HISTONE DEACETYLATION.....	103
DNA DEMETHYLATION?	105
MBD4 deficiency in cancer.....	106
MBD4 GERMLINE MUTATION IN AN OUTLIER UM PATIENT	107
ARTICLES.....	109
INTRODUCTION TO ARTICLE 1.....	109
ARTICLE 1: GERMLINE MBD4 MUTATIONS AND PREDISPOSITION TO UVEAL MELANOMA	111
CONCLUSION TO ARTICLE 1	112
INTRODUCTION TO ARTICLE 2.....	115

ARTICLE 2: DIFFERENT PIGMENTATION RISK LOCI FOR HIGH-RISK MONOSOMY 3 AND LOW-RISK DISOMY 3 UVEAL MELANOMAS.....	116
CONCLUSION TO ARTICLE 2	117
INTRODUCTION TO ARTICLE 3.....	120
ARTICLE 3: FUNCTIONAL CHARACTERIZATION OF CLPTM1L/TERT 5p15.33 RISK LOCUS IN UVEAL MELANOMA IDENTIFIES RS452384 AS A FUNCTIONAL VARIANT REGULATING CLPTM1L AND TERT EXPRESSION THROUGH ALLELE-SPECIFIC BINDING OF NKX2.4.....	121
CONCLUSION TO ARTICLE 3	122
CONCLUSIONS AND PERSPECTIVES	125
RÉSUMÉ DE LA THÈSE VERSION FRANÇAISE.....	134
REFERENCES.....	147

TABLE OF FIGURES

Figure 1. Mutational signatures of single-base substitutions across human cancer types.	19
Figure 2. Contribution of genetic inheritance in breast cancer.	24
Figure 3. Genetic architecture of cancer risk.	27
Figure 4. Genome-Wide Association Study design.	35
Figure 5. From GWAS to functional genomics.	39
Figure 6. Principles of expression quantitative trait loci (eQTL) mapping.	40
Figure 7. Melanogenesis pathways leading to the synthesis of pheomelanin and eumelanin in melanosomes.	46
Figure 8. Anatomy of the eye showing the three different types of uveal melanoma: melanoma of the choroid, ciliary body, and iris.	49
Figure 9. Somatic mutational profile in uveal melanoma.	56
Figure 10. Signaling pathways downstream of $G\alpha_q$ pathway alterations in uveal melanoma.	60
Figure 11. Nuclear and cytoplasmic functional roles of BAP1.	64
Figure 12. Integrative study of the genomic landscape in UM reveals 4 molecularly and clinically distinct subgroups.	71
Figure 13. Overall survival Kaplan-Meier curves of UM-related death according to chromosomes 3 and 8 status.	74
Figure 14. Molecular subsets of uveal melanoma based on genetic, cytogenetic and transcriptomic characteristics.	76
Figure 15. Evolutionary route of metastatic uveal melanoma.	78
Figure 16. Manhattan plot of a GWAS conducted in UM.	91
Figure 17. Canonical mammalian base excision repair (BER) pathway.	97
Figure 18. Methyl-binding domain (MBD) protein family.	100
Figure 19. MBD4 germline mutation in an outlier UM patient.	108
Figure 20. Low eumelanin to pheomelanin ratios and defective eumelanin synthesis may lead to the generation of reactive oxygen species and DNA damage in the eye.	119

ABSTRACT

Uveal melanoma (UM) is a rare tumor, yet it is the most common primary intraocular malignancy in adults with ~500 new cases per year in France. It arises from the malignant transformation of melanocytes of the uvea, composed of the choroid (90% of cases), ciliary body (6%) and iris (4%). UM is a genetically simple tumor characterized by two major somatic driver events: mutually exclusive mutations of *GNAQ*, *GNA11*, *PLCB4* or *CYSLTR2*, initiating tumorigenesis by constitutive activation of the $G\alpha_q$ signaling pathway, and a second event through mutually exclusive mutations that define metastatic risk and patient outcome: loss-of-function mutations of *BAP1* associated with its bi-allelic inactivation by monosomy 3 (M3) leads to UM with high metastatic risk and poor prognosis, while disomy 3 (D3) is associated with either *SF3B1* or *EIF1AX* mutations and respectively confer intermediate and low risks of metastasis. Prognosis is dismal after occurrence of metastasis in 30–50% of cases, almost invariably to the liver, with a median survival of ~12 months. Strikingly, individuals of European ancestry with light eyes and fair skin are particularly at risk of developing UM, with a relative risk (RR) of up to 20 in populations of European ancestry compared to those of African or Asian ancestries. However, the absence of ultra-violet (UV) mutational signature and stable UM incidence over the past decades rule out a role for pigmentation protecting against UV light to explain this peculiar epidemiology. Besides, familial cases of UM are seen in 1% of UMs, yet germline mutations of *BAP1* are the only strongly predisposing genetic risk factor known so far and only explain a fraction of familial cases, suggesting that genetic risk of UM mostly remains unaccounted for.

We hypothesized that some susceptibility alleles could predispose populations of European ancestry to UM. We sought to investigate the genetic predisposing factors in UM following two approaches: (i) a candidate-gene approach by targeted-sequencing of *MBD4* in a large consecutive cohort of 1,093 UM patients, which allowed us to demonstrate that *MBD4* is a UM predisposing gene associated with high-risk M3 UM, conferring a RR = 9.2 of developing UM compared to the general population; and (ii) a genome-wide association study in UM followed by a functional genomics study to identify common, low-penetrant UM risk alleles. Through this second approach, we confirmed the UM risk region at chromosome (chr)

5 on the *TERT/CLPTM1L* locus and further identified two pigmentation susceptibility loci, *HERC2/OCA2* (chr15) and *IRF4* (chr6). Importantly, these two loci were differentially associated with high-risk, M3 and low-risk, D3 UMs respectively, distinguishing two UM subgroups marked by specific genetic risk factors influencing tumor biology and metastatic potential. Lastly, we functionally characterized the *TERT/CLPTM1L* UM risk locus. After demonstrating an allele-specific regulation of gene expression at this locus, we narrowed down the genomic region to one functional variant, rs452384, which exhibited allele-specific nuclear factor binding properties. Using a quantitative proteomic approach, we evidenced binding of NKX2.4 transcription factor specifically to the rs452384-T allele. We further show that knockdown of *NKX2.4* in UM cell lines results in a subtle yet significant increase in *TERT* and *CLPTM1L* expression, representing the first steps towards deciphering the tumorigenic mechanism of this risk locus in UM. Finally, we suggest that rs452384 may mediate its activity through differential telomere length regulation, potentially arising prior to UM tumorigenesis. Taken together, these results shed light on multiple moderate and low-risk genetic predisposition factors in UM, which bring new insights on the underlying biological mechanisms in UM predisposition and has clinical relevance as patients with *MBD4* deficiencies may respond to immunotherapy, resulting in the addition of *MBD4* to clinical genetics panels.

This thesis is divided into four introductory chapters, reviewing the cancer genetics notions underlying the present work, melanogenesis, uveal melanoma and *MBD4*. The results are then described in successive articles, two of which have been published, the last one about to be submitted for consideration in a scientific journal. These results are individually discussed, before general conclusions and future perspectives are made.

ABBREVIATIONS

BER: base excision repair pathway
CNV: copy number variation
CSG: cancer susceptibility gene
CM: cutaneous melanoma
D3: disomy 3
GWAS: genome-wide association study
eQTL: expression quantitative trait loci
LD: linkage disequilibrium
LOH: loss of heterozygosity
MAF: minor allele frequency
M3: monosomy 3
MUM: metastatic uveal melanoma
NGS: next-generation sequencing
OR: odds ratio
PUM: primary uveal melanoma
QTL: quantitative trait loci
RR: relative risk
SNP: single nucleotide polymorphism
SNV: single nucleotide variant
TCGA: The Cancer Genome Atlas
TF: transcription factor
TMB: tumor mutation burden
TSG: tumor suppressor gene
UM: uveal melanoma
UV: ultra-violet
UVR: ultra-violet radiation

INTRODUCTION

GENETIC BASIS OF CANCER

It is now well known that cancer is a genetic disease. Carcinogenesis arises from mutations of cancer driver genes and from epigenetic changes, both of which are preserved as cancer cells divide, acting in combination with environmental factors to drive tumor formation. The resulting deregulation of expression or function of critical genes is a primordial feature of cancer, required for tumor formation and ultimately metastatic spread.¹ These genes confer to cancer cells some common features, such as uncontrolled cell division and growth, enhanced survival, evasion of apoptosis and differentiation, cell cycle deregulation, and acquisition of angiogenic and metastatic capacities.^{2,3} While most DNA mutations driving cancer occur sporadically (somatic mutations), some are passed on hereditarily from parent to child (germline mutations). After reviewing the key features underlying cancer genetics, we will focus on cancer predisposition genes and the inherited form of cancer.

CANCER GENETICS

Cell transformation during tumorigenesis is driven by accumulation of genetic and epigenetic alterations, also highly influenced by environmental and lifestyle factors, which ultimately result in the gain- or loss-of-function of key cancer driver genes and proteins that confer a selective advantage to tumor cells over healthy cells.

ACTIVATING AND LOSS-OF-FUNCTION EVENTS

All cancers arise from successive steps of mutations within genes that drive tumorigenesis and confer to tumor cells a growth advantage. Somatic mutations accumulate throughout time in virtually all dividing cells, normal and neoplastic, at a constant rate per cell division; yet, there is high heterogeneity in the number of somatic mutations both across and within cancer types,⁴ where the mutational rate greatly varies according to exposure to intrinsic and extrinsic mutagens, cell type, presence of rare inherited diseases, and potentially an increase in somatic mutation rate in cancer cells after neoplastic change.⁵ However, not all cancer

types require an accelerated somatic mutation rate and enhanced genomic instability to develop.⁶ More than 97% of somatic mutations in tumors in fact do not confer any selective advantage to the cancer cell and simply accumulate throughout time in the tumor without favoring the neoplastic process: these are the so-called “passenger” mutations, in contrast to “driver” mutations, found within a subset of genes that are causally implicated in cancer development by conferring selective growth or tumor fitness to the cell.⁷

ONCOGENES

On one hand, these driver mutations may have an “activating” effect, either by leading to the formation of a hyperactive protein (through gene amplification events), its constitutive expression (essentially through point mutations leading to enhanced transcription from constitutively activated promoters) or to the formation a fusion protein with oncogenic functions (through chromosome translocations).^{1,7} Such mutations reside within oncogenes, in which an activating somatic event in a single allele is sufficient to confer growth advantage to the tumor cell and express the gain-of-function phenotype. Since the discovery of the first oncogene (*c-src*) in the 1970s through the study of retroviruses,⁸ a multitude of oncogenes have been characterized, with different biological functions such as growth factors (such as *c-Sis*, *EGF*, *FGF*), growth factor receptors (mainly receptor tyrosine kinase, such as *EGFR*, *HER2*, *VEGFR*), signal transducers (including *KRAS*, *GNAQ*, *HRAS*), transcription factors such as *MYC*, and regulators of apoptosis like *BCL-2*.

TUMOR SUPPRESSOR GENES

On the other hand, loss-of-function events can also lead to tumorigenesis, through genetic changes leading to the inactivation of tumor suppressor genes (TSGs). Inactivating mutations of those genes typically include stop-gain or missense mutations, indels of a few nucleotides, mutations affecting gene splicing or large gene rearrangements. Unlike oncogenes that are frequently mutated at the same amino acid positions (‘hotspot’ mutations), TSGs mutations do not follow a clear pattern and typically span the entire length of the gene.⁷ These mutations are usually associated with a second somatic event inactivating the second allele, such as loss of an entire chromosome arm or a segment containing the TSG.¹ The association of two loss-of-function events (either by combination of genetic mutations targeting both alleles, or combining genetic and chromosomal events) targeting both parental alleles are

usually required to observe the phenotype and lead to tumor growth, in a Knudson's two-hit model.^{9,10} This phenomenon is essentially evidenced by search of 'loss of heterozygosity' (LOH), which aims to identify the loss of an allele at the tumoral level compared to a matched DNA sample from blood. In some cases however, inactivation of only one wild-type allele of a TSG (one 'hit') is sufficient to confer a selective advantage to the tumor cell: this refers to haploinsufficiency, a form of genetic dominance where loss of only one allele, associated with reduced levels of the encoded protein, results in the disease phenotype as the remaining wild-type allele is insufficient to execute normal physiological functions and achieve the wild-type phenotype.¹¹ An example is the TSG *TP53*, commonly in multiple cancers and associated with Li-Fraumeni syndrome, or *53BP1* in glioblastoma, where LOH is not observed in a subset of patients and despite retention of one wild-type allele, the decrease in protein levels are enough to partially or strongly abolish protein function and drive tumorigenesis.¹²⁻¹⁴

The biological functions of TSGs are essentially opposite to that of proto-oncogenes. The analogy of a brake pedal on a car depicts well their role in regulating and inhibiting cell division or DNA replication, or in inducing apoptosis when necessary. Well-known proteins coded by TSGs and implicated in such pathways include RB, p16, APC and p53. Bi-allelic inactivation of a TSG may result in the formation of a truncated protein with partial functions, a full-length protein with inactivated function, or complete lack of expression of the protein.

TSGs are sometimes referred to as gatekeepers, due to their shared function in the regulation and/or inhibition of cell growth. Another subset of TSGs (sometimes classified as a third type of cancer genes) are the caretakers, or stability genes.¹ They follow similar tumoral inactivation pattern as TSGs but share a common function in ensuring genome stability and keeping DNA damage to a minimum.¹ Major DNA pathways include the mismatch-repair (MMR), nucleotide-excision repair (NER), base excision repair (BER) pathways and homologous recombination (HR), which have distinct repair mechanisms and are each implicated in the repair of specific types of damage, but ultimately all play a crucial role in maintaining genetic stability and keeping alterations to a minimum. Caretakers are frequently mutated in cancer, most often resulting in the increased mutation rate of other genes, including oncogenes and TSGs that ultimately affect net cell growth. Common examples of caretakers mutated in cancer include mismatch repair proteins MLH1 or MSH2, pleiotropic repair protein BRCA1 involved in multiple repair pathways, or double-stranded DNA break

repair protein ATM; their inactivation results in a higher rate of acquisition of somatic DNA alterations.¹⁵

It should be noted that the dichotomy between gatekeepers and caretakers is imperfect, as some proteins may have overlapping functions in the two categories, such as p53. Altogether, mutations within caretakers result in unrepaired DNA damage, which in turn leads to the accumulation of further genetic alterations, including structural rearrangements and point mutations that are critical to cancer, as discussed below.

GENETIC ALTERATIONS

Similar to ageing, malignant transformation in cancer involves an accumulation of mutations and epimutations passed on through cell divisions to daughter cells, which spontaneously accumulate in normal cells at low levels and contribute to cancer risk.¹⁶ During tumorigenesis, which is also influenced by environmental factors, this accumulation correlates with the appearance of alterations within cancer drivers (TSG inactivation or oncogene activation), although most acquired changes are considered to be passengers. In the context of this work, we will focus on the genetic component of alterations that can lead to cancer, however epigenetic changes (i.e., changes to the genome that are not affecting the base sequence of DNA but are still transmitted through cell divisions) can also exert major effects genome stability and expression of tumor-related genes, such as through epigenetic inactivation of TSGs via gene silencing. For a review of epigenetic alterations that are fundamental to cancer (such as histone modifications, chromatin remodeling and DNA methylation), please see Jones *et al.*¹⁷ Although we will examine here the causes and types of DNA mutations in normal tissues and cancer, which can be seen as “small-scale” DNA damage, a second class of genetic alterations can affect multiple genes at once through large-scale events. Chromosomal abnormalities are a major source of genomic instability, highly recurrent in (and characteristic of) cancer.^{1,18} Chromosomal abnormalities are usually a result of errors in mitosis or DNA damage and encompass numerical changes, such as chromosomal gain or loss, and structural changes including chromosome translocations, deletions, large amplifications and inversions, and whole-genome duplications.

DNA MUTATIONS

DNA mutations in tumors vary by type and number; however, these mutations are initially acquired in normal cells, whether it be spontaneously (such as due to unrepaired DNA) or environmentally induced (such as following exposure to a mutagen).

DNA mutations directly modify the DNA sequences, affecting one or a couple of nucleotides. They include single-base substitutions (SBS) and insertions and deletions (indels) of one or a few base pairs that sometimes result in frameshifts. The genome is exposed to damage from both endogenous and exogenous sources, and while a majority is effectively repaired by multiple DNA repair mechanisms present in the cell, some fail to be recognized and remain unrepaired. Mutations do not affect each nucleotide position equally and often follow specific mechanisms. Therefore, distinct patterns of mutagenesis exist, with dominance of some specific mutations within the existing repertoire.¹⁹ Major sources of endogenous DNA damage include reactive oxygen species (ROS) resulting in oxidative DNA adducts, of which 8-oxo-dG is the most abundant, deamination of bases containing primary amines (such as hydrolytic deamination of cytosine and 5-methylcytosine, generating uracil and thymine respectively, with a net effect of C.G→T.A transitions), depurination resulting from spontaneous base loss, and errors in replication such as those arising from replication slippage or due to intrinsic proofreading properties.²⁰ Added to these, inherently error-prone DNA polymerases²⁰ or DNA repair pathways such as transcription-coupled nucleotide excision repair (NER) can directly cause further DNA mutations, while deficiencies or alterations of DNA repair pathways (such as mismatch-repair pathway in colorectal cancer)⁷ will result in failure of DNA repair. In addition to these intrinsic mechanisms, DNA mutations also arise from external influences and exposure to environmental factors, such as ultra-violet (UV) light exposure, tobacco smoke, and alkylating agents often used as chemotherapeutics.^{20,21}

Although some mutations are inherited from parents (i.e., germline mutations), an overwhelming majority of DNA sequence changes in a cancer genome are somatic mutations, acquired throughout lifetime and resulting from exposure to intrinsic and extrinsic mutagens.⁵ The small fraction of these mutations that are not appropriately repaired are converted into fixed DNA and accumulate over time at a rate that varies according to different

mutagens and to cancer type specific neoplastic changes.⁵ The somatic mutation load is consequently highly heterogeneous between samples and across tumor types,²² ranging from approximately 0.001 mutations per megabase (Mb) with pediatric tumors and leukemias at the low-end of the spectrum, to over 500 Mb in high-outliers. These include hypermutated tumors with DNA repair deficiencies (such as tumors with mismatch repair (MMR) defects exhibiting microsatellite instability (MSI), DNA polymerase defects, nucleotide excision repair (NER) deficiencies leading to skin-cancer prone xeroderma pigmentosum), but also tumors exposed to potent DNA mutagens such as melanoma and lung cancers,^{7,23} which are characterized by high exposure to UV and smoke, respectively.

MUTATIONAL SIGNATURES IN CANCER

Different mutational processes (and their combination) result in characteristic patterns, or “signatures”, that are jointly shaped by DNA damage (induced by endogenous and exogenous DNA mutagens), infidelities in DNA replication, and deficiencies in DNA repair mechanisms.²³ Major advances in next generation sequencing (NGS) technologies and large-scale cancer genome studies²⁴ dramatically increased the power to detect multiple mutational signatures within a single sample.^{19,20} Alexandrov and colleagues used a large mutational catalog of over 30 different cancers and incorporated information of the closest 5’ and 3’ bases from the mutations to extract mutational signatures according to a 96 substitution classification, shedding light on cancer etiology.²³ These signatures represent a combination of multiple factors and can also be instructive for the understanding of natural processes such as ageing. Since the 21 distinct mutational signatures described in 2013,²³ many were recently reported for single-base (SBS), doublet-base, clustered-base substitutions and small indels²⁵ from the large-scale sequencing analysis from the Pan-Cancer Analysis of Whole-Genomes Consortium (PCAWG),²⁶ the International Cancer Genome Consortium (ICGC) and The Cancer Genome Atlas (TCGA). The repertoire of 49 mutational signatures from SBS signatures are shown (Figure 1); many mutational signatures still remain without known etiology.²⁵



Figure 1. Mutational signatures of single-base substitutions across human cancer types.

Proportion of mutations (%) are shown for each of 96 classes of mutation types (<https://cancer.sanger.ac.uk/cosmic/signatures/SBS>) comprising the 6 base substitutions with all possible combinations of 5' and 3' flanking bases. Figure from Alexandrov *et al.*²⁵

Remarkably, some signatures are highly recurrent across tumor types while others are cancer specific. Signature SBS1 is virtually present in all cancer types and is thought to arise from the spontaneous deamination of 5-methylcytosines,²⁷ leading to CpG>TpG transitions and subsequent T/G mismatches that are not repaired before DNA replication. These mutations are usually repaired by the base excision repair (BER) pathway, initiated by the removal of mispaired bases by DNA glycosylases and further processed by short-patch or long-patch repair (BER is detailed in “MBD4 and cancer” chapter). Deamination of methylcytosines is a natural process that occurs virtually in every cell since the zygote state. The number of mutations characterized by signature SBS1 correlates in some types of tumors with age at diagnosis. It has since then been identified as a clock-like mutational process that correlates with age,²⁸ where mutations seen in cancer genomes predominantly derive from normal pre-

neoplastic cells and for which the mutation rate correlates with the number of mitotic cell divisions since fertilization of the egg.

SBS5 is also highly recurrent across tumor types, also correlates with age²⁵ and is part of signatures showing transcriptional strand bias (more mutations present on the untranscribed strand), probably due to mechanisms inherent to transcription-coupled nucleotide repair (NER). Multiple other highly recurrent signatures such as SBS2 and SBS13 account for APOBEC activity, which are cytidine deaminases (converting cytosines to uracils) likely to be linked with cancer development.²⁹ On the other end of the spectrum, some signatures are highly cancer specific, such as SBS7 linked to ultraviolet (UV) exposure, mostly found in skin melanoma and to some extent, head and neck squamous cell carcinoma. UV mutational signature is mainly characterized by a high frequency of C>T transitions at pyrimidine dimer sequences containing cytosines, prevalent on the untranscribed strand and repaired by transcription-coupled NER.³⁰

ACQUIRED VS. INHERITED MUTATIONS IN CANCER

Although most cancers are not caused by genetic inheritance (often referred to as sporadic cancers, solely driven by the accumulation of acquired somatic mutations), inherited (germline) mutations transmitted to progeny may also confer predisposition to cancer, whether it be within a highly penetrant cancer gene or through transmission of common genetic variation (discussed in the next section on genetic susceptibility to cancer).

SOMATIC MUTATIONS

Somatic mutations encompass all mutations that are acquired throughout life in a non-germ cell of the body. As previously mentioned, cancer development is driven by the accumulation of somatic mutations, some already present in the preneoplastic phase, and the others acquired later during tumorigenesis. Initiation of the neoplastic process is more specifically driven by somatic changes within an oncogene or TSG, leading to the initial clonal expansion, followed by the acquisition of multiple other somatic mutations (both passenger and driver) that result in further clonal expansion and tumor progression through progressive gain of the cancer cell phenotype and some of its hallmark features, resulting in excessive growth, sustained proliferation and resistance to apoptosis.^{3,31,32}

Sporadic cancers are solely driven by somatic changes and are thought to account for 70 to 80% of all cancers, although estimates of the proportion of inherited cancer forms vary according to definition, as discussed later. As mutations are regularly acquired throughout life, risk of sporadic cancer increases with age, which is the biggest risk factor for cancer, although the relationship is nonlinear and affected by multiple processes such as selective pressure, microenvironment changes, alterations to the immune system and others.^{33,34} While cancer evolution may be seen, much like ageing, as a multistage process characterized by the acquisition (and accumulation) of specific somatic events through time, somatic mutations acquired during development via somatic mosaicism somehow lie in-between acquired mutations and predisposition to cancer.

SOMATIC MOSAICISM

Genetic mosaicism describes the occurrence of two distinct populations of cells that differ in their genetic content. Unlike gonadal (or germline) mosaicism that affects gamete-producing tissues, somatic mosaicism arises from postzygotic mutations during development, are not transmitted to progeny and vary in terms of severity of phenotypic consequences according to the type of mutation and whether it has a dominant effect, as well as how early in development the mutation occurs. This timing reflects the proportion of cells that will carry the mutation as the latter will be inherited in all cells descending from their lineage.^{35,36} In other words, genetic mosaicism reflects the presence within a subset of cells of a somatic mutation, distinct from the inherited germline DNA. Much like cancer progression, fixation of this mutation and selection through subsequent clonal expansion are key features of genetic mosaicism, for which ageing is also a known risk factor.³⁷ Since cancer cells acquire specific fitness features and growth advantage through somatic mutations of driver genes that are not acquired in the rest of the cells, cancer can be seen as a classic example of somatic mosaicism. Somatic mutations acquired early in development are now sometimes seen as major contributors to cancer development rather than exceptional occurrences.³⁶ An example of somatic mosaicism highlighting the different phenotypic consequences depending on developmental stage and cell type affected by the mutation are identical mutations of *GNAQ*, which can either lead to mild port-wine stains, severe neurocutaneous disorder Sturge-Weber syndrome, or act as activating mutations in uveal melanoma.³⁵

Transitioning to germline mutations that are passed to offspring, part of the supposedly *de novo* cases of pathogenic mutations (i.e., mutations not seen in parents and only detected in the offspring) are in fact inherited from a parent, which may have mosaicism affecting either only the germline or both germline and somatic cells (gonadal mosaicism). Germline mosaicism is commonly seen in X-linked and autosomal dominant disorders. Although acquisition of somatic mutations is a core cancer driver process, these examples highlight the role of inherited germline mutations in disease.

INHERITED GERMLINE MUTATIONS AND FAMILIAL FORMS OF CANCER

Inherited (or germline) mutations are mutations that are passed on hereditarily from parent to child, present in virtually all cells since the zygote state, including germ cells. In the case of inherited cancer risk, this includes all genetic variation transmitted from parents, from pathogenic germline mutations within highly penetrant cancer genes (such as TSGs on oncogenes) to common genetic predisposition variants (SNPs) of low penetrance that also participate in the inherited polygenic architecture of cancer.

Germline mutations do not cause cancer *per se* but rather predispose individuals to develop cancer in their life. In the case of high-penetrance pathogenic variants, their germline mutation represents the first hit of the series of oncogenic events that lead to tumorigenesis, while somatic inactivation of the remaining allele (second hit) will tend to be the rate-limiting step in tumor initiation (in the case of haploinsufficiency, the germline monoallelic mutation is sufficient to contribute to the cancer phenotype). Inherited mutations of cancer driving genes (oncogenes, TSGs or stability genes) lead to an elevated risk of cancer, and individuals affected by cancer hereditary syndromes tend to develop multiple tumors and often at a much earlier age than in sporadic cases, as it is the case in hereditary retinoblastoma compared to its sporadic counterpart (both driven by mutations of the *RB1* gene).^{1,10} Inherited mutations associated with loss-of-function events (through monoallelic mutations of a TSG) are much more common than oncogenic gain-of-function, as the remaining wild-type allele remains functional during development³⁸ while oncogenic activating mutations often have a dominant effect that can disturb embryonic development. However, both types of mutations still exist, an example being high-risk susceptibility genes in cutaneous melanoma (CM), *CDKN2A* and *CDK4*, where the former encodes two tumor suppressors

implicated in cell cycle inhibition, while the latter encodes an oncogenic Ser/Thr protein kinase involved in cell cycle progression.^{39,40} Although patterns of germline mutations of these two genes differ, they both strongly predispose to CM and lead to very similar phenotypes in melanoma families.⁴¹ Other typical examples of mutations of cancer genes inherited in a Mendelian fashion (where inheritance of a single autosomal dominant gene confers strong disease susceptibility) include *TP53* in Li-Fraumeni syndrome (increased risk of breast, sarcoma and other tumor types), *PTEN* in Cowden syndrome (predisposition to glioma, uterus tumors and others), germline mutations of DNA repair genes *BRCA1* or *BRCA2* strongly predisposing to breast cancer,⁴² or mismatch repair gene *MLH1* in susceptibility to non-polyposis colorectal cancer.⁴³

Germline mutations of the above-mentioned genes are essentially rare mutations, associated with a very strong risk of developing cancer when inherited (i.e., with high penetrance, discussed in the “Genetic susceptibility to cancer” section below). However, these only represent a portion of hereditary predisposition to cancer. The observation of familial aggregation of cancer, occurring within family members more often than expected by chance, enables the identification shared genetic traits within affected members. This approach has been extensively used to identify the first highly penetrant inherited genes, as a higher genetic contribution is expected in these cases compared to sporadic cancers. Paradoxically, although inherited pathogenic variants within known cancer genes are associated with a strongly increased risk of cancer (inherited cancer syndromes), these genes only account for a small minority of familial forms of cancer. An example illustrating this point is familial breast cancer, which accounts for 15–20% of all breast cancer cases.⁴⁴ Germline mutations of *BRCA1* or *BRCA2* confer cancer risk associated with the highest penetrance of all known breast cancer predisposing genes (risk of developing breast cancer at age 70 of 64.5% with *BRCA1* mutation and 61.0% with *BRCA2* mutations),⁴⁵ yet these two genes added to other high penetrance breast cancer genes only account for approximately 20% of familial risk (Figure 2) and 4–5% of all breast cancers,⁴⁴ suggesting the implication of multiple other inherited variants, probably with much lower penetrance, in cancer risk.

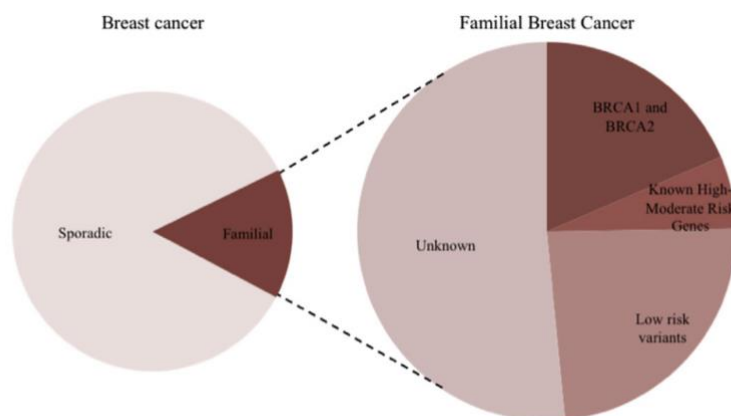


Figure 2. Contribution of genetic inheritance in breast cancer.

Familial (hereditary) breast cancer accounts for 5-10% of breast cancer. Among identified predisposing genes, *BRCA1*, *BRCA2* and other high penetrance genes account for around 20% of familial breast cancer, moderate risk variants 5%, and the rest are either low-risk variants or have not yet been identified. Figure from Aloraifi *et al.*⁴⁶

INHERITED EPIGENETIC ALTERATIONS

As a last point before discussing the different forms of genetic susceptibility to cancer, there is also growing interest in the epigenetic inheritance of cancer susceptibility. As epigenetic changes are also key drivers of cancer progression and are tightly linked to genetic mutations, epigenetic changes may also account for some of the 'missing heritability' in cancer. Main epigenetic determinants of cancer⁴⁷ include DNA methylation and its effects on transcriptional silencing, nucleosome remodeling, changes in histone post-translational modifications, changes in the expression pattern of regulatory microRNA (miRNA), and multiple others – all of which, when deregulated, can ultimately affect expression of the same TSGs and oncogenes that are otherwise frequently mutated in cancer, albeit without any observable mutation in these genes. It is now recognized that epigenetic alterations are also heritable changes that favor cancer evolution.⁴⁸ Not only are these epigenetic changes mitotically and meiotically heritable (passed on to daughter cells),⁴⁹ some data also suggest that they account for some of the earliest heritable changes in tumor evolution, particularly epigenetic silencing via DNA methylation repatterning.⁴⁷ Furthermore, there is growing evidence of intergenerational epigenetic inheritance, characterized by epigenetic changes in the parent germ line that are transmitted to offspring.⁵⁰⁻⁵³ In this respect, heritable epigenetic changes that can confer cancer risk include not only cancer susceptibility genes encoding chromatin regulators, but also some specific histone modifications or methylation patterns

that are passed on hereditarily and that ultimately favor tumor initiation or progression through aberrant regulation of wild-type cancer genes.⁵⁰

GENETIC SUSCEPTIBILITY TO CANCER

Genetic predisposition to cancer is a relative notion, corresponding to the increased risk of a given cancer conferred by a genetically inherited trait compared to the average risk in the general population. Large familial and twin studies⁵⁴⁻⁵⁶ have played critical roles in providing estimates of the genetic heritability of cancer (i.e., the fraction of cancer risk that is explained by genetic variation)⁵⁷ in the general population and in familial cancer. Familial forms of cancer refer to the co-occurrence of cancers in at least two first- or second-degree relatives within a family, and often arise from the combination of shared genetic traits and environmental factors.^{54,58} Family history is consequently a major cancer risk for most tumor types. It is currently thought that between 5 and 10% of all cancers are associated with an inherited (germline) mutation in a cancer gene, such as a TSG, that predisposes to cancer and often clusters within families.^{59,60} However, depending on the cancer type, estimates of the genetically heritable component of cancer derived from twin studies often largely exceed this number.⁵⁴⁻⁵⁶ This highlights the complexity and ambiguity in the estimation of cancer heritability within populations, where “inherited cancer” sometimes only refers to hereditarily transmitted genes with high penetrance, which in fact only represent a small portion of heritable cancer.

Cancer is a complex, polygenic disease, arising from the combination of multiple genetic variants. Even in familial occurrences where pathogenic germline mutations often cluster, a significant proportion of shared genetic traits arises from a combination of common variants. A significant portion of cancer risk is most likely explained by low to moderate risk variants, also inherited through germline variation, which probably account for much more than the 5–10% of pathogenic germline variants predisposing to cancer. In fact, these low-risk germline variants are now thought to greatly contribute to somatic mutation burden and other somatic events influencing clinical outcome.^{61,62} Accordingly, research on heritability is transitioning from traditional twin and family studies towards large case-control association studies, which provide the opportunity to account for common genetic variation by studying the contribution of single-nucleotide polymorphisms (SNPs) to cancer heritability.

Genetic susceptibility to cancer therefore refers to all inherited genetic factors, which can be distinguished based on the strength of their penetrance. Moreover, cancer susceptibility genes (CSGs) and risk alleles are often identified through distinct approaches depending on their conferred risk in genetic predisposition and their frequency in the population.

STRONG AND WEAK PREDISPOSITION

The penetrance of a genetic variant (or allele) refers to the proportion of individuals carrying this variant (or genotype) that also express the associated phenotype.⁶³ If a variant in a cancer gene has a 50% penetrance, then half of the patient population harboring this variant will develop the cancer. Penetrance can be assessed using different disease models depending on the expected relationship between genotype and phenotype, including dominant, recessive, additive and multiplicative models.⁶⁴ The relative risk (RR) is a commonly used measure of the strength of association between a genotype and a phenotype. Although the mathematical relationship between RR and penetrance varies according to the disease model, the RR of a disease allele is inherently linked to its penetrance as it compares the penetrance in individuals carrying the different genotypes at the variant position. The vast differences penetrance and RR of a predisposition gene or variant reflect the distinctions previously made between strongly predisposing cancer genes (with autosomal dominant patterns of single-gene inheritance) and weakly predisposing variants that behave much more as polygenic determinants of cancer risk. In addition, there is an inverse relationship between the RR of a CSG and its frequency in the population (Figure 3), implying that highly penetrant genes are rare and only account for a minority of cancer inheritance. On the other hand, genes or SNPs with low to moderate penetrance are more common in the general population, individually contribute little to cancer phenotype, but account for a much more significant proportion of polygenic cancer risk through the combination of multiple common susceptibility alleles.⁶⁵ Identification of low-risk predisposition loci may unravel some familial risk of cancer but may also explain the inherited susceptibility within cancers that do not usually cluster in families, owing to their lower penetrance with cancer. Complementary approaches are often used to unravel CSGs with high, moderate or low penetrance, detailed below.

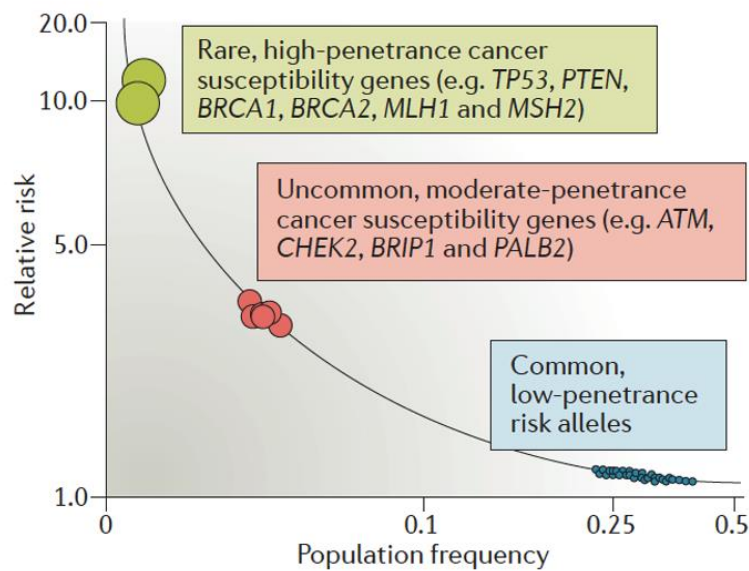


Figure 3. Genetic architecture of cancer risk.

The relationship between the relative risk conferred by a predisposing gene or variant and its minor allele frequency in the general population reflects its cancer penetrance and distinguishes between three forms of cancer predisposition that can be identified with different approaches. Figure from Sud *et al.*⁶⁵

APPROACHES TO IDENTIFY SUSCEPTIBILITY GENES

CSGs can be split according to their frequency in the general population and their penetrance. Rare, high-risk variants usually have a minor allele frequency (MAF) of <0.005 (based on the Human Genome Variation Society nomenclature), and typically confer RR values of >4 ; common, low-risk variants tend to have a MAF >0.05 , and usually do not confer RR of more than 1.5 or 2; moderate-risk variants lie within these two ranges and do not form a distinct group *per se*. Different strategies have been employed to identify CSGs, according to their frequency and penetrance. Importantly, many of these have evolved in the past decade with the development of high-throughput sequencing (HTS) technologies.

RARE, HIGH-PENETRANCE GENES

The observation that multiple cancers cluster within families led to the development of genetic linkage studies in the 1980s, searching for genes responsible for monogenic diseases. While multiple highly penetrant CSGs were initially discovered using this approach, the limits to this method combined with revolutionizing advances in sequencing technologies led to the

shift towards large-scale sequencing and both family-based and population-based association studies.

FAMILY-BASED GENETIC LINKAGE STUDIES

Most highly penetrant CSGs are associated with cancer predisposition syndromes and therefore common cancers usually segregate strongly (>3 affected members) or moderately (2 affected individuals) in families. The seek for strong CSGs in the past 30 years has led to the discovery of over 70 CSGs conferring RRs of 5 to 100,⁶⁶ with the strongest peak in the 1990s initially arising from the broad use of linkage analysis at that time.⁶⁷ This family-based technique aims to track the genomic position of strongly penetrant markers within family members affected by a certain trait, using a genome-wide agnostic approach demonstrating the co-segregation of these genetic markers with the trait. This technique has been particularly useful for the identification of highly penetrant genes with Mendelian inheritance, which are also often associated with early onset of cancer, enabling to search for such genes in families with higher risk of germline mutations (as it is strongly the case for example in inherited forms of breast, ovarian or colorectal cancer). A gene responsible for family clustering is usually mapped by successive rounds of linkage narrowing down the whole genome to small intervals of <1 centiMorgan showing association,^{68,69} followed by fine-mapping, sequence analyses and assessment of potential biological relevance of candidate genes. A wide number of genetic linkage and positional cloning studies allowed to identify main TSGs and DNA repair genes involved in cancer predisposition syndromes, such as *BRCA1/2* in ovarian and breast cancers, *MSH2* in colorectal cancer, *CDKN2A*, *PTEN* and many others (Figure 3).^{39,65,70-73}

SHIFT TOWARDS ASSOCIATION STUDIES AND HIGH-THROUGHPUT SEQUENCING

More recently, approaches to discover new CSGs with high to moderate penetrance have shifted strategies^{67,74} owing to major advances in sequencing techniques (HTS) such as next-generation sequencing (NGS) of whole exomes or genomes (WES and WGS), and access to large databases of genetic and genomic information. Comino-Mendez *et al*⁷⁵ were the first to report the identification of a strongly penetrant CSG through exome sequencing: *MAX*, implicated in hereditary pheochromocytoma. Case-control association studies were then

developed, enabling the detection of genetic variants with high and low frequencies. However, since inherited variants of high penetrance are expected to be rare in the general population, some strategies were developed to overcome their expected low statistical power, such as family-based association studies and the use of high throughput sequencing and large-scale population studies.

High-throughput sequencing in large population studies

The advent of NGS high-throughput technologies, including WES but also deep and ultra-deep WGS, provides an unprecedented opportunity to investigate low-frequency and rare genetic variants that may have been missed in previous studies due to lack of sequencing depth, and accelerated the discovery of germline disease-causing mutations.⁷⁶ Deep sequencing studies, but also resequencing of existing data, are now a frequently used strategy and allowed significant increase in molecular diagnostic, especially for rare diseases.^{77,78} However, large-scale and high-depth sequencing-based techniques are often costly, and deep sequencing is impractical for large populations. Depending on the objective of the study, ways to address these limitations in the detection of rare variants include pooled DNA sequencing strategies,⁷⁹ exome sequencing, targeted sequencing or low-depth WGS, and rare-variant genotyping arrays;⁸⁰ but the recent release of thousands of sequencing data (genome and exome), matched to thousands of phenotypes, in large international projects now provides an opportunity to perform high-depth, large-scale population-based analyses to unravel rare disease variants of high penetrance. Major pan-cancer studies in the TCGA or in the Pediatric Cancer Genome project (PCGP) have allowed the identification and validation of multiple pathogenic germline variants, mostly TSGs such as *BRIP1*, *ATM*, *BRCA1/2*, *PALB2*, *NF1* or *TP53*.^{81,82} In 2019, the catalog of CSGs with high or moderate penetrance was extended to 152 different CSGs (only a small half of which originate from familial studies), covering over 10,000 cancer patients and 33 cancer types from rare variants of the TCGA germline data,⁸³ reporting new associations of known CSGs with certain types of cancer, and new predisposition variants altogether. As a last example, Rasnic and colleagues took advantage of the UK Biobank (UKB) collection to search for cancer-exclusive ultra-rare and highly-penetrant variants using a UKB DNA array and reported novel CSG candidates and new variants within known CSGs.⁸⁴ The same strategy was applied using exome sequences from the UKB and the Healthy Nevada Project to detect rare variants associated with >4,000

phenotypes simultaneously, illustrating the advantages of performing large population studies of deeply phenotyped cohorts combined with recent sequencing technologies.⁷⁸

Case-control studies for rare variants

The advent of large population-based association studies, now routinely used for the identification of common genetic variants (discussed later), can also be applied to the search of low-frequency and rare genetic variants, owing to the progress in statistical methods aiming to overcome the weak statistical power expected to arise from single-variant associations involving rare alleles of low MAFs.^{78,80} These methods are mostly based on the grouping of rare variants, such as at the gene scale in collapsing tests, and assessing their combined effect. Examples of widely used statistic methods include the burden tests, which assesses the cumulative effects of variants via aggregation of rare variants within a genomic segment considered as a 'single burden variable' in the test statistics,^{85,86} variance-based sequence kernel association tests (SKAT), which has the advantage of detecting both the direction and magnitude of effect of a variant associated with a trait,⁸⁷ and multiple derived techniques.⁷⁸

Family-based sequencing studies

In a similar approach to genetic linkage studies, a cost-effective manner to optimize chances of identifying rare, highly penetrant variants yet within relatively small cohorts is to aggregate phenotypically-similar individuals (such as family probands) to identify rare variants through exome or genome sequencing.^{78,88} Here again, resequencing of clinical exome or genome sequencing data can be resourceful in the search of rare variants and disease markers,⁷⁷ and multiple techniques of genome-wide linkage analysis using family-based WGS have been described.⁸⁸ Moreover, rare-variant detection in genetic association studies can also be applied to family-based studies, using extended versions of statistical methods previously mentioned such as gene-based association methods including burden and kernel-based tests.^{89,90} An example of such study design is the gene-based association study of familial cutaneous melanoma,⁹¹ which identified strong signals in *CDKN2A*, *BAP1* and *EBF3*, loss of the latter correlating with cancer progression.

While familial studies can detect components of genetic predisposition to cancer, they are often incomplete in explaining the ‘missing heritability’ in cancer. Previously identified high penetrance susceptibility genes only account for a minority of the total heritable fraction of cancer, especially when only a few family members are affected.^{44,65,92} Recent extensive efforts to identify additional CSGs with high penetrance in breast and colorectal cancer (CRC) have shown that only a minority of other CSGs with relatively high penetrance are involved, such as *POLE* in CRC,⁹³ or *CHEK2* in breast cancer⁹⁴ but these do not compare to the high penetrance of mismatch repair genes or *BRCA1/2* in these diseases, and only account for <1% of familial risk.⁹² It is likely that the remaining inherited components of familial cancers involve multiple genes with smaller effects, as it is the case for multiple moderately penetrant CSGs subsequently identified. Linkage analysis is extremely limited in its application to multifactorial genetic traits, as they have smaller effects (low disease penetrance) and a weaker relationship with specific genomic loci, highlighting the need for other approaches.

MODERATE PENETRANCE - CANDIDATE GENE APPROACH

Relatively rare or uncommon variants of moderate penetrance typically have a MAF<2% and confer RRs of ~2–5,⁶⁵ although they do not constitute a homogeneous group and it is more likely that they occur on a continuum of allele frequency and RR. Moderately penetrant genes are often identified through direct interrogation of genes part of a biological pathway, mainly through linkage studies, with presumably known biological functions. Candidates may also be chosen based on known common cancer phenotypes (such as defective DNA repair or telomere shortening) or because somatic mutations of a particular gene have been identified.⁶⁷ This direct examination of biologically relevant candidates can be referred to as candidate gene analysis. Numerous candidate genes have been investigated using this method, which has proven to be quite inefficient in the identification of new CSGs, often lacking significant association (or study replication) with cancer predisposition.^{67,69,95} Likely explanations for the limits of this approach include the fact that (i) it is a largely speculative strategy whereas many biological processes underlying complex diseases are probably still unknown, and (ii) not many genes are expected to confer significant predisposing effects when searching for complex traits. Nevertheless, candidate-gene approach has still allowed the successful identification of multiple CSGs with moderate penetrance,⁹⁶⁻⁹⁹ and may continue to do so as biological mechanisms are unraveled. Two-staged studies combining

sequencing of selected individuals (such as members of a family with segregation of cancer, or individuals selected on a genetic criteria) followed by large-scale association studies have proven to be successful. This strategy was employed for example in the detection of pathogenic variants in *BRCA1/BRCA2*-negative breast and ovarian cancer patients¹⁰⁰ and in the identification of a new CSG, *RECQL*, in breast cancer susceptibility.¹⁰¹ Other main CSGs with moderate penetrance that were identified through candidate approaches include *ATM*,¹⁰² *PALB2* (partner of *BRCA2*),¹⁰³ and *RAD51D*.¹⁰⁴

Nevertheless, alleles that confer a RR of 2 typically have a MAF lower than 1%, and as for highly penetrant CSGs, only account for a small percentage of genetic inheritance in cancer. Overall cancer heritability is estimated at 33%,⁵⁵ and the combined effect of known CSGs with high and moderate penetrance may not account for more than 9% cancer risk, as determined in a large gene-panel study testing for germline mutations.¹⁰⁵ The recent PCAWG study estimated at 17% the proportion of cancer patients having a rare pathogenic germline variant in a CSG.²⁶ Although these numbers vary by cancer type and so does heritability, a large part of the missing heritability of cancer is thought to arise from the combination of common multiple alleles of weak penetrance.

WEAK PENETRANCE - CASE-CONTROL ASSOCIATION STUDIES

The implication of multiple common alleles of weak penetrance in cancer predisposition makes the search of polygenic components of cancer an attractive and highly informative strategy. In the current era of HTS, databases of cancer mutation catalogs, and large-scale sequencing projects,^{26,106-108} large population-based cohort studies can overcome the limits of linkage studies or candidate gene approaches in the complex identification of multiple variants of weak penetrance. The advantages of family-based studies for high to moderate penetrance CSGs do not apply here, as the low RR of common variants strongly decrease the likelihood of observing familial aggregation. Instead, large population-based case-control association studies provide higher statistical power required to detect association of these low-risk variants with disease.

Association studies represent an agnostic approach in the search of common genetic variation by comparing the frequency of many (thousands to millions) genetic variants between two groups of individuals, such as cancer cases and matched healthy individuals (controls). While

case-control studies can also be used in candidate-gene approaches, a main advantage is the ability to simultaneously interrogate many genomic locations throughout the entire genome (or exome), in a hypothesis-free approach. Genome-wide association studies (GWAS) became a highly utilized and successful tool in human genetic research starting from the mid-2000s, in cancer and other diseases. This was notably a consequence of improved insight regarding common polygenic variation coupled with technological improvements in genotyping, and the emerging knowledge of SNPs due to the launch of projects providing access to large databases of common sequence variation such as the Human Genome Project^{109,110} and the International HapMap Consortium.¹¹¹ Coupled with the development of increasingly performant genotyping products tailored to multiple applications,¹¹² association studies shifted from studying a limited number of candidate genes to the analysis of up to a few millions of SNPs throughout the whole genome. Another advantage of GWAS compared to candidate gene studies is that it also allows the identification non-coding risk variants, which in fact make up 90% of all GWAS variants.¹¹³ Although these are challenging to functionally address, they represent a complementary information to the identification of exonic sequences within CSGs and may help to unravel biological mechanisms in cancer.

GENOME-WIDE ASSOCIATION STUDIES

The final part of this introductory chapter to the genetic basis of cancer is dedicated to a more in-depth explanation of the concepts underlying GWAS study design and the broad range of associated analysis methodologies that currently drive the discovery of cancer susceptibility genes, their underlying functional variants and biological mechanisms that mediate risk.

GWAS STUDY DESIGN

GWAS studies test the preferential association of a considerable number of genetic variants with a particular trait. When this trait is a disease, genotyping data from a cohort of affected individuals (cases) is compared to that of an unaffected population (controls), as appropriately matched as possible to avoid false positives and biases from confounding factors. Variants (alleles) that occur more or less frequently within cases compared to controls are susceptibility alleles and represent genomic regions associated with risk. Commercially available genotyping platforms (such as SNP arrays) are used to genotype

germline DNA from patients and controls at specific SNP positions. These SNPs are most often common variants (i.e., with a MAF>5%) although association studies can now be used to detect variants from an entire frequency range.¹¹² The underlying concept is that alleles that belong to a same haploblock co-segregate together in a non-random way during mitotic and meiotic recombination events. A haploblock contains SNPs that are in high linkage disequilibrium (LD), and each underlying haplotype refers to the specific combination of alleles within the shared haploblock (Figure 4). LD can be affected by multiple factors, including population evolution, genetic drift, admixture, recombination and mutation history.^{65,114} It is commonly measured by the squared correlation between two alleles, r^2 . Microarray chips used in GWAS analyses make the most of this correlational genomic structure, as genotyping information of an affordable number of SNPs ('tag SNPs') are then sufficient to capture genetic variation of multiple other SNPs in LD. Genotype imputation¹¹⁵ can further fill some of the remaining gap of genetic variation through statistical inference from known LD and haplotype patterns, using panels from WES or WGS reference projects.^{116,117}) Although chip microarray followed by imputation is more commonly used to obtain genotyping data for association study, WGS is an alternative (and more costly) strategy offering higher coverage density and thus a higher MAF spectrum, notably useful for the detection of ultra-rare variants.¹¹²

Manhattan plots graphically summarize the association tests by plotting the statistical association of SNPs (individual dots) against their chromosomal position (Figure 4). Genome-wide significance threshold (horizontal line) is set at a particularly stringent P value, usually $P < 5 \times 10^{-8}$, accounting for a Bonferroni correction of multiple testing (very high number of individual association tests) and of uncertainty of potential false discovery of SNPs. GWAS meta-analyses are a common way to increase statistical power. Complementary to the P-value, the odds ratio (OR) is an important effect size statistic measuring of the strength of association between a genetic variant and outcome such as cancer.⁶⁵ GWAS studies carried out in the last ten years have led to the unanimous conclusion that complex traits are characterized by multiple risk loci, combining polymorphisms in many genes that each contribute to genetic variation, altogether explaining a significant portion of heritability.¹¹²

By design, GWAS studies identify multiple risk signals, each composed of multiple SNPs in LD, and do not indicate a particular gene target or mechanism that underly risk loci to explain the complex trait, nor do they evidence the combination of SNPs that confer significant disease risk. To face this challenging gap between GWAS output and biological processes, analysis methods have been developed and are rapidly evolving to address a wide range of aspects underlying association studies.¹¹² These methodologies serve different purposes, some of which are briefly outlined below along with a non-exhaustive list of the methods employed in each case.

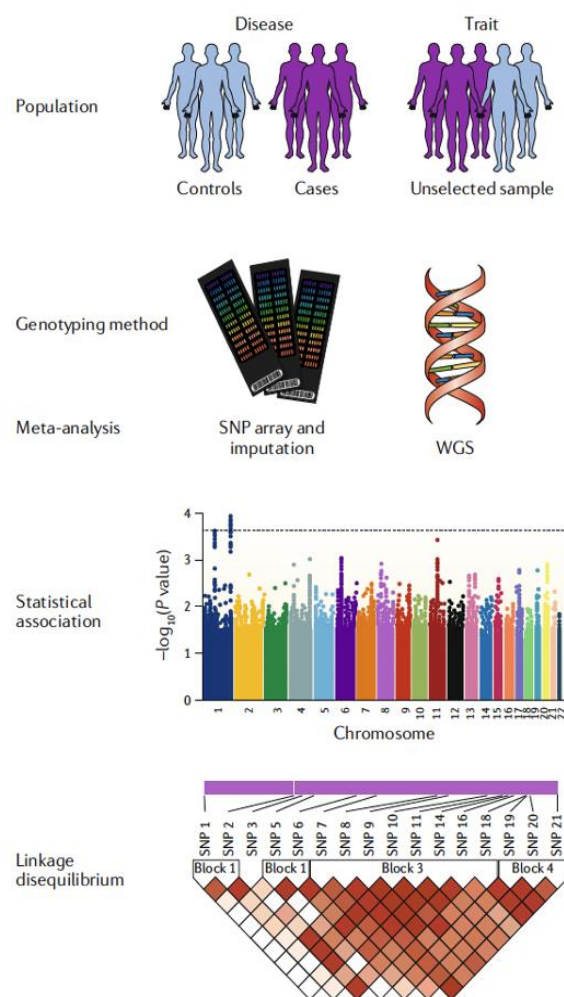


Figure 4. Genome-Wide Association Study design.

Genotyping data from an unselected population or two populations of cases and controls can be obtained from SNP arrays or whole-genome sequencing and tested through GWAS for association with a particular trait. Figure from Tam *et al.*¹¹⁸

METHODOLOGIES ASSOCIATED WITH GWAS STUDIES

Mixed-model analyses

Multiple models derived from the mixed linear model (MLM)¹¹⁹ were developed to better account for population structure and remove potential confounding factors in the association study. In this model, a genetic relationship matrix (GRM) is built to model sample structure, taking into account parameters such as population relatedness and geographic population structure (stratification), two main GWAS confounders.¹²⁰ The contribution of the GRM to phenotypic variance is taken into account by applying a correction to the statistics analysis, allowing to remove potential false positives due to population bias and resulting in increased statistical power. Since the initial introduction to MLM,¹²¹ variants of this method have been developed, accounting for factors such as phenotypic correlation in multi-trait models¹²² and more recently, adaptations to large-scale data.^{120,123}

Gene- and pathway-based association tests

Conventional GWAS is limited in its ability to address genetic and phenotypic heterogeneity in complex diseases as it identifies multiple small-effect common variants that individually do not account for a significant portion of heritability^{124,125} and that may be missed due to the marginal correlation between a single SNP and a trait, even with large sample sizes. Gene-based techniques overcome this problem by aggregating the effect of multiple variants within a gene into a single test,¹²⁴ which also enables detection of both common and rare variants. Enrichment analyses derived from this method group genetic variants within gene sets or pathways to identify biologically related variants that, when pooled together, reach the significance threshold and indicate enrichment of a particular pathway that otherwise would have been missed.¹²⁶ These methods focus on SNPs within coding regions, making it easier to infer biologically relevant disease pathways. Pathway analyses can be conducted following an agnostic approach (genome-wide pathway analysis, GWPA) or by testing hypotheses in candidate pathway analyses.¹²⁷

GWAS based on summary statistics

An incredible amount of GWAS summary statistics is now publicly available within databases, summarizing critical information from published GWAS studies, such as P-value and effect sizes (OR) of associated SNPs. These summary statistics not only provide valuable information for GWAS data interpretation and the genomics field, they can also be used as input in new

association studies.¹¹² Avoiding the need to perform individual GWAS studies, summary statistics-based association studies pool individual studies in meta-analyses, gaining statistical power. Summary statistics can be integrated with other available data, such as LD patterns and gene expression, and has proven successful in a wide range of fields, including identification of novel variants,^{128,129} estimations of SNP heritability and polygenicity,¹³⁰ quantification of pleiotropy and its impact on shared heritability across multiple traits^{131,132} and many others.¹¹² A typical example of integrative approach using summary statistics is combination of GWAS data to reference genotype-expression reference panels (eQTL, mentioned later) to unravel gene expression-trait associations in derivatives of transcriptome-wide association studies (TWAS).^{129,133,134}

Mendelian randomization

Analytical methods involving mendelian randomization (MR) allow to infer the causality of particular traits in the etiology of a disease.¹³⁵ By measuring the variation of known genetic variants and using them as instrumental variables, MR techniques can infer on which genetic variant (or multiple variants) have a causal effect on disease when combined to GWAS data, including summarized data.¹³⁶ Recent studies have combined MR to TWAS data (which unlike MR, does not estimate causality), thereby unraveling genetic determinants in multiple complex traits.¹³⁷

Polygenic risk scores

An important application of GWAS is that data linking genetic variants to a particular trait or disease can be gathered and used to predict risk of complex diseases.¹³⁸ A widely used GWAS-derived genetic predictor that captures cumulative common variation contributing to the risk of a disease is the polygenic risk score (PRS). PRSs are estimated from the coefficient (effect sizes) of individual alleles at multiple loci in a discovery series, which are then used in an independent set to calculate an overall risk score for each individual part of this independent series. Although they are currently mostly used in a research-based context,¹¹² PRSs can help bridge the gap between GWAS data and clinical relevance by predicting for example cancer subtypes (such as in breast cancer),¹³⁹ or providing elements for risk stratification and thus potentially clinical decision making,¹⁴⁰ and are thus being extensively tested for accurate prediction of multiple cancer types.^{141,142} By combining the weighted effect of individual

alleles on risk, PRS provides an aggregated risk value that can account for a significant portion of heritability, infer genetic overlap between traits (pleiotropy), but also predict molecular phenotypes such as gene expression.

As analysis methods continue to develop, the applications of GWAS will likely expand and bring further insights for clinical studies. However, since common genetic variation accounts for a significant portion of polygenic risk and since genetic variants identified by GWAS often fall within noncoding regions of the genome, assessing their functional impact is of utmost importance to understand biological mechanisms implicated in disease. While the “post-GWAS” era of functional evaluation of risk variants has not evolved at the speed of GWAS studies and currently faces challenges, advances in sequence-based analyses and large genomic projects now allow variant prioritization and functional testing of candidate variants.

POST-GWAS FUNCTIONAL STUDIES: CHALLENGES AND STRATEGIES

Only a small minority of GWAS studies so far have led to the characterization of causal variants and biological pathways underlying susceptibility loci.¹⁴³ Many reasons explain the difficulties in bridging the gap between genomic risk loci and functional mechanisms,^{65,113,118,143,144} which further hinders their translation into clinically-relevant information. First, most risk loci identified by GWAS fall within non-coding parts of the genome (such as introns or intergenic regions) and either contain multiple genes or are far away from the nearest known gene, making it challenging to identify target genes. Second, risk loci are often composed of dozens of SNPs in high LD, but the association test does not distinguish the functionally relevant polymorphism (functional SNP) from those that simply co-segregate due to physical proximity and co-inheritance from a same haplotype (Figure 5).¹⁴⁵ Finally, biological effects underlying a risk locus (such as gene regulation) in common polygenic diseases may be highly cell-type specific or driven by interactions between different cell types or variants.¹¹³

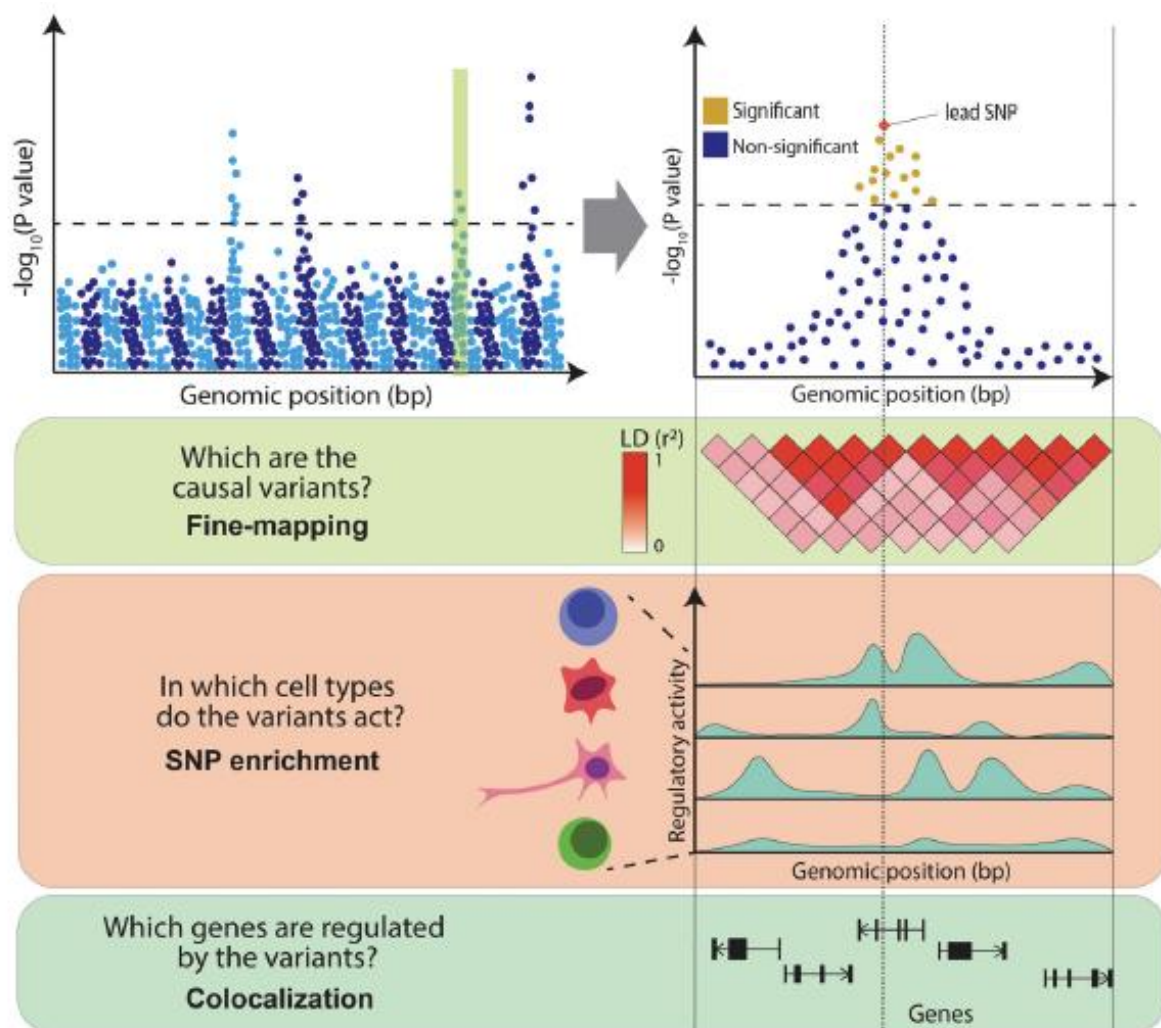


Figure 5. From GWAS to functional genomics.

Challenges underlying association loci pinpointed by GWAS studies include high LD between variants (in red, which high r^2 values), variable levels of regulatory activity among cell or tissue types, and multiple genes sharing a same risk locus. Figure from Cano-Gamez *et al.*¹¹³

Fortunately, many publicly available databases and statistical methods are now available to tackle (*in silico* and experimentally) the challenges in post-GWAS studies to answer the following questions: which are the functional variant(s), in which cellular or tissue context and by what mechanism do they mediate their effect, and which are the target genes regulated by the variants (Figure 5)? The process of refining GWAS association signals to localize and prioritize likely functional variants is referred to as fine-mapping. Epigenomic annotations from large databases (such as ENCODE¹⁴⁶ and RoadMap Epigenomics¹⁴⁷) allow

visualization of regulatory marks overlapping GWAS variants within their genomic environment, which can pinpoint variants that may have functional relevance.¹⁴⁸ As causal GWAS SNPs often accumulate within DNA regulatory elements,¹⁴⁹ looking at histone marks, chromatin accessibility marks, DNase I hypersensitive sites and transcription factor (TF) binding may be highly informative. Since functional non-coding variants often disrupt TF binding sites,¹⁵⁰ searching for putative TF binding sites through online databases like JASPAR¹⁵¹ and TRANSFAC¹⁵² can help prioritize variants likely to exert gene regulatory functions. Finally, integrative tools such as HaploReg¹⁵³ and RegulomeDB¹⁵⁴ can prioritize candidate variants by pooling genomic information from public databases.

Linking regulatory SNPs to gene expression or other genomic changes is particularly important to unravel disease pathways. Integration of SNP genotype with molecular phenotypes (such as gene/protein expression, methylation and splicing) can be assessed through quantitative trait loci (QTL) analyses. Gene expression QTLs (eQTLs) are commonly used (Figure 6), in large part due to the Genotype-Tissue expression project (GTEx)¹⁵⁵ database encompassing large datasets from over 50 tissues. An extension of this method is the previously mentioned combination of GWAS dataset with an eQTL reference panel (TWAS), allowing the comparison of inferred gene expression between cases and controls and directionality of association.¹¹²

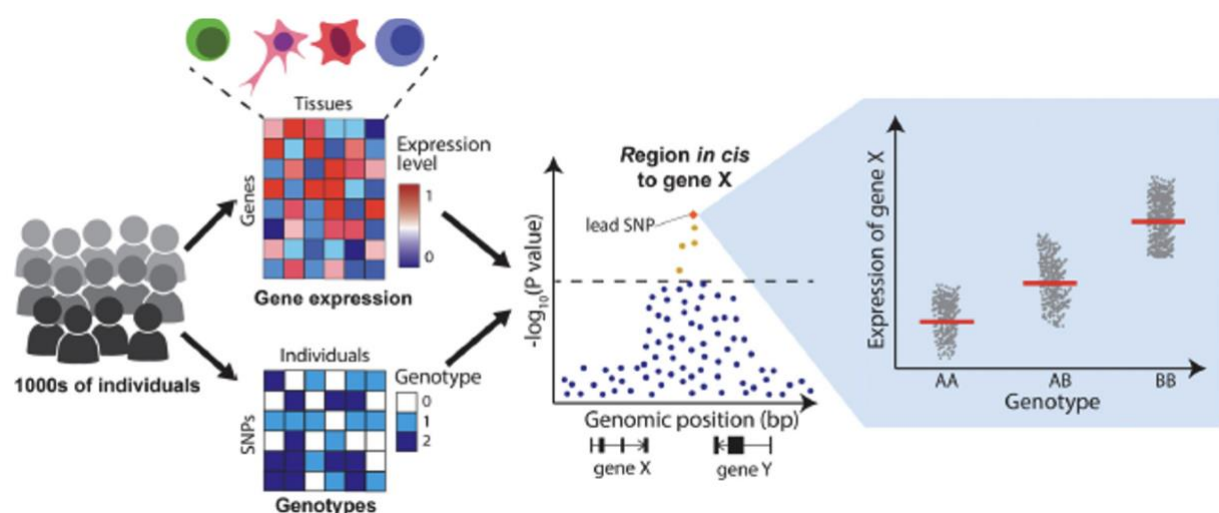


Figure 6. Principles of expression quantitative trait loci (eQTL) mapping.

Individuals for which both genotyping and gene expression information is available can be subjected to GWAS studies followed by eQTL mapping, allowing to test SNPs significantly associated with a trait (in yellow/orange) for gene regulatory effects according to their genotype. Figure from Cano-Gamez *et al.*¹¹³

In silico fine-mapping methods are critical for variant prioritization, yet experimental functional follow-up is most often a requirement to unambiguously validate a functional SNP and establish causality with the complex trait. A wide range of techniques are available to test the effect of SNPs within *cis*-regulatory elements, such as the following: approaches testing allele-specific gene regulation (cell-based reporter assays, their high-throughput counterpart massively parallel reporter assays (MPRA),¹⁵⁶ clustered regularly interspaced short palindromic repeats (CRISPR) gene editing techniques); approaches determining the molecular function and binding partners of variants (electrophoretic mobility shift assays (EMSAs), proteomic approaches such as quantitative mass spectrometry); approaches evaluating gene-gene long-range interactions (4C and Hi-C derived techniques). For an exhaustive list of available techniques to decipher risk loci, please refer to Gallagher *et al*¹⁴³ or Sud and colleagues.⁶⁵

SUSCEPTIBILITY LOCI IN CANCER

There are currently over 400 reported associations of genomic regions with cancer,⁶⁵ mostly coming from studies of European populations. Breast,¹⁵⁷ ovarian¹⁵⁸ and prostate¹⁵⁹ cancers are so far the tumor types with the highest number of identified susceptibility loci, which does not only reflect the statistical power of underlying GWAS studies (some breast and prostate cancer studies now include over 100,000 individuals), but also represents the variable heritability of different tumor types. On the other end of the spectrum of common cancers, large-scale association studies in lung cancer¹⁶⁰ have “only” revealed 18 susceptibility loci (90 in breast¹⁵⁷), which highlights the significant contribution of non-genetic factors in this cancer, such as smoking. Major GWAS studies and meta-analyses were also reported for pancreatic,¹⁶¹⁻¹⁶³ colorectal,^{164,165} renal,^{166,167} esophageal,¹⁶⁸ thyroid cancers,^{169,170} cutaneous melanoma^{171,172} and many others, including hematological malignancies and pediatric tumors⁶⁵ (for an extensive list of identified susceptibility loci, genes and SNPs in a large set of tumors, please refer to Liang *et al*¹⁷³). In non-European populations, cancer GWAS studies are sparser¹⁷⁴ with only 11% samples coming from East Asians, 4% Africans and 1% of Latin American ancestry.

Multiple of the previously mentioned GWAS-associated methods have been used in cancer research, such as genome-wide pathway analysis in renal cell carcinoma (RCC), which recently

implicated the PI3K/AKT pathway as an RCC heritability trait.¹⁷⁵ Another successful approach is the combination of different GWAS studies studying association with multiple phenotypes other than cancer, such as traits thought to play a role in a particular cancer type; this was recently done by Landi and colleagues¹⁷¹ who combined meta-analysis of cutaneous melanoma with pigmentation traits (GWAS on nevus count and hair color), identifying secondary loci and potentially important new pathways in melanoma pathogenesis.

PLEIOTROPIC CANCER RISK LOCI

Around a third of risk variants identified are associated with multiple tumor types, indicating a pleiotropic effect.⁶⁵ This has become increasingly clear with the emergence of GWAS studies looking at correlations between pairs of cancer types and pan-cancer studies of up to 18 distinct tumor types.¹⁷⁶⁻¹⁷⁹ Significant correlations between susceptibility loci were found for colorectal, pancreatic and lung cancers, kidney and testicular cancers, and many others. Similar studies in rare cancers have not been conducted and have yet to be reported. A major goal in pinpointing these cancer hotspots is to identify shared genetic and molecular mechanisms of potential relevance to known hallmarks of cancer. Risk loci identified in multiple cancer types include *MDM4* (chromosome 1q32),^{180,181} *TET2* (4q24),^{157,182} *TERT/CLPTM1L* (5p15),¹⁸³⁻¹⁸⁶ *CDKN2A/CDKN2B* (9p21)^{187,188} and multiple others. Pathway analysis of risk loci across cancers can highlight cancer-specific mechanisms; a nice illustration of this point is the identification of 7 independent telomere-related loci¹⁸⁹ across different cancer types, 5 of which map to genes known to be involved in telomere biology and maintenance (*TERC* at 3q26, *TERT* at 5p15, *OBFC1* at 10q24, *RTEL1* at 20q13 and *NAF1* at 4q32). 5p15 *TERT* SNP rs2736100 is notably a risk variant from multiple cancers including glioma, lung cancer, melanoma and others.⁶⁵ Applying this pathway-based strategy across other GWAS risk loci shared between cancers may be highly instructive in the understanding of biological mechanisms of carcinogenesis .

MELANOCYTES AND MELANOGENESIS

The work presented in this thesis focuses on uveal melanoma, which arises from the transformation of melanocytes of the eye. To better understand mechanisms the function of uveal melanocytes and eye pigmentation, some of the key processes of melanocytogenesis and melanogenesis are first introduced here.

MELANOCYTES

Human melanocytes are derived from the neural crest, from which their undifferentiated, non-pigmented precursor cells melanoblasts migrate at different times after closure of the neural tube towards multiple parts of the body where they differentiate into melanocytes, start producing melanin, and exert different functions depending on the site. Other than in the epidermis, hair follicles and the eye, melanocytes are also distributed in the stria vascularis of the cochlea where they are involved in equilibrium and hearing,^{190,191} in the brain where they are associated with neuroendocrine functions and where neuromelanin protects against oxidative damage,¹⁹² in the heart where they are thought to contribute to the mechanical properties of the heart valve,¹⁹³ and in the meninges¹⁹⁴ where their role is currently unknown. The primary function of the melanocytes is to produce melanin, which gives their pigmented color to skin, hair and eyes, although a small number of other cell types also have the ability to produce melanin, such as epithelia of the iris and retina of the eye, some neuronal cells and adipocytes.^{195,196}

Studies on embryology, development of the melanocytic lineage and melanogenesis have mainly been carried out for skin melanocytes, and therefore what is currently known about the differentiation and biological function of melanocytes is essentially derived from skin melanocytes and epidermal melanin. We will therefore take epidermal melanocytes as a model for melanocytic embryology, development, and melanin synthesis before stating what is currently known regarding uveal melanocytes and iris pigmentation.

EMBRYOLOGY AND MELANOCYTOGENESIS

Developmental studies in mice have led to most of today's understanding of melanocyte development, although work on chicken and zebrafish models also provided important embryological and genetic insights, respectively.¹⁹⁷⁻²⁰² Melanocytogenesis and melanogenesis involve a succession of steps including development of melanocyte precursors (melanoblasts) in the neural crest, migration to peripheral sites and differentiation into melanocytes, proliferation and maturation of melanocytes at the target place, where they start to synthesize melanin within maturing melanosomes, their melanocyte-specific organelle.

Neural crest cells (NCCs) are pluripotent cells that initially reside in the neural tube, a structure derived from the differentiation of the neuroectoderm, from which they migrate into the developing embryo. NCCs originating from the trunk give rise to melanocytes, but also other cell types including neuronal and glial cells in the peripheral nervous system (PNS). Cells of the PNS including neurons and Schwann cells arise from NCC migration from the dorsoventral pathway while the dorsolateral pathway is the main source of melanocytes, starting with the specification of their melanoblast precursors on embryonic days 8.5 to 9.5 in the trunk region,²⁰³ although there are conflicting results as to when NCCs first acquire melanoblasts features.²⁰⁴

From SOX10-positive melanoblast precursors to mature melanoblasts, to their subsequent migration to various sites (skin, hair follicles, eye, meninges, ear and mucous membranes) leading to differentiated melanocytes, most knowledge comes from research on epidermal melanoblasts. Melanoblasts proliferate and differentiate into melanocytes following a cascade of molecular events in which important receptors (such as c-kit), transcription factors (SOX10, PAX3, MITF) and melanogenic enzymes (TYR, TYRP2) play a role.²⁰⁴ The MAPK-signaling pathway is required in early melanoblast development and migration, during which ephrins and stem cell factor (SCF) are key growth factors, while Notch- and Wnt-signaling pathways are essential for subsequent melanocyte development, and participate in the regulation of key transcription factors. In animal models, expression of *MITF* has been shown to play a key role in the migration of melanoblasts, along with *KIT*, *EDN3* and *EDBNRB2*.²⁰³ In addition to its role in melanoblast migration, MITF is often also seen as the master regulator

of melanocyte development, including proliferation, differentiation, and survival. Its sustained expression activates proteins that are key to melanocyte survival (such as BCL-2) and function (such as TYR for melanin synthesis and genes required for the formation of melanosomes).

Timing of migration and specification of NC-derived cells to their final site (i.e., their fate) is still unclear. Some models support the presence of multipotent progenitors with stem cell properties in early neural crest population that subsequently differentiate according to different environmental exposures,²⁰⁵ while others suggest an acquisition of cell fate earlier in development before leaving the neural tube.²⁰⁶ Some existing data support both hypotheses and suggest the two models are complementary. The migration of melanoblasts to their final site is determined by a new phase of proliferation and differentiation into melanocytes between embryonic days 8 and 14.²⁰⁴ Melanoblast specification to the epidermis is driven by expression of KIT, MITF and DCT, and following colonization of hair follicles, a portion dedifferentiate by losing MITF and KIT expression to form a nest of melanocyte stem cells that can later on replenish hair follicles in pigment-producing cells.^{197,204}

MELANOGENESIS

Melanogenesis refers to the process by which the melanin pigment is synthesized, which is the main function of melanocytes. This predominantly occurs within melanosomes, subcellular lysosome-like organelles that develop within the cytoplasm of melanocytes.²⁰⁷ Melanin is synthesized and packaged within granules in the final steps of melanosome maturation. Within the epidermis and hair follicles, melanin is subsequently transferred from mature melanosomes into neighboring keratinocytes.²⁰⁴ The two major types of melanin that are produced within the melanosomes of melanocytes are brown-black pigmented eumelanin and red to yellow pheomelanin.²⁰⁸ The size, shape and packaging of melanin granules differ between the two pigments,²⁰⁹ where eumelanin is more densely packed than pheomelanin.

In mammals, synthesis of these two pigments is coordinated by a cascade of tightly regulated and interacting pathways involving a multi-enzyme complex, including three major enzymes,

tyrosinase (TYR), tyrosinase-related protein 1 (TYRP1) and 2 (TYRP2) (Figure 7).²¹⁰ Both

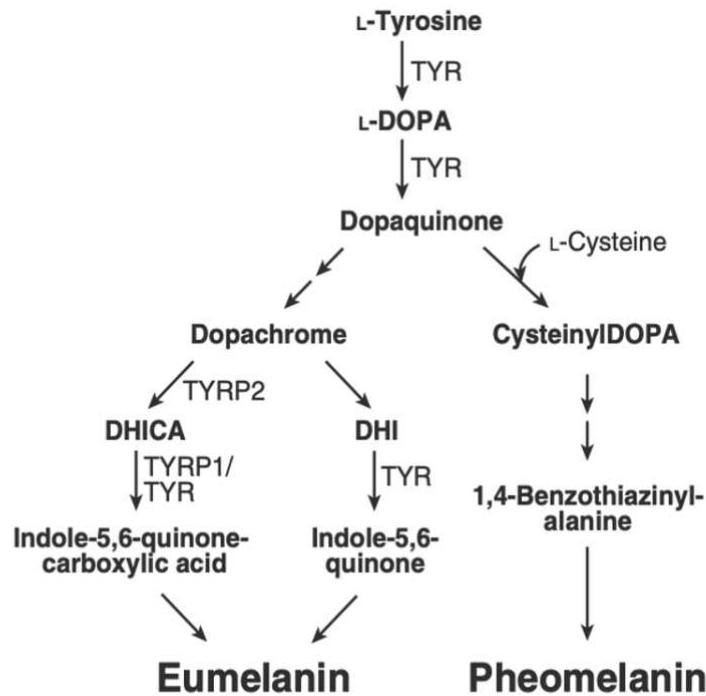


Figure 7. Melanogenesis pathways leading to the synthesis of pheomelanin and eumelanin in melanosomes. Tyrosinase enzyme (TYR) is the rate-limiting enzyme in the synthesis of melanin precursors L-DOPA and dopaquinone. In the absence of L-cysteine, dopaquinone will be converted to eumelanin (black-brown pigment) in successive steps involving TYRP2, TYRP1 and TYR enzymes; if cysteine is present, dopaquinone will instead be converted into pheomelanin (yellow-red pigment). Figure from Wiriyasermkul et al.²⁰⁷

melanin types are synthesized from a common dopaquinone precursor, itself a product of the

rate-limiting enzymatic activity of TYR in the hydrolysis of L-tyrosine to L-DOPA followed by the oxidation of L-DOPA to dopaquinone. Pathways then diverge depending on the availability of substrates and implication of downstream enzymes to synthesize the two melanin types. On one hand, the extensively studied eumelanin synthesis pathway is mainly driven by enzymatic cascades involving TYRP1, TYRP2 and TYR; these three enzymes physiologically interact with one another, and disruption of either of them lead to serious pigmentation defects. Importantly, eumelanin precursor dopachrome increases TYR activity, which in turn increases L-DOPA precursor synthesis, thereby regulating melanin synthesis in a positive feedback loop mechanism controlling homeostasis.²¹¹ On the other hand, pheomelanin synthesis relies on the availability of sulfhydryl compounds in melanosomes, mainly L-cysteine, which is not required for eumelanin synthesis. L-cysteine results in a

preferential conversion of dopaquinone to cysteinylDOPA instead of dopachrome, ultimately yielding pheomelanin. The switch between the two pathways is particularly important as the ratio of pheomelanin to eumelanin as well as the amount of eumelanin are major determinants of skin and hair pigmentation.²⁰⁸ This ratio is determined by substrate availability (tyrosine and cysteine) and TYR activity,²¹² although TYRP1 has also been suggested to increase the eumelanin : pheomelanin ratio. A higher amount of melanin and particularly eumelanin will result in darker skin, black or brown hair, while individuals with a higher ratio of pheomelanin to eumelanin will have higher chances of having lighter skin, blonder or red hair and freckles.

Multiple molecular pathways influence the production of melanin, through signaling cascades and transcription factors that regulate the expression of key genes implicated in the melanogenesis pathway. A major regulator of pigmentation is melanocortin-1 receptor (MC1R), a G-protein coupled receptor that induces the production of cyclic AMP (cAMP) upon activation by an agonist, which in turn leads to the transcriptional activation of key transcription factors including microphthalmia transcription factor (MITF), a master regulator of pigmentation and melanocyte development.²⁰⁹ MITF has binding sites on many pigmentation genes, recognizing the so-called E-box sequences²⁰⁸ and inducing the transcription of multiple genes, including those involved in melanin production (such as *TYR* coding for the tyrosinase enzyme, *TYRP1* and *TYRP2*), melanocyte differentiation (such as *PMEL*), but also cell survival, proliferation and motility, making it a central regulator of melanogenesis.

Binding of different agonists of MC1R lead to the preferential synthesis one of the two melanin types, participating in their regulation. As such, in epidermal melanocytes, α -melanocyte stimulating hormone (α -MSH) and adrenocorticotrophic hormone (ACTH) bind MC1R as agonists and stimulate eumelanin synthesis via upregulation of cAMP, resulting in a switch in production from pheomelanin to eumelanin, while MC1R antagonist agouti signaling protein (ASP) inhibits binding of α -MSH, preventing downstream expression of MITF and subsequent activation of TYR, TYRP1 and TYRP2, thus leading to the switch from eumelanin to pheomelanin synthesis.^{213,214}

Although the above-mentioned enzymes, receptors and transcription factors are major regulators of melanin synthesis, including the switch between the production of eumelanin

and pheomelanin which is of interest for the present work, the general regulation of melanogenesis is driven by multiple other signaling pathways, mostly MITF-dependent, including MAPK-ERK, CREB, Wnt and PKC-dependent pathways but also the SOX family of transcription factors in development of melanocytes. For an extensive review, please see Harris *et al.*²¹⁵ and D'Mello *et al.*²⁰⁸

Finally, melanin synthesis is a process that is tightly linked to melanosome development. Melanosome biogenesis is composed of four stages.²⁰⁷ Stage I premelanosomes (non-pigmented precursors of melanosomes) develop from the early endosome and start producing fibrils composed of the melanocyte protein PMEL. In stage II, fibrils are elongated and organize themselves as parallel sheets within an ellipsoidal-shaped premelanosome. Melanosomes are then formed from their precursors in stage III, when melanin synthesis begins, orchestrated by the translocation of critical melanogenic enzymes TYR, TYRP1 and TYRP2, but also membrane ion transport proteins OCA2, SLC45A2, ATP7A and TPC2, from the Trans-Golgi network to melanosomes; the pH change in stages III-IV from acidic to neutral allow TYR enzyme to be functional and start synthesizing melanin, which is progressively deposited on the matrix of PMEL fibrils, which are fully covered in stage IV. The resulting darkened, mature melanosome can then be secreted to keratinocytes, resulting in the spread of color throughout the skin and hair.²⁰⁷

FUNCTION OF CUTANEOUS MELANOCYTES

The transfer of melanin-containing melanosomes to adjacent keratinocytes allowing their homogeneous distribution throughout the skin is a major mechanism of protection against ultra-violet radiation (UVR) induced by exposure to sunlight. UVR exposure causes cell damage through carcinogenic actions of UVA, resulting in the generation of reactive oxygen species (ROS), and UVB, through direct DNA mutagenesis.²¹³ Skin pigmentation is one of the endogenous mechanisms that has evolved to protect against UVR damage, as the melanin granules that are transferred from skin melanocytes accumulate within keratinocytes where they readily absorb UVR, protecting deeper skin layers (dermis) from damage. It is therefore not surprising that melanogenesis in the skin is enhanced after UVR exposure, resulting in darker (tanned) skin. Melanin has also been shown to have antioxidant properties and to act

as a free-radical scavenger.^{209,216} More specifically, eumelanin, more densely packed than its lighter counterpart, is much more protective against UVR and ROS than pheomelanin; some studies even describe eumelanin as the sole photoprotective agent while pheomelanin is phototoxic, pro-oxidant and further generates ROS notably by stimulating lipid peroxidation in UV-dependent and -independent pathways.^{207,217,218} Consequently, individuals with light skin (low eumelanin levels) are significantly more at risk of cutaneous melanoma (CM) than those with darker skin,²¹⁹ and although the relationship between UVR exposure and CM is still not fully elucidated,²²⁰ exposure to sun and subsequent sunburns are a strong risk factor for CM.²¹⁹

UVEAL MELANOCYTES

The wall of the human eye is made of three layers: the cornea and sclera on the outer layer, the uveal tract and the retina (Figure 8). The uveal tract is a highly vascularized and pigmented layer further composed of three parts, the iris (most anterior part), the ciliary body, and the choroid in the most posterior part. The latter is adjacent to the retina, directly supporting and nourishing it with diffusible nutrients and oxygen. The retina is further subdivided in two components, the neural retina mostly implicated in visual transduction, and the retinal pigment epithelium (RPE). The RPE (along with iris and ciliary pigment epithelia contiguous with RPE) and the uveal melanocytes are the two types of cells that provide pigmentation to the eye, although they differ in many aspects: RPE cells originate from the neural ectoderm, stop producing melanosomes early in childhood and therefore adult RPE cells do not produce

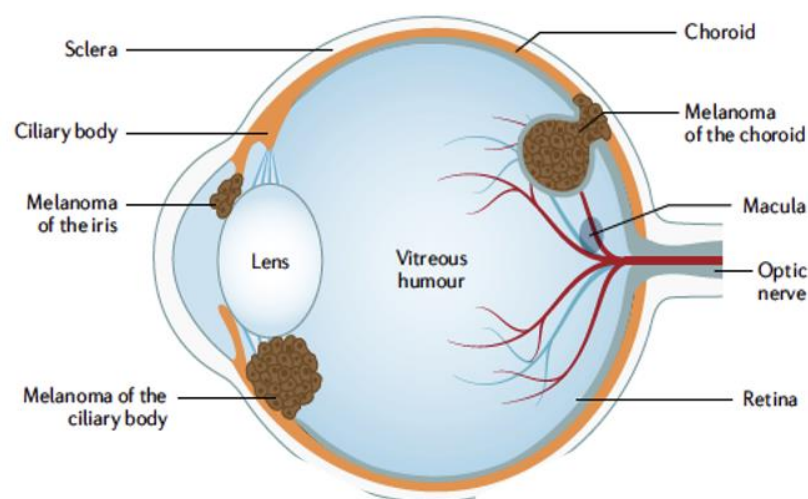


Figure 8. Anatomy of the eye showing the three different types of uveal melanoma: melanoma of the choroid, ciliary body, and iris.

Figure from Jager *et al.*²²³

melanin.²²¹ Uveal melanocytes, however, are derived from the neural crest and maintain a constant production of melanin throughout life.²²²

Despite their common neural crest origin and melanoblast precursor from an embryonic point of view, uveal melanocytes largely differ from their skin counterparts not only in terms of biological behavior and function, but also in terms of etiopathogenesis upon neoplastic transformation of their respective cells of origin, uveal and cutaneous melanocytes, leading to two very distinct types of melanomas on a biological, genetic (driver genes) and clinical point of view.²²⁴⁻²²⁶ While both cutaneous and uveal melanocytes are derived from the NCCs arising from the cervical trunk and midbrain-hindbrain junction within the neural crest,¹⁹⁸ adult uveal melanocytes are thought to follow their own developmental pathway, migrate at a different time from the neural crest, follow a unique pre-migration specification, have a mesodermal destination, and display different patterns of tissue and gene expression.²²⁷⁻²²⁹ Uveal melanocytes migrate towards the eye's uveal tract (choroid and stroma of the iris and of the ciliary body) where they start producing pigment. Uveal melanogenesis is thought to occur between the 20th week of embryonic development and 6 months after birth, explaining why iris color may vary until that age.

Although the eumelanin : pheomelanin ratio in iridal melanocytes influences eye color in a similar fashion to skin or hair pigmentation (detailed in the next section “determinants of eye pigmentation”), major differences in biological characteristics of uveal and cutaneous melanocytes exist. Unlike cutaneous melanocytes, uveal melanocytes are only in contact with each other, they do not transfer their pigment to keratinocytes, and they do not adapt their pigment content to environmental stress such as UVR.

In the skin, epidermal melanocytes protect deeper layers from UVR damage resulting from sun exposure, as previously seen; the same mechanism has been proposed in the eye, where uveal melanocytes would protect the most posterior parts of the eye from UVR damage. However, (i) the eye behaves as a “camera obscura” allowing little sunlight to enter through the pupil to the inner eye, and the most posterior parts are not exposed to sunlight;^{230,231} (ii) unlike in the skin, uveal melanocytes do not transfer their melanin pigment to neighboring cells to enhance pigment density as a photoprotective mechanism, and the iris color does not

change following exposure to sunlight; and (iii) RPE, the most fragile layer, is anterior to uvea and adult RPE cells do not produce melanin.

Apart from the iris where melanocytes are thought to confer photoprotective effects (via light absorption, scattering, and quenching of ROS) due to direct exposure of the anterior part of the eye to sunlight,^{204,217,229} the function of posterior uveal melanocytes (choroid and ciliary body) that are not exposed to UVR have yet to be elucidated. Previous studies hypothesized some functions, such as a role for choroidal melanocytes as free radical scavengers as they are thought to be exposed to high oxidative tension.^{195,232} More recently, roles for uveal melanocytes in maintaining general homeostasis of the eye through immune regulation, inflammatory and angiogenic processes have been suggested,²³³ but future work is needed to fully elucidate these mechanisms.

DETERMINANTS OF EYE PIGMENTATION

The role of uveal melanocytes and genetic factors influencing the determination of eye color are of particular interest in the context of this work as the latter is a risk factor for uveal melanoma, as discussed in the next chapter.

Similarly to skin and hair pigmentation, iris color is influenced by both by the quantity and quality of melanin within uveal melanocytes, where higher amounts of eumelanin and a higher ratio of eumelanin to pheomelanin leads to darker irises (black/brown) than relatively lower eumelanin to pheomelanin ratios associated with a lighter eye color (blue).²³⁴ Differences in eye color are not associated to a difference in the number of melanocytes, which remains constant between iris color groups,²³⁵ but rather to differences in melanosome composition and melanocyte structure. It is thought that melanosomes are composed of an inner pheomelanin core surrounded by an eumelanin outer core, whose thickness impacts eye color.²¹³ However, MC1R and its agonist MSH are not expressed in uveal melanocytes and are therefore not thought to play a role in the switch between eumelanin and pheomelanin synthesis in the eye, unlike in skin and hair tissues in which MC1R influences pigmentation.²³⁶ Instead, two main genes have been shown to majorly influence iris color: *OCA2* and *HERC2* on chromosome 15.^{237,238} Individuals with blue or hazel eye color typically have specific polymorphisms within these two genes, the most critical ones being the “blue-eye” alleles of

rs12913832 in *HERC2* and of rs1800407 in *OCA2*, which were further shown to strongly determine eye color via epistatic effect through gene-gene interactions.^{238,239} *OCA2*, or oculocutaneous albinism II, encodes a protein sometimes referred to as the P protein. As previously discussed, *OCA2* is a membrane-associated ion transport protein majorly implicated in melanogenesis and more particularly in the melanosome maturation process.²⁰⁷ *OCA2* is translocated from trans-Golgi network during stage III of melanosome formation, along with other key membrane transport proteins SLC45A2, TPC2 and ATP7A and core melanin-synthesizing enzymes TYR, TYRP1 and TYRP2, a critical step in melanosome biogenesis during which melanin synthesis is initiated. Polymorphisms or deficiencies in any of the membrane transport proteins mentioned can result in oculocutaneous albinism types I to IV²¹⁰ (type II specifically caused by defects of *OCA2*) that involve pigmentation defects of the eye, skin and hair.

HERC2 (and more particularly its 86th intronic region) lies within an enhancer region that regulates *OCA2* expression. The *HERC2* protein itself is not implicated in pigmentation, but its deletion causes hypopigmentation due to its regulatory activity on *OCA2* synthesis.²³⁸ In individuals with *HERC2* rs12913832 “blue eye” (C) polymorphism, chromatin loop formation between this enhancer and *OCA2* promoter sequence is strongly attenuated, resulting in reduced *OCA2* expression.²⁴⁰ Decreased levels of *OCA2* in turn lead to compromised synthesis of the darker pigment, eumelanin, without affecting pheomelanin synthesis and thus resulting in lighter pigment and thus lighter eye color. This is thought to occur via an accumulation of tyrosinase enzyme TYR that inhibits eumelanin synthesis pathway without affecting pheomelanin.²⁴¹

Although these two genes are the main determinants of iris color, polymorphisms within other genes also influence the amount of melanin and eumelanin/pheomelanin ratio. The IrisPlex System is a tool that was developed to accurately predict blue and brown eye color based on the genotyping information of 6 SNPs: already mentioned rs12913832 on *HERC2* and rs1800407 on *OCA2*, to which were added *IRF4* rs12203592, *TYR* rs1393350, *SLC24A4* rs12896399 and *SLC45A2* rs16891982.²⁴² The system has since then been optimized to simultaneously determine eye, hair and skin color from 41 predictive SNPs (17 for skin and 24 for simultaneous eye and hair color prediction).^{243,244}

POPULATION GENETICS AND PIGMENTATION

As a closing remark, population genetics have extensively been studied to decipher the genotypes underlying variations in distribution and frequency of different phenotypes across the globe. In addition to other phenotypes that account for geographic variation of human traits, such as the immune system (linked to various distribution of alleles of the human leukocyte antigen – HLA – complex),²⁴⁵ pigmentation traits (including skin, hair and eye color) vary dramatically in human populations both temporally and geographically.²⁴⁶ These pigmentation differences are tightly linked to melanin and melanogenic genes previously stated, and most probably arise from a combination of environmental selective pressures leading to genetic drift.

There are multiple hypotheses to explain the heterogeneous spread of light skin in European populations in “recent” (~65,000–100,000 years ago) evolution, and of light eye color in a subset of these populations (mostly Northern and Eastern Europe), including sexual selection and environmental factors linked to varying UVR exposure levels.²¹⁷ The former suggests that hair and eye polymorphisms were strong determinants of sexual selection in populations migrating from African to Europe, due to the rarity of light eye color and skin depigmentation, creating a “bottleneck” in Northern Europe of individuals with blue eyes.²⁴⁷ Second, natural selection of lighter skin color during evolution when populations moved from Africa to Europe and Asia is also linked to varying exposures to UVR and thus the need to optimize the synthesis of vitamin D in countries receiving less sunlight, as its deficiency could lead to osteomalacia and rickets that could impair human fertility and reproduction success.²⁴⁸ Vitamin D can therefore be seen as a selective pressure mechanism in evolution. Pheomelanin being more loosely packed than eumelanin, it allows more UVR to penetrate the skin (and eye), which may explain a higher synthesis of vitamin D with people of lighter skin and eye color, and further explain the evolution of pigmentation genes and traits in Northern Europe, where there is less sunlight exposure, to adapt their vitamin D synthesis.²¹⁷

UVEAL MELANOMA

This chapter will provide a general overview of uveal melanoma (UM), introducing some major clinical and genomic aspects as well as malignant transformation and treatment strategies, before focusing on UM epidemiology and genetic risk factors.

CLINICAL ASPECTS

Uveal melanoma (UM) is the most common form of primary intraocular malignancy in adults, although it is a rare cancer with a stable incidence of 5.1 per million in the United States.²⁴⁹ While it represents 3 to 5% of all melanomas, UM is very different from cutaneous melanoma (CM) and other melanoma subtypes in terms of etiology, genetics, biology, epidemiology and clinical features.^{224,226} UM arises from the malignant transformation of melanocytes of the uvea, itself composed of the choroid, ciliary body and iris. Choroidal melanoma is the most common form of UM (90%), far more frequent than melanomas of the ciliary body (6%) and of the iris (4%).²⁵⁰ Symptoms include blurred vision and visual field loss, although a high proportion of UMs are asymptomatic, often resulting in late diagnosis also due to the fact that tumors initially grow in areas of the eye (between the sclera and the retina) that are not visible at the surface of the eyeball. Although primary tumors are most often successfully treated by brachytherapy, external radiation therapy (including proton beam therapy) or enucleation in case of excessively large tumors, metastases eventually occur in the following years between 30 and 50% of cases, predominantly in the liver (90% of UM cases) via dissemination through blood vessels, and remain without effective therapy until very recently, resulting in a dismal prognosis and a median survival of 9 to 16 months.²⁵¹⁻²⁵³ Most UMs are unilateral and unifocal, although exceedingly rare bilateral cases do arise (1 in 50 million estimated lifetime risk), sometimes decades after UM in the first eye, and the risk of developing bilateral UM after a first occurrence in one eye is estimated at 0.2%.^{254,255} Unilateral multifocal UMs (occurrence of a second tumor) are even rarer and are often associated with ocular melanocytosis or germline mutations of *BAP1*, both strong UM risk factors discussed later.²⁵⁶ UM tumors vary in location, size, shape and color, which impacts their diagnosis.²⁵⁷ Pigmentation of UM tumors varies from non-pigmented to very strongly pigmented, and pigmentation levels within a same tumor can also be heterogeneous. While

melanosomes within melanotic UM produce both pigment types (eumelanin and pheomelanin), amelanotic tumors are characterized by a weak tyrosinase (TYR) activity and exclusively produce pheomelanin.²¹⁷

GENOMIC LANDSCAPE OF UVEAL MELANOMA

From a genetic point of view, UM tumors are remarkably stable, characterized by very few structural variants compared to other melanoma subtypes and a strikingly low tumor mutation burden, with a markedly lower single nucleotide variant (SNV) average mutation rate of 0.5 per megabase (Mb)²⁵⁸ and among the lowest number of somatic SNVs compared to other cancer types (Figure 9A&B).²⁵⁹ In addition, while a controversial role for ultra-violet radiation (UVR) exposure was suggested in UM development (later discussed in “UM epidemiology” and “environmental risk factors” sub-chapters), Furney and colleagues’ work and other main WGS studies in UM demonstrated a clear absence of UV-induced mutational signature (Figure 9C&D),²⁵⁹⁻²⁶¹ classically characterized by C>T transitions in a pyrimidine dinucleotide context²⁶² and representing 80 to 90% of mutations in CM.^{263,264} A notable exception, developed later among the list of environmental UM risk factors, is the strong prevalence of UV mutational signature in iris melanoma,²⁶⁵ itself accounting for a small minority of UM cases. This is due to the direct exposure to sunlight of the iris compared to more posterior parts of the eye, resulting in major genetic and genomic differences between anterior and posterior UM subtypes.

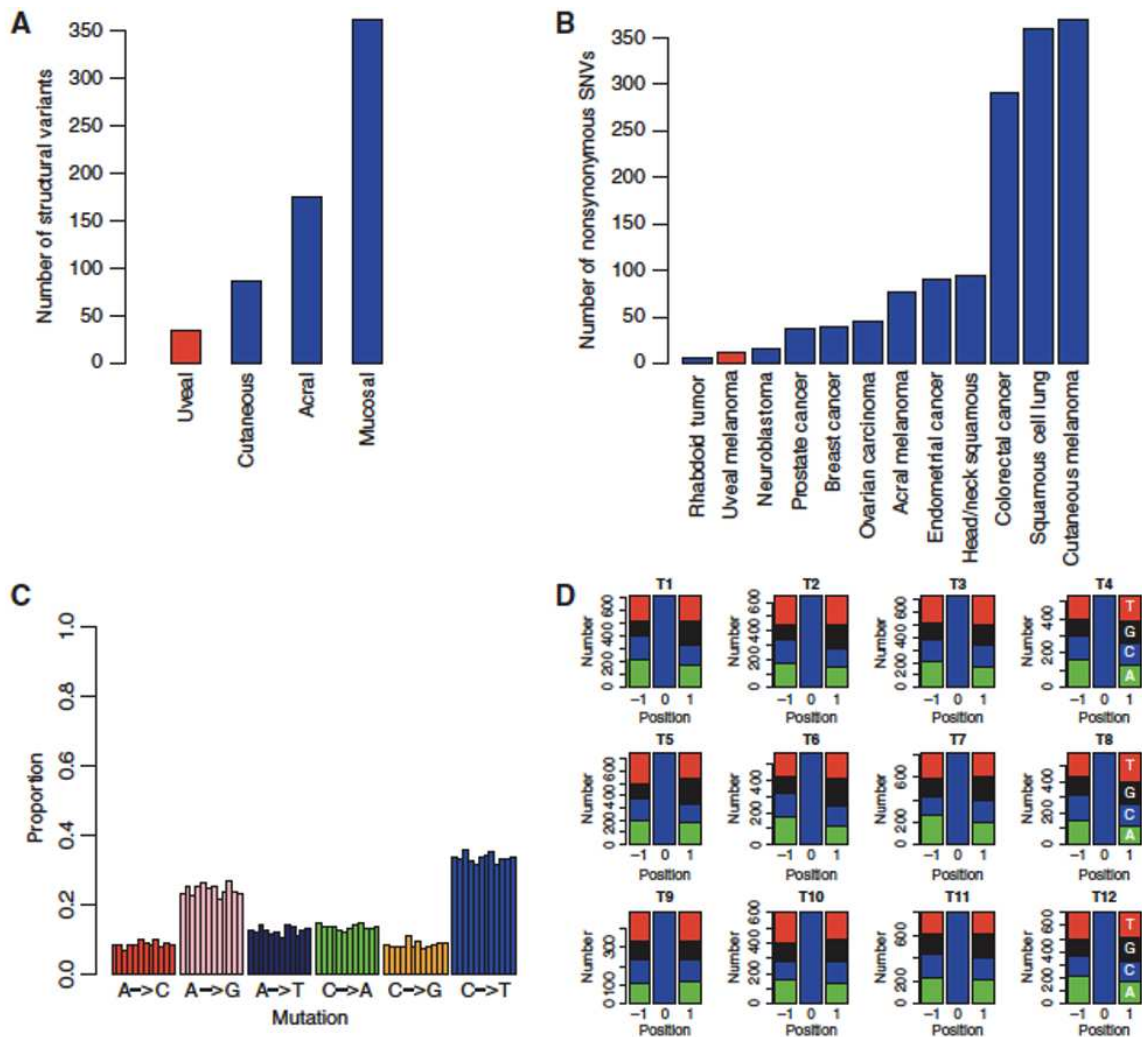


Figure 9. Somatic mutational profile in uveal melanoma.

A. Comparison of the number of somatic structural variants between different melanoma subtypes. **B.** Number of single nucleotide variants (SNVs) in uveal melanoma compared to other cancer types. **C.** Proportion of the different types of nucleotide transitions in uveal melanoma mutational signature (each bar representing one UM tumor sample). **D.** Mutational context of observed C>T transitions in 12 UM tumors with 3' and 5' directly flanking bases revealing an absence of UVR-induced mutational signature. Figure from Furney *et al.*²⁵⁹

SOMATIC MUTATIONS

The main somatic mutational alterations in UM combine two genetic events: mutations of genes within the $G\alpha_q$ signaling pathway (*GNAQ*, *GNA11*, *PLCB4* or *CYSLTR2*), and a 'BSE' event, standing for mutations in *BAP1*, *SF3B1* or other splicing genes, or *EIF1AX*. As discussed later, there is still debate regarding the relative and absolute timing of these two oncogenic events in tumor evolution, although it is generally accepted that mutations of the $G\alpha_q$ signaling pathway are thought to arise first.

MUTATIONS OF THE $G\alpha_q$ PATHWAY

Gain-of-function mutations of *GNAQ*, *GNA11*, *PLCB4* or *CYSLTR2* genes occur in a mutually exclusive manner and lead to the constitutive activation of the $G\alpha_q$ signaling pathway. Mutations of either of these driver genes, as a “first set” of driver events, are present in virtually all UM cases, are thought to arise very early in UM due to presence of the mutation at all stages of malignant progression, and are thought to initiate tumorigenesis, although “ $G\alpha_q$ mutations” are insufficient alone to drive malignant transformation.^{266,267}

GNAQ AND GNA11

GNAQ and *GNA11* are paralogous oncogenes and mutations in either of these genes are present in 85% to 90% of all UM cases, making them the most frequent alteration of the $G\alpha_q$ signaling pathway in UM.^{268,269} The two genes are ubiquitously expressed and share 90% amino acid homology.²⁷⁰ Activating mutations of these genes essentially occur at either of two hotspots, most frequently at codon Q209 in exon 4 (95% of cases), but also at R183 position in exon 5 (5% of cases), and in very rare occasions, G48 mutational hotspot.^{260,268,271,272} These mutations are also found in nevi of Ota where they favor progression into melanoma,²⁷³ and *GNAQ* R183Q somatic mosaic mutation can also lead to port-wine stains or Sturge-Weber syndromes depending on the developmental time of mutation acquisition.²⁷⁴ Importantly, almost all choroidal nevi carry a *GNAQ* or *GNA11* clonal mutation,²⁷⁵ which reinforces the hypothesis that UM can develop from the malignant transformation of a pre-existing nevus and that *GNAQ/GNA11* mutations are the initiating driver event in UM.

GNAQ and *GNA11* encode the $G\alpha$ subunits of the $G\alpha_q$ heterotrimeric proteins (composed of three subunits, α , β and γ), which are essential membrane-bound signal transducers with intrinsic GTPase activity that act directly downstream of G-protein coupled receptor (GPCRs). Without the presence of a ligand, GPCRs are in their basal state and are bound to $G\alpha$, itself also bound to guanosine diphosphate (GDP) and to its partners $G\beta$ and $G\gamma$, forming an inactive G protein trimer. Upon activation by binding of an extracellular ligand, GPCRs undergo a conformational change and act as guanine nucleotide exchange factors, leading to the release of $G\alpha$, which switches from $G\alpha$ –GDP state to its activated form $G\alpha$ –GTP (guanosine triphosphate), thereby promoting the subsequent dissociation of activated

G α -GTP from G $\beta\gamma$, both of which can then activate their respective downstream signaling cascades.

G_q proteins are one of the four main families of G proteins, along with G_i, G_s and G_{12/13}, which regulate distinct downstream biological pathways. G_q family consist of four members, with four respective α subunits G α_q , G α_{11} , G α_{14} and G $\alpha_{15/16}$ encoded by the *GNAQ*, *GNA11*, *GNA14* and *GNA15* genes, respectively. Upon upstream GPCR activation, all G_q proteins activate phospholipases C β (PLC β), their canonical second messenger, which in turn cleaves phosphatidylinositol 4,5-biphosphate (PIP₂) into diacyl glycerol (DAG) and inositol 1,4,5-triphosphate (IP₃). IP₃ release results in an increase in cytosolic calcium concentration, which acts together with DAG to activate protein kinase C (PKC), which leads to further activation of downstream molecules via their phosphorylation. As mentioned later, PKC activation can trigger various signaling pathways,²⁷⁶ such as the MAPK/ERK pathway, and lead to cell growth and oncogenic behavior upon excessive activation.

Mutations involving Q209 and R183 hotspots within *GNA11*/*GNAQ* lie in the GTPase domain of G α_q proteins and result in decreased GTPase activity, locking G α_q proteins in their GTP-bound activated state, thus leading to excessive activation of downstream signaling pathways, albeit to different extents.²⁷⁰ Although differences in the biological effects of Q209 and R183 are still unclear, Q209 amino acid change from glutamine to proline, leucine or arginine are all thought to completely abrogate the GTPase function of G α_q as Q209 is located in the switch II region required for the hydrolysis of GTP to GDP and for the binding of regulator of G protein signaling (RGS) proteins, which catalyze this process.²⁷⁰ Mutations at codon R183 however, may only result in decreased GTPase activity as this amino acid position is not required for RGS binding. Although *GNAQ* and *GNA11* are paralog genes, mutations of the former are thought to arise in 45% of primary UM (PUM) and 22% of metastases (MUM), while mutations of the latter are present in 32% of PUM and 57% of MUM, suggesting potential differences in the effects of these two genes;^{268,269} yet, a subsequent large study pooling reports of somatic mutations in PUM suggested a more balanced contribution of *GNAQ* and *GNA11*²⁷¹, and future studies will be needed to further shed light on the potentially differing mechanisms of these two genes in UM.

PLCB4

15 to 20% of UM cases do not harbor a *GNAQ* or *GNA11* mutation. Deep sequencing of UM samples revealed a recurrent mutation in the *PLCB4* gene (D630 hotspot) in a subset of samples harboring wild-type *GNAQ* and *GNA11*,²⁵⁸ suggesting that mutations of these genes are mutually exclusive in UM. Consistent with this finding, *PLCB4* encodes the phospholipase C β 4 protein, isoform 4 of the PLC β enzyme acting directly downstream of G α_q in the GPCR–G α_q induced signaling cascade, through direct interaction with G α_q ,^{258,277} resulting in DAG and IP₃ formation and activation of downstream cascades through PKC signaling. *PLCB4* hotspot mutations are rare in UM (2.5–4%)^{260,278} and are thought to lead to constitutive activation of the same signaling pathway than *GNA11* and *GNAQ* via gain-of-function,²⁵⁸ although this has yet to be demonstrated.

CYSLTR2

Shortly after the report of *PLCB4* mutations in UM, mutational analysis of WGS or WES data of 136 UM samples previously published in multiple cohorts identified a recurrent mutation in *CYSLTR2* in almost half (4/9) patients not harboring any *GNAQ*, *GNA11* or *PLCB4* mutations and none in patients mutated for either of these 3 other genes, once again indicating mutual exclusivity and suggesting implication in a common biological pathway.²⁷⁸ Hotspot mutations at L129 on *CYSLTR2* are present in ~4% of UMs.²⁶⁰

CYSLTR2 encodes cysteinyl leukotriene receptor 2, a GPCR activated by leukotriene binding and known to activate G α_q .²⁷⁹ L129 lies within a highly conserved region known to stabilize the resting (inactive) state of leukotriene receptors, and alteration of this position results in the constitutive activation of this family of GPCRs, once again aberrantly activating the GPCR–G α_q signaling cascade through PKC signal transduction pathways.²⁷⁸

CONSEQUENCES OF ALTERATIONS OF THE $G\alpha_q$ SIGNALING PATHWAY

GNAQ, *GNA11*, *PLCB4* and *CYSLTR2* code for proteins that are involved in the same signal transduction pathway and therefore activating mutations in these genes are thought to affect the same biological pathways downstream $G\alpha_q$ pathway constitutive activation in UM tumorigenesis. Knowledge of the pathways that are dysregulated in UM downstream of the

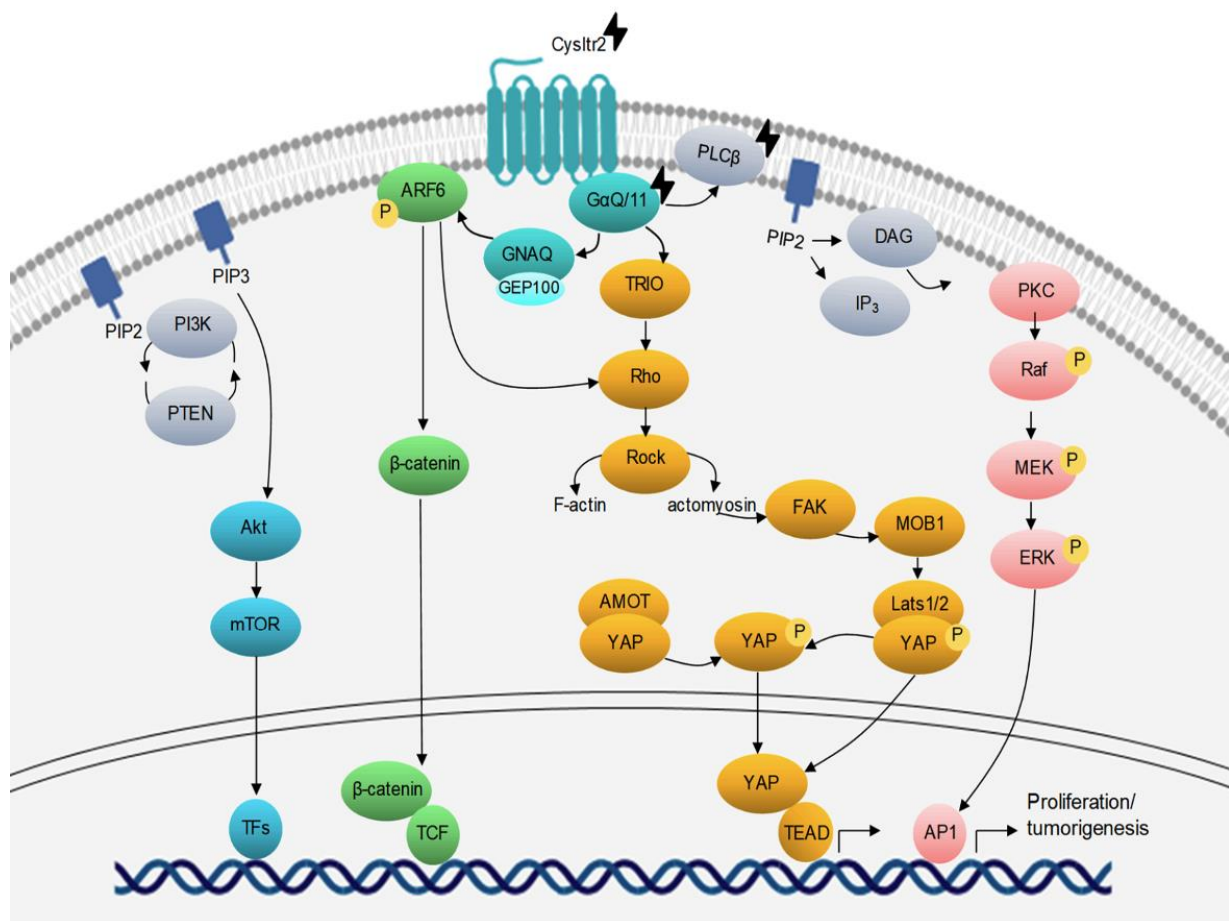


Figure 10. Signaling pathways downstream of $G\alpha_q$ pathway alterations in uveal melanoma.

Recurrent mutations of *GNAQ*, *GNA11*, *PLCB4* or *CYSLTR2* in UM result in alterations depicted in black. These further lead to the dysregulation of multiple downstream signaling pathways, including activation of the Ras-Raf-MAPK signaling cascade, resulting from activation of phospholipase C β (PLC β), subsequent cleavage of phosphatidylinositol (PIP2) into diacylglycerol (DAG) and inositol triphosphate (IP $_3$), allowing the release of cytoplasmic calcium which with DAG activates protein kinase C (PKC), initiator of the MAPK cascade. A PLC β -independent pathway is also overactivated in UM, involving the downstream activation of Yes-associated protein (YAP) following the Rho small GTPase activation through TRIO directly downstream $G\alpha_q$. Other potentially targetable pathways in UM include the β -catenin pathway upon small GTPase ARF6 activation by $G\alpha_q$ and the phosphatidylinositol-3 kinase (PI3K), Akt, mTOR signaling cascade. The end results of these pathways are most often nuclear translocation of factors or co-factors leading to activation of target genes implicated in cell growth and proliferation. Figure from Vivet-Noguer *et al.*²⁸⁰

G α_q pathway mutations are of particular importance as they may reveal targetable molecules to test as drug candidates in therapeutic strategies.²⁸⁰

Figure 10 above outlines the signaling pathways downstream of G α_q pathway activation that may play a role in UM, although their respective contributions in UM oncogenesis are still unclear.²²³ The first pathway described as activated in UM is the MAPK pathway, where some studies showed increased levels of phosphorylated MEK and ERK due to previously mentioned activation of PKC.^{281,282} Unlike in CM, deregulation of this pathway occurs without *BRAF* mutations in UM.^{281,283} Yet, other studies did not find MAPK pathway to significantly contribute to UM,²⁸⁴ and MEK1/2 inhibitors (acting downstream of PKC) showed no clinical benefits.²⁵² Another study demonstrated a role for GTPase ARF6 in G α_q signaling transduction, leading to activation of multiple pathways including β -catenin, MAPK and Yes-associated protein (YAP) pathways, where ARF6 inhibition by small molecules could result in reduced proliferation and tumorigenesis in a UM mouse model.²⁸⁵ Several studies support an important role for activation of oncogenic driver YAP in UM,^{286,287} where its dephosphorylation allows nuclear trafficking and binding to transcription factors such as those from the TEAD family. In addition to its Hippo-dependent activation (MST/LATS/YAP cascade, which when deregulated results in inactive LATS1/2 tumor suppressor kinase and activated YAP/TAZ, associated with oncogenic processes in multiple cancers when overexpressed),^{288,289} YAP activation has also been shown to occur via non-canonical signaling cascade initiated by TRIO (guanine nucleotide exchange factor) activation of small GTPase Rho.²⁸⁷ Recently, focal adhesion kinase (FAK) has been identified in a synthetic lethal gene interaction screen, as part of this non-canonical TRIO signaling, resulting in the inhibition of Hippo signaling (which normally inhibits excessive YAP activation) and in the phosphorylation of MOB1, leading to YAP activation.²⁹⁰ Although YAP knockdown results in decreased cell growth in *GNAQ/11* mutated UM cell lines,²⁸⁶ no drug targeting YAP has been used in clinic for UM treatment so far. However, co-targeting of FAK (upstream of YAP) and MEK-ERK in a recent study demonstrated synergistic growth-inhibitory effects in UM cells and resulted in tumor collapse in UM xenografts and in *in vivo* metastatic UM models,²⁹¹ providing a new potential strategy for targeted therapy.

A last potentially targetable pathway in UM is the phosphatidylinositol-3 kinase (PI3K) pathway, which initiates a signaling cascade involving Akt and mTOR (mammalian target of rapamycin).^{292,293} Although PI3K inhibition resulted in weak effects on proliferation in *GNAQ/11* mutated UM cell lines and in patients,^{294,295} the finding that mTOR pathway could be upregulated upon MEK inhibition²⁹⁴ led to studies combining inhibition of MEK with PI3K/mTOR leading to significantly improved inhibition of UM cell growth and tumor growth in vivo.^{280,292}

While the mechanisms of action resulting from the oncogenic activation of the $G\alpha_q$ signaling pathway have yet to be refined, this first driver event characterized by mutually exclusive mutations of *GNAQ*, *GNA11*, *PLCB4* and *CYSLTR2* are observed in almost all UM patients and are likely the initiating driver event in UM tumorigenesis,²⁷¹ consistent with the fact that these mutations do not lead to differential prognosis or clinical outcome,^{260,278,296} suggesting a role for these mutations in tumor initiation rather than metastatic progression.

BSE EVENTS (*BAP1*, *SF3B1* AND *EIF1AX*)

The low rate of malignant transformation of choroidal nevi into melanoma compared UM occurrence suggests a role for another key mutational event in UM malignant transformation. This second genetic aberration is often referred to as the 'BSE' event, characterized by almost mutually exclusive mutations of *BAP1*²⁹⁷⁻²⁹⁹ *SF3B1* or other splicing genes^{259,260,300,301} and *EIF1AX*,^{260,261,300} which are strongly associated with prognosis.

BAP1

BAP1 (*BRCA-1 associated protein 1*) is located on human chromosome 3p21.3. Bi-allelic inactivation of *BAP1* is seen in almost 50% of primary UM, and in a majority (85%) of metastatic UM.²⁹⁷ *BAP1* acts as a TSG in UM, where loss-of-function mutations can occur throughout the length of the gene exons,²⁷¹ and for which bi-allelic inactivation occurs in a two-hit Knudson model by combining a deleterious mutation in one allele and loss-of-heterozygosity (LOH) of the second allele by monosomy 3 (or less frequently isodisomy 3), one of the most recurrent cytogenetic alterations in UM as discussed later, associated with poor prognosis.³⁰² It is now well-known that UM tumors with inactivated *BAP1* have a greatly increased risk of tumor progression, metastatic relapse and are associated with the poorest prognosis.^{260,297} Somatic inactivation of *BAP1* is implicated in multiple other cancers, including

mesothelioma, clear renal cell carcinoma (RCC) and cutaneous melanocytic tumors.³⁰³ Importantly, germline mutations of *BAP1* are also seen in 2–4% of all UM cases³⁰⁴⁻³⁰⁶ (detailed in “Genetic susceptibility” section) where they confer strong predisposition to UM and in this context represent the first of two hits required for complete inactivation. Germline *BAP1* mutations are also frequent in other cancers and define a tumor predisposition syndrome, *BAP1*-TPDS, which includes UM,³⁰⁷ mesothelioma, RCC, melanocytic tumors, basal cell carcinoma and possibly multiple other tumor types, often associated with a group of highly aggressive tumors conferring poor prognosis.³⁰⁸⁻³¹³

BAP1 is a multi-function protein that has been extensively studied in the past decade, elucidating key biological functions but also key mechanisms in tumor suppressor activity,^{314,315} although clear oncogenic mechanism of *BAP1* in UM is still unclear. *BAP1* is a nuclear ubiquitin carboxy-terminal hydrolase (UCH) belonging to the family of deubiquitinating enzymes, which cleave ubiquitin from proteins thereby regulating multiple processes such as proteasomal degradation and cellular localization. *BAP1* is a 90kDa protein comprising 17 exons and multiple functional domains, including its N-terminal catalytic UCH domain, an HCF1 binding domain (HBM) and a C-terminal binding domain (CTD) with a partly overlapping nuclear localization domain (NLS).³¹⁵ *BAP1* sequence also harbors numerous binding regions for interaction with ASXL1/2/3 (Additional Sex Combs Like), *BARD1*, *BRCA1* and others. An initial role for *BAP1* was proposed in double-stranded DNA break repair, owing to its expected binding to *BRCA1* protein via its RING finger domain (Figure 11).³¹⁶ However, recent mass spectrometry studies did not identify an association between *BAP1* and *BRCA1* nor *BARD1*, making the postulated role of *BAP1* in DNA repair unclear with conflicting results,³¹⁷⁻³¹⁹ and may not be the most UM-relevant pathway as no particular genomic instability is seen in *BAP1*-mutated UM context. Multiple other biological functions have now been elucidated (Figure 11),³¹⁴ most of which involve *BAP1*’s deubiquitination activity, including nuclear functions in chromatin modification, regulation of transcription factors, gene transcription and DNA damage response/genome stability, and cytoplasmic roles in the regulation of apoptosis, ferroptosis and oxidative phosphorylation. For an extensive review of *BAP1* biological functions and implications in cancer, please refer to Carbone and colleagues³¹⁴ and Louie *et al.*³¹⁵

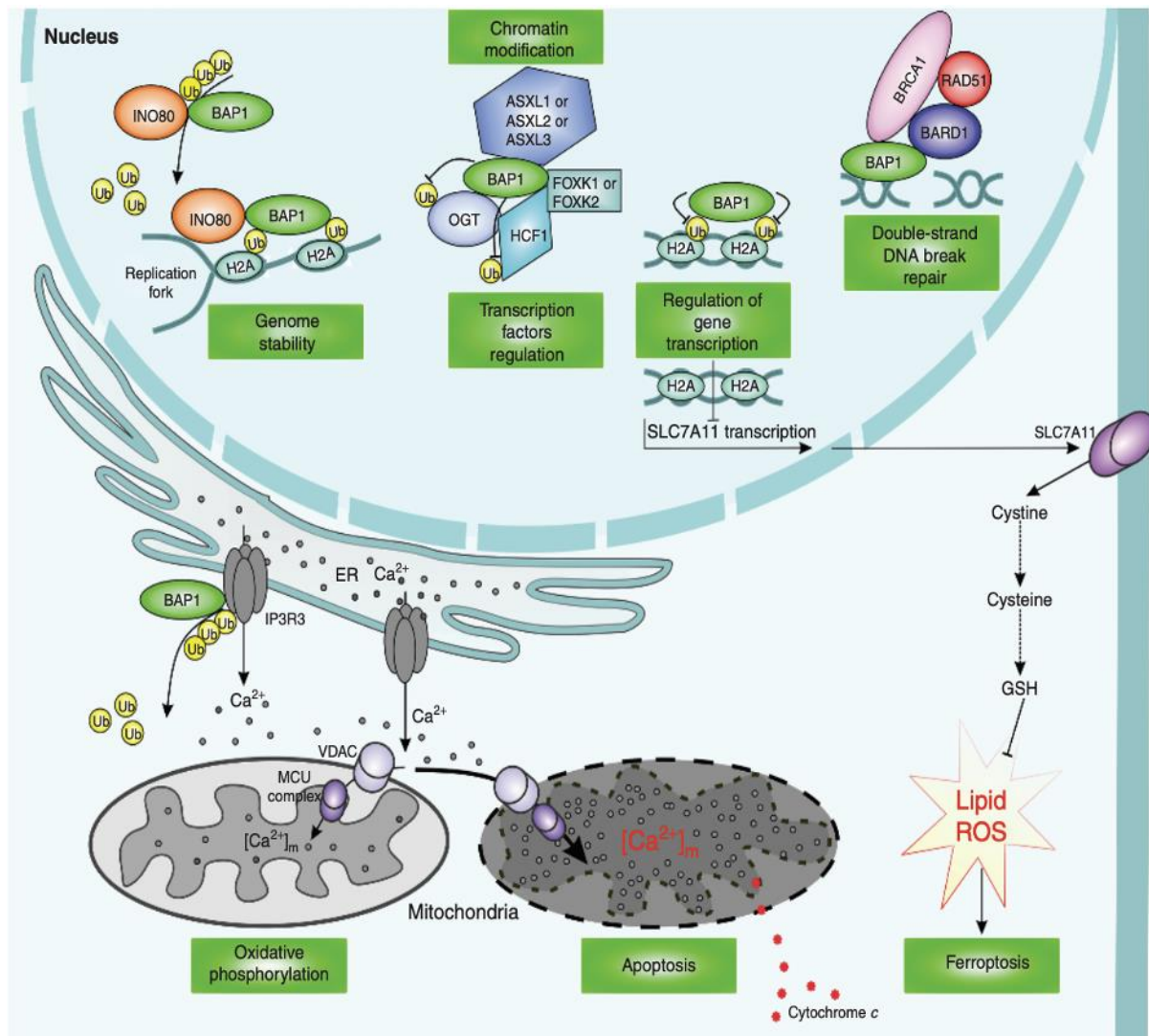


Figure 11. Nuclear and cytoplasmic functional roles of BAP1.

BAP1's role in genome stability occurs via recruitment of INO80 to replication fork via H2A-Ub interaction, ensuring efficient DNA replication.³²⁰ Epigenetic and genetic regulation mediated by BAP1 occurs via interaction with ASXL1/2/3 in combination with FOXK1-HCF1,³²¹ leading to deubiquitination and stabilization of HCF1 and OGT, localization of this complex to multiple gene-regulatory elements and activation or repression of target genes.³²² BAP1/ASXL heterodimerization also plays a major role in gene regulation via epigenetic modifications such as deubiquitination of histone factors.^{323,324} Functions of BAP1 as a tumor suppressor also occur via regulation of H2A-Ub levels, as deubiquitinated H2A can bind to the *SLC7A11* promoter, and expression of this gene induces lipid reactive oxygen species (ROS) formation and ferroptosis via extra-cellular import of cystine.³²⁵ BAP1 may also participate in Rad51-dependent double-stranded DNA break repair, which involves many proteins that may be regulated by BAP1 via deubiquitylation.³²⁶ Finally, BAP1 is a critical regulator of the IP3R3 endoplasmic reticulum (ER) channels in the cytoplasm, via regulation of its ubiquitylated levels.³²⁷ Proper amounts of IP3R3 are required for calcium transfer to the mitochondria required for oxidative phosphorylation, while excessive mitochondrial calcium levels may result in apoptosis, preventing unrepaired DNA damage to accumulate and potentially cancer development.^{328,329} Figure from Carbone *et al.*³¹⁴

Among the many functions of BAP1, one of its roles with perhaps most relevance to UM is its fundamental implication in epigenetic remodeling. Human BAP1, as its *Drosophila* ortholog *Calypso*, bind to ASXL1/2/3 proteins (ASX in *Drosophila*) to form the Polycomb repressive deubiquitinase complex PR-DUB,^{330,331} which can modulate chromatin via BAP1-catalyzed deubiquitination of monoubiquitinated histone H2A on lysine 119 residue (H2AK119Ub1), thereby inducing gene expression.³³⁰ PR-DUB function is directly antagonist to that of the Polycomb repressive complex 1 (PRC1), which leads to gene silencing via chromatin modification and more precisely mono-ubiquitination of H2AK119Ub1. These family of transcriptional regulators are crucial during development, where they are known to regulate transcription of gene involved in differentiation, self-renewal, embryonic development and pluripotency.³³² Supporting a role for BAP1 in chromatin remodeling and epigenetic regulation of gene transcription and silencing balance in UM, Field and colleagues³³³ recently demonstrated that BAP1 inactivation in UM tumors with high metastatic risk leads to significant changes in methylation, where hypermethylation at chromosome 3 results in the downregulation of multiple genes in different pathways. In this respect, loss of BAP1 in UM could participate in cancer progression by deregulating the transcription of critical genes via epigenetic modifications. Lastly, deregulation of the Hippo pathway in BAP1-deficient tumors has recently been described in pancreatic tumors,³³⁴ characterized by enhanced ubiquitin-dependent degradation of LATS tumor suppressors. As deregulation of the Hippo pathway may potentially be implicated in UM, future studies assessing a potential role for BAP1 in this context may be interesting.

SF3B1 AND OTHER SPLICING FACTORS

Another BSE event that occurs in 25% of primary UMs is characterized by hotspot mutations of a splicing gene, of which *SF3B1* is the most frequently mutated (20% of all UM cases).^{259,300,301} These missense mutations mainly occur at hotspots R625, H666 and to a lesser extent K700, the latter being the preferred mutated site in myelodysplastic syndrome and chronic lymphocytic leukemia.³³⁵⁻³³⁷ In UM, *SF3B1* mutations are almost mutually exclusive with *EIF1AX* and with *BAP1* mutations and monosomy 3, and are associated with low to moderate metastatic risk compared to *BAP1*,^{260,301} although studies with longer patient follow-up demonstrated a higher rate of late-onset metastases (>5-10 years after UM

diagnosis) rather than prolonged survival.^{301,338} *SF3B1* encodes splicing factor 3B subunit 1, a critical component of the U2 small nuclear ribonucleoprotein (snRNP) complex of the spliceosome, and is involved in early stages of RNA splicing where it allows branch point (BP) recognition (upstream of the 3' splice site) and recruitment of U2 snRNP by interacting with intronic RNA and U2AF, mandatory for correct splicing.^{339,340} *SF3B1* change-of-function mutations promote recognition of an alternative BP, disrupting normal splicing at ~1% of splicing junctions and resulting in usage of cryptic 3' splice sites which contributes in tumors to transcriptomic and proteomic diversity.^{259,340-343}

Recent studies of the biological mechanism underlying *SF3B1* mutants in multiple cancers demonstrated loss of interaction of *SF3B1* with SUGP1, another spliceosomal protein, which upon knockdown recapitulated K700 *SF3B1* mutant splicing defects.^{344,345} Decreased interaction between SUGP1 and *SF3B1* within the spliceosome, required for correct BP recognition, was also seen in *SF3B1* mutants other than at K700 hotspot.³⁴⁴ The mechanisms by which *SF3B1*-related splicing defects promote UM malignancy have yet to be elucidated. Possible consequences of cryptic splice site usage are mRNA degradation if the intronic insertion is out-of-frame, but since the latter are in-frame in 1/3rd of abnormal transcripts, change-of-function or activated proteins may also play a role.²²³ It was shown that mutant *SF3B1* results in recognition of an aberrant BP within *BRD9*, which encodes a core component of the non-canonical BAF remodeling complex, resulting in its depletion and alteration of BAF chromatin localization, leading to tumorigenesis via altered regulation of HTRA1 and genes involved in apoptosis and cell growth.³⁴⁶⁻³⁴⁸ This study also showed that *BRD9* demonstrates potent tumor suppressor activity in UM. Investigations of target genes in other malignancies affected by *SF3B1* mutations have also identified *TERC*, *MAPK37*, *KLF8*, *ABCB7* and multiple others in myelodysplasia that may represent potential candidates in UM as well.³⁴⁸ Another candidate target gene is the proto-oncogene *MYC*, thought to be activated via post-translational modifications resulting from *SF3B1*-mutant mediated decay of transcripts encoding protein phosphatase PP2A Ser/Thr complex, which normally regulates *MYC* and *BCL2* dephosphorylation, resulting in their aberrant stabilization.³⁴⁹ Finally, a recent study demonstrated that *SF3B1*-mutant UMs share common tumor-specific neo-antigens,³⁵⁰ representing a potential source of therapeutic strategies for cancers involving *SF3B1*-related splicing defects.

In addition to *SF3B1*, analysis of UM cases within the TCGA cohort reported in-frame deletions of *SRSF2* (*serine and arginine rich splicing factor 2*),²⁶⁰ which was confirmed by others.^{266,343} Similar to *SF3B1*, mutations of this gene are also found in hematological malignancies.^{343,351} *SRSF2* is another major component of the splicing machinery, and mutations are also thought to deregulate mRNA splicing³⁵² although the target genes and proteins in UM have yet to be identified. Other rare alterations of splicing-related genes in UM include loss-of-function mutations of *RBM10* and VUS in *SF3A1* and *SRSF7*.^{266,343}

EIF1AX

Somatic mutations of *EIF1AX*, which encodes the eukaryotic translation initiation factor eIF1A, were initially reported in UM by Martin and colleagues.³⁰⁰ They essentially occur in the absence of *SF3B1* or *BAP1* mutations, mostly in tumors with disomy 3 and exceptionally in monosomy 3 UM. UM cases harboring somatic *EIF1AX* are associated with good outcome and very low metastatic risk.^{260,353} Missense mutations and small deletions of *EIF1AX* arise in ~15–20% of UMs, within the unstructured N-terminal tail (NTT) of eIF1A.^{271,300} As a major component of the 43S preinitiation complex, eIF1A allows the transfer of methionyl initiator tRNA to the small 40S ribosomal subunit, required for translation initiation. The functional consequences of *EIF1AX* mutations in UM are unknown, and usage of alternative start codons to overcome weak recruitment of ternary complex at the first start site resulting from *EIF1AX* mutants have been suggested,^{300,354} which could result in differential ratios of protein isoforms. The downstream protein targets have yet to be determined in UM,³⁵⁵ as well as the processes by which *EIF1AX* alterations result in tumorigenesis without (or negatively) influencing metastatic risk. Of note, *EIF1AX* mutations are also found in a subset of thyroid carcinomas, in a mutually exclusive manner to other oncogenic driver events in this tumor type.^{356,357} Recently, *EIF1AX* and *RAS* mutations (that strongly associate in thyroid carcinoma) were shown cooperatively drive thyroid tumorigenesis through stabilization of ATF4, a cellular stress sensor, and of c-MYC.³⁵⁸ An integrative analysis of aggressive thyroid cancers also revealed co-mutations of *EIF1AX* and oncogenes *AKT1/PIK3CA* (often associated with mutually exclusive mutations of *BRAF* or *NRAS*), correlating with aggressiveness.³⁵⁹ These mechanisms may shed light on UM oncogenic process driven by *EIF1AX* mutations, although *NRAS* and *BRAF* mutations do not occur in UM.

CHROMOSOMAL IMBALANCES

Cytogenetic abnormalities play a major role in the genomic landscape of UM. UM tumors harbor recurrent copy number variations (CNV) of either partial or whole chromosomes, some of which, combined with the previously stated genetic alterations, correlate with patient outcome and distinguish between tumors with good and poor prognosis.^{261,271,302,360-365} Fluorescence *in situ* hybridization (FISH), microsatellite analysis, SNP microarrays, comparative genomic hybridization (CGH) arrays, multiplex ligand-dependent PCR amplification (MLPA) and droplet digital PCR (ddPCR) are among the technologies that were widely used to detect cytogenetic changes in UM.²⁷¹ These include monosomy 3 (total or partial loss) and 8q amplification, which are the most frequent CNV events in UM (~50% of UM cases), but also frequent 6p gains, followed by 1p loss, 6q loss, 6q gain and 8p loss. Recently, WGS has allowed to establish a comprehensive landscape of genomic events in UM, and this high-resolution technology will surely start to replace the above-mentioned techniques to detect CNVs.²⁶¹ Monosomy 3 and 8q gain are the most important cytogenic events in UM as they are both strongly associated with poor prognosis and high metastatic risk.^{260,261,366}

MONOSOMY 3

It is now well documented that monosomy 3 is heavily correlated with poor prognosis and is the single, strongest factor to predict high metastatic risk.^{302,367-369} This poor prognosis mediated by monosomy 3 is thought to be linked, at least in part, to deleterious mutations of *BAP1* located on chromosome 3p21. While this represents a classical TSG “two-hit” inactivation, TSGs are also often inactivated by mutations of both alleles, which is not the case in UM where *BAP1* inactivation is intriguingly almost always the result of the combination of monosomy 3 and a somatic mutation in the remaining allele, unlike other cancers associated with *BAP1* loss such as mesothelioma or renal clear cell carcinoma where bi-allelic deleterious mutations or loss of chromosome 3 short arm are often seen.³¹¹ This raises the hypothesis that other TSGs may lie on chromosome 3 and play a role in tumorigenesis.³⁷⁰ A recent study demonstrated that two “smallest overlapping regions” were frequently deleted within chromosome 3 in the (infrequent) occurrence of partial deletions of chromosome 3 (rather

than monosomy 3) in UM,³⁷¹ one containing *BAP1* (along with 290 other genes) on chromosome 3p short arm, associated with poor prognosis, and another on 3q long arm, showing no clear impact on metastasis-free progression but potentially still harboring genes involved in UM tumorigenesis, as it is the case for *MBD4* (3q21.1) which is discussed in the following chapter. The occurrence of isodisomy 3³⁷² during clonal evolution and intra-tumor heterogeneity of chromosome 3 status^{369,373,374} are also observed, the latter of which does not seem to affect prognosis but may rather define a distinct subgroup demonstrating extensive immune infiltration.³⁷⁵

8Q GAIN

8q gain is also a common event in UM (almost 50% occurrences) and most often co-occurs with monosomy 3, as another important prognostic factor in UM.³⁷⁶ 8q chromosomal aberration can be the result of direct 8q gain, trisomy 8, or isochromosome 8 characterized by concomitant loss of 8p and gain of 8q. A possible explanation for this common 8q amplification in UM with high-metastatic risk is the presence of oncogene *MYC* on chromosome 8q sub-telomeric region, but this has not been proven yet³⁷⁷ and no correlation was found between 8q gain levels and *MYC* expression levels.²⁶¹ *PTPA3* and *POUF51* were also suggested to play a role in tumorigenesis following the same hypothesis, but further studies are required as data is limited.^{261,377} As we will see in the next section, 8q gains allow to distinguish between two sub-populations of UM associated with monosomy 3 and *BAP1* mutations.

INTEGRATIVE GENOMIC LANDSCAPE OF UM

The combination of these genetic and cytogenetic events in UM tumors are major determinants of metastatic risk in UM progression. Indeed, transitioning towards the next section which addresses important prognostic factors in UM, Robertson and colleagues' integrative genomic study of 80 primary UMs in the TCGA Consortium clearly illustrates the almost perfect dichotomy in UM at the DNA, methylomic, transcriptomic, lncRNA and miRNA levels according to chromosome 3 status (and most often, *BAP1* expression) (Figure 12).²⁶⁰ This finding corroborates and complements the pioneer studies of Oken and colleagues,³⁷⁸ who reported two molecular subsets in UM based on gene expression profile mainly

distinguished by changes in gene expression on chromosomes 3 and 8q (detailed in “Gene expression profiling” section), and that of Royer-Bertrand and colleagues²⁶¹ who assessed the genetic landscape of UM based on WGS of UM cases, clearly showing that loss of chromosome 3 (most often associated with *BAP1* somatic mutations) defines UM “classes A and B” (further distinguished based on 6p loss) while “classes C and D” tumors typically harbor *EIF1AX* and *SF3B1* mutations, are associated with fewer chromosomal rearrangements and mostly disomy 3, although class D can be distinguished from class C based on one major aneuploidy event, characterized by gain of distal parts of chromosome 8q. These findings were recapitulated in multiple other studies, such as that of Field and colleagues.²⁶⁶

Taken together, these findings confirm that *GNAQ/GNA11* somatic events are seen in virtually all UMs and are thus critical initiating steps in UM tumorigenesis, but do not distinguish different molecular subsets in UM, while major cytogenetic and ‘BSE’ alterations do. Importantly, this dichotomy, mainly based on chromosome 3 status, is not only relevant to most genomic characteristics of UM tumors, but also strongly determines metastatic potential and patient outcome,^{261,266,297,360,378} as outlined in UM “Prognostic Factors” chapter below.

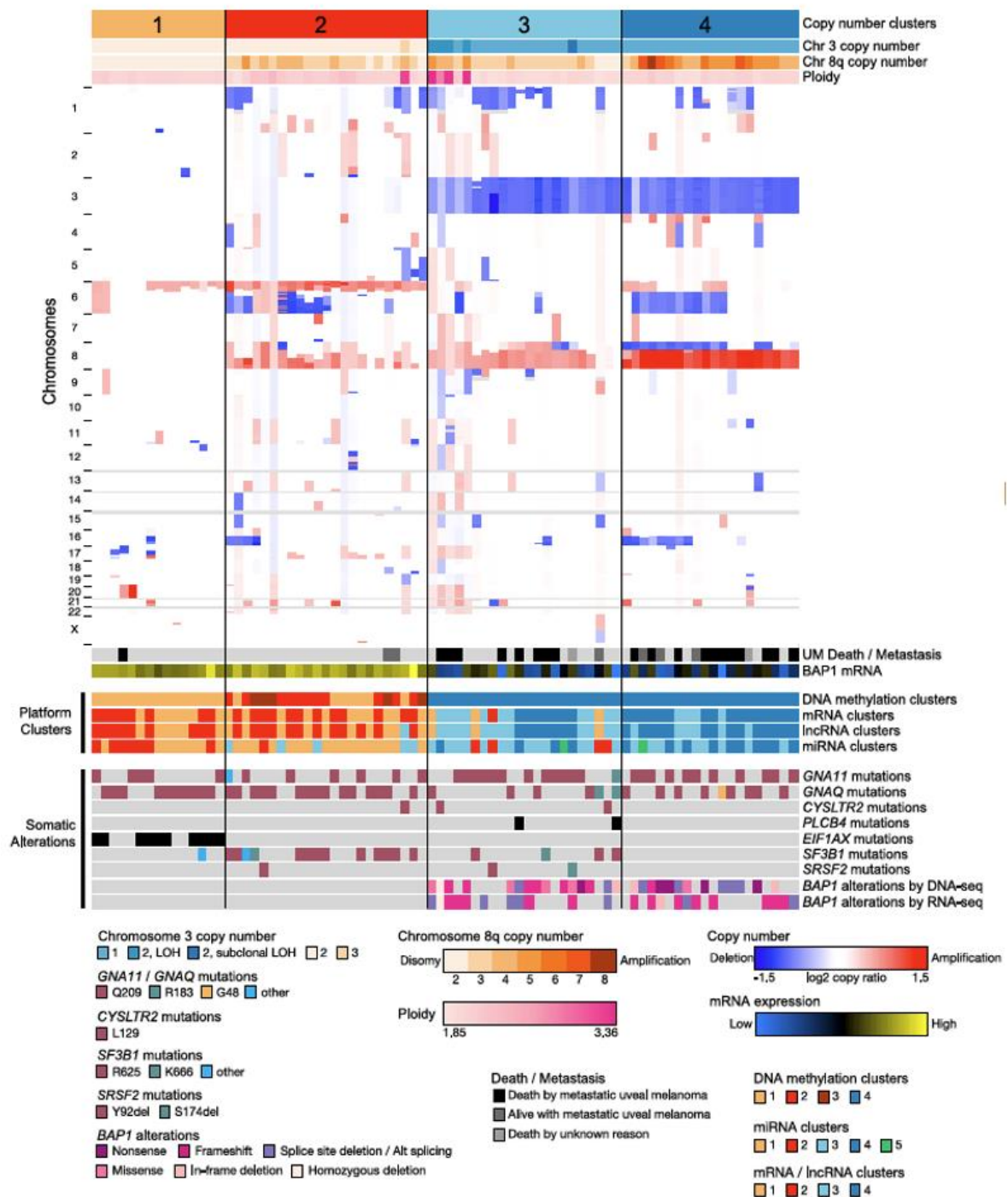


Figure 12. Integrative study of the genomic landscape in UM reveals 4 molecularly and clinically distinct subgroups.

Unsupervised clustering of 80 primary UMs (TCGA dataset) identifies different UM subsets based on chromosomal instability/somatic CNAs (top) (mainly chromosome 3 loss and 8q gain), and (lower track) clinical outcome, mRNA expression levels of *BAP1*, distinct unsupervised clusters on the methylomic, mRNA, long non-coding RNA (lncRNA) and micro-RNA (miRNA) levels. Somatic mutations of UM driver genes are also shown (bottom). These findings indicate that chromosome 3 status, along with *BAP1* somatic inactivation, clearly define a biological dichotomy in UMs at all genomic levels (genetic, methylomic, mRNA, lncRNA and miRNA), which also defines metastatic risk. Figure from Robertson *et al.*²⁶⁰

EPIGENETIC ONCOGENIC MECHANISMS

Another key observation from these genomic integrated analyses in UM, which bring together previous findings from multiple studies, is the importance of epigenetic factors (in addition to oncogenic driver genes) in UM progression and in defining metastatic risk.^{379,380} Some of the oncogenic epigenetic mechanisms thought to contribute to UM tumorigenesis include: (i) the deregulation of certain miRNAs (such as miR-143, miR-193b and Let-7b) that lead to over-expression of oncogenes or inactivation of TSGs, and can differentiate between UMs with high- and low-metastatic risk;^{381,382} (ii) second, aberrant DNA hypermethylation at CpG islands within gene promoters, resulting in silencing of TSGs, have been reported in UM for *RASSF1A* (on chr3p), *RASEF*, *TIMP3*, *p16* and multiple others thought to contribute to UM progression.³⁸³⁻³⁸⁵ More widespread hypermethylation has also been described by Field and colleagues,³³³ as previously stated, specifically on chromosome 3 in UM tumors with *BAP1* inactivation (high metastatic risk group), through which *BAP1* itself was demonstrated to be epigenetically regulated (hypermethylated) and which may account for loss of melanocytic differentiation observed in this UM subgroup; (iii) conversely, DNA hypomethylation, such as at specific CpG sites of the Preferentially Expressed Antigen in Melanoma (*PRAME*) gene, which correlates with metastatic risk in UM associated with 8q gain³⁸⁶ and is often seen as an epigenetic biomarker for metastatic UM; (iv) and finally, histone modifications, where loss of BAP1 deubiquitinating enzyme is thought to result in the gain of stem-like properties via hyper-ubiquitination of histone 2A (H2A),^{387,388} which may be reversed by treatment with histone deacetylase inhibitors (HDACi), at least in xenograft models³⁸⁷ and very recently in a phase 2 clinical trial.³⁸⁹ Taken together, epigenetic changes in UM are essential oncogenic processes that, along with genetic driver events, define distinct molecular subgroups in UM, which also translate into clinical significance as they also help define metastatic risk.

PROGNOSTIC FACTORS

It is estimated that 25% UM patients will develop metastases within 5 years, ~35% within 10 years,³⁹⁰ and almost half by 25 years, indicating that patient follow-up over extended periods of time can be critical. While the primary tumor is most often efficiently treated, the mean

survival is estimated at ~1 year after diagnosis of metastasis.²⁴⁹ Multiple classification systems have been developed to determine the metastatic risk and patient outcome as accurately as possible, according to characteristics of UM tumors such as clinical, histopathological, cytogenetic and transcriptomic aspects. Accurate prognosis is critical for early diagnosis of metastasis, determination of duration of patient follow-up, and selection of patients for clinical trials.

CLINICAL AND HISTOPATHOLOGICAL FEATURES

Location of the UM tumor within the eye is strongly prognostic, as ciliary body location is more associated with metastatic risk than choroidal UM, and considerably more than iridal UM (10-year metastatic rates of 33%, 27% and 5% respectively).^{391,392} Other important clinical risk factors are older age,³⁹³ large basal diameter, tumor thickness,³⁹⁴ and extraocular involvement.³⁹⁵ The tumor node metastasis (TNM) staging system takes into account these factors to classify UM tumors into four categories (T1-T4), allowing to predict metastatic rates.³⁹⁶ Epithelioid cell type (over spindle cell type), high mitotic activity and extravascular connective tissue loops are among the histopathological factors associated with poor prognosis UM.^{391,394,395,397}

CYTOGENETIC CLASSIFICATION

Chromosome 3 status by itself is a strong prognostic factor in UM, where monosomy 3 (detected in half of UM patients) is associated with high risk of metastatic progression, and poor clinical outcome.^{302,398} 8q gain, the most frequent cytogenetic event in UM, is also associated with poor prognosis.^{365,366,376} Multiple studies of the prognostic values of cytogenetic alterations in UM have revealed other important chromosomal events with prognostic value, including 6q loss and 1p loss associated with poor prognosis, and 6p gain with good prognosis (found in an almost mutually exclusive manner with monosomy 3),³⁹⁹ although they are most often dependent or linked to chromosome 3 loss and 8q gain.⁴⁰⁰⁻⁴⁰³ A recent retrospective study of prognosis of 1,059 UM patients based on their cytogenetic profile reported chromosomal imbalances with the strongest risk of metastasis to be complete monosomy 3, partial monosomy 3, 8q gain, 8p loss, followed by 6q loss.⁴⁰⁴ In addition, combination of complete monosomy 3, 8q gain and disomy 6 was associated with the highest metastatic risk (39% risk in 5-year Kaplan-Meier estimates), compared to the

lowest risk on the other end of the spectrum characterized by disomy of chromosomes 3, 6 and 8 (4% risk). While the timing of occurrence of monosomy 3 and 8q gain remains unclear,^{399,405} 8q gain is most often associated with monosomy 3 and results in worsened prognosis.^{365,366,376,394} Damato and colleagues³⁹⁴ determined that the 5-year UM-related mortality rate was 6% with normal chromosomes 3 and 8, 31% with chromosome 8 gain, 40% with monosomy 3 and 66% with combined monosomy 3/ gain of 8 (Figure 13). It should be noted, however, that cytogenetic factors alone cannot accurately predict UM prognosis. While cytogenetic alterations strongly correlate to tumor size and to some extent ciliary body involvement and epithelioid cell type,^{363,394} they should be taken into account with previously mentioned clinical and histopathological factors for accurate determination of prognosis,³⁹⁵ as well as genetic factors mentioned below which further define metastatic risk in UM.

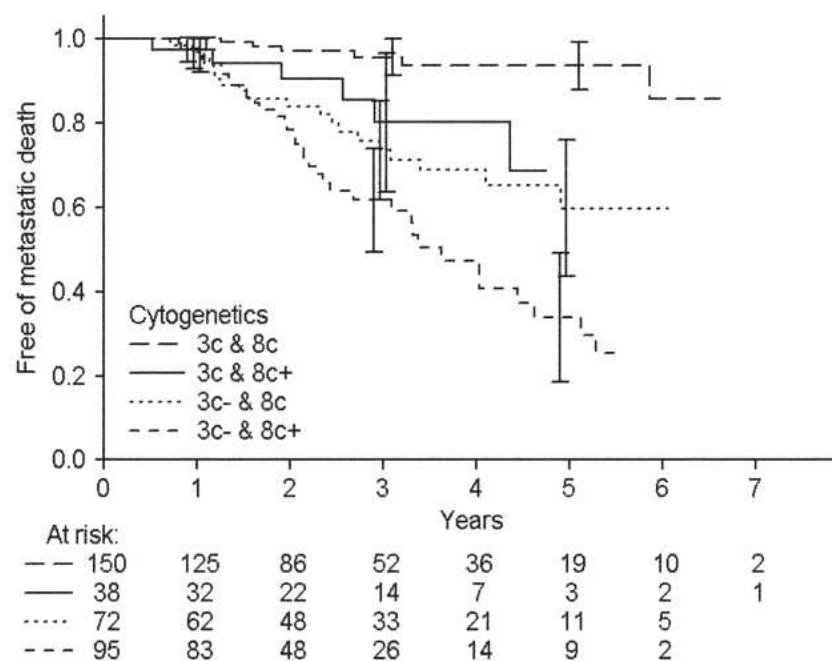


Figure 13. Overall survival Kaplan-Meier curves of UM-related death according to chromosomes 3 and 8 status. Dotted lines (bottom to top) indicate normal chromosomes 3 and 8 copy number (3c & 8c), chromosome 8 gain (3c & 8c+), chromosome 3 loss (3c- & 8c) and combination of chromosome 3 loss and chromosome 8 gain (3c- & 8c+). Figure from Damato *et al.*³⁹⁴

GENE EXPRESSION PROFILING

Gene expression profile (GEP) classification in UM was initially introduced by Harbour and colleagues, revealing two molecular subsets based on the mRNA levels of certain specific genes: class 1, associated with low metastatic potential and thus better prognosis, and class 2 characterized by a very high rate of metastasis.^{378,406,407} 62 discriminating genes were initially reported, with class 2 mainly characterized by downregulated genes on chromosome 3 and upregulated genes on chromosome 8, in accordance with the cytogenetic determinants of high- and low-risk UMs.³⁷⁸ Since then, the list of discriminating genes was subsequently reduced to 12, and 3 “control genes”;⁴⁰⁶ a commercially available test, DecisionDx-UM, is now used in multiple American centers for accurate determination of metastatic risk in UM, outperforming the prognostic value of chromosome 3 status, TNM classifications and other clinical and histopathological factors.⁴⁰⁸⁻⁴¹¹

It should be noted that a significant portion of the discriminating genes lie on chromosome 3 (and to some extent, chromosome 8),⁴¹² suggesting that global expression changes seen in Class 2 GEP tumors may simply reflect copy number variation rather than determining mechanistic pathways implicated in UM metastatic progression. However, (i) the prognostic value of GEP has proven highly accurate so far, as evinced by a metanalysis of results from 12 research centers, allowing to determine GEP classes in 97% of cases, including 62% class 1 and 38% class 2 tumors, after which only 1% of class 1 UM tumors metastasized during the 18 month follow-up period compared to 26% in class 2;⁴⁰⁸ and (ii) GEP classification also provides insights into UM biology. Studies by Onken *et al*⁴¹³ and Chang *et al*⁴¹⁴ demonstrated that while class 1 tumors closely resemble normal, differentiated uveal melanocytes in terms of gene expression, class 2 tumors often undergo stem-like cell behavior, switching back to primitive neural crest-like progenitors, suggesting a role of deregulation of differentiation transcriptomic programs in metastatic progression. In line with this finding, Field and colleagues recently reported class 2-specific methylomic repatterning triggered by *BAP1* loss, downregulating genes notably involved in axon guidance and melanogenesis,³³³ after Harbour and colleagues evidenced the role of *BAP1* deleterious mutations specifically in class 2 tumors.²⁹⁷

Subset	Metastatic potential	mRNA GEP	Chromosome 3	Chromosome 8q	Chromosome 6p	Key mutation	Inflammation
A	Low	Class 1	Two copies	Two copies	Partial or total gain	EIF1AX	No
B	Intermediate	Class 1	Two copies	Partial gain	Gain	SF3B1	No
C	High	Class 2	One copy	Three or more copies	No change	BAP1	No
D	High	Class 2	One copy	Three or more copies; isochromosome 8q	No change	BAP1	Yes

Figure 14. Molecular subsets of uveal melanoma based on genetic, cytogenetic and transcriptomic characteristics.

Figure from Jager *et al.*²²³

Class 1 was further subdivided into classes 1a and 1b, where 1b is associated with intermediate prognosis between classes 1a and 2 (Figure 14).⁴¹⁵ In particular, *PRAME* has been identified as a prognostic biomarker of metastatic risk within class 1 tumors,³⁸⁶ although its mechanism of action is still unclear. Finally, UM tumors with high metastatic risk (class 2) have also been proposed by other groups to be further separated into class 2a and 2b, based on extensive immune infiltration only observed in one group (2b). These two groups were distinguished molecularly solely based on 8q gain in class 2b, which coincided with expression of a set of immune genes located on chromosome 8q.³⁷⁵

Taken together, the combination of different prognostic factors, including mRNA GEP and cytogenetic changes, can help classify UM tumors in four subsets associated with different metastatic potential, which also coincide with main UM “BSE” driver genes (Figure 14).

METASTATIC PROGRESSION AND TUMOR EVOLUTION

The fact that most metastatic UMs (MUM) are treatment refractory while primary UMs (PUM) can most often be efficiently treated suggest that UM tumors acquire new genetic and epigenetic properties during their metastatic progression, highlighting the need for studies analyzing tumor evolution between PUM and MUM.

MALIGNANT PROGRESSION MODELS

There are essentially two progression models that have been proposed to describe UM evolutionary path towards metastasis. The first one is a Darwinian model of gradual evolution of cancer, marked by successive waves of mutations and clonal expansions, ultimately

resulting in malignant tumor, as it is the case for CM progression from benign nevi to melanoma.⁷ In this model, uveal melanocytes would gradually evolve to pre-neoplastic nevi, marked by a first early oncogenic event of activating mutations of the $G\alpha_q$ signaling pathway, thought to initiate tumorigenesis as *GNAQ/GNA11* are found in uveal nevi and benign blue nevi, unlike 'BSE' mutations.^{268,275} The most recent common ancestor (MCRA) of UM would then harbor *GNAQ/GNA11* mutations and start acquiring subclonal somatic mutations at a low rate, until a second (and later) oncogenic event characterized by combinations of BSE mutations and ultimately cytogenetic events, resulting in further tumor progression (such as *BAP1* bi-allelic inactivation by mutation of one allele and loss of chromosome 3, associated with strong metastatic risk and death). However, the timing and order of events required for the progression from benign nevi to UM is still not fully elucidated.^{223,268,269,275}

Recently, another model has been proposed, supporting a punctuated evolution marked by rapid, almost simultaneous bursts of oncogenic events driving tumor fitness.²⁶⁶ This model by Field and colleagues arises from their observation that *BAP1* and other main genomic aberrations in UM (BSE events but also their associated recurrent CNAs such as monosomy 3 and 8q gain), are essentially clonal events present in all tumor cells, followed by clonal stasis and neutral evolution until clinical detection (although some additional CNAs were still found to be subclonal events in some cases, such as 6p gain or 1p loss).²⁶⁶ Owing to major improvements in sequencing techniques and bioinformatic tools, they were also able to report previously undetected *BAP1* (but also *SF3B1* and other splicing genes) alterations (such as large indels and complex rearrangements). In this preferred model, heterogeneity in UM driver events, including copy number alterations, arises early and over a short period of time, implying that metastatic potential is determined early in tumor development, with very little acquisition of other driver events after MCRA appearance and thus little clonal selection after this point. Although the order of appearance and relative timing of the two main driver events in UM remains unknown, it is still thought that $G\alpha_q$ alterations are the initiating event required for subsequent UM tumorigenesis (but insufficient alone to drive tumorigenesis), but may either occur prior to the BSE event, or after but in this case likely acts as a key determining step for BSE mutations to be "unsilenced" and trigger malignant transformation.²⁶⁶

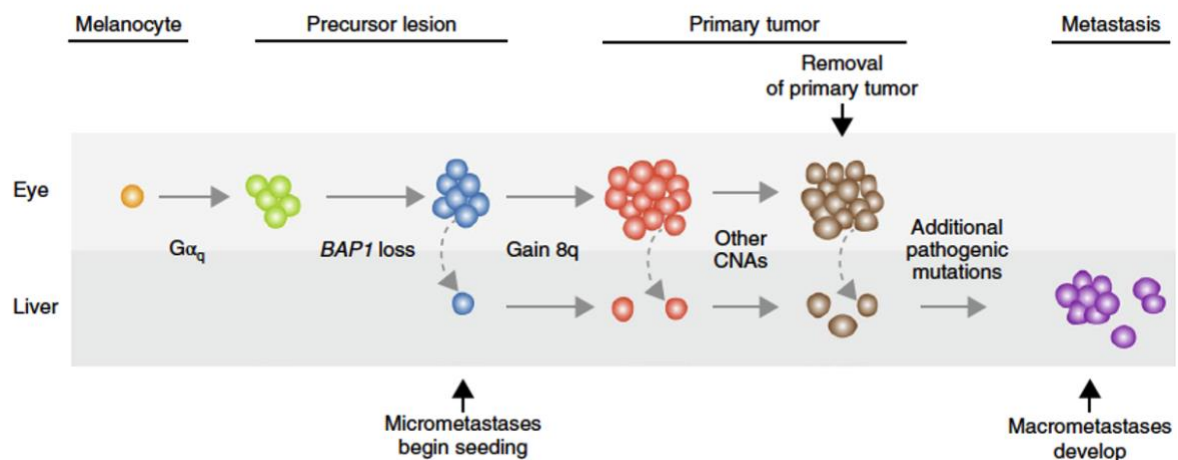


Figure 15. Evolutionary route of metastatic uveal melanoma.

The model for sequential order of events leading to metastatic UM proposed by Shain and colleagues places the initiating $G\alpha_q$ first, leading to precursor lesion, rapidly followed by *BAP1* loss, gain of 8q and other copy number alterations such as 8p loss, 6q loss or 6p gain, leading to the primary tumor. Macrometastases only develop at a much later change accompanied by other (not necessarily driver) pathogenic mutations, such as *CDK2NA* loss and *EZH2* mutations (mentioned below in “tumor evolution” section). Figure from Shain *et al.*⁴¹⁶

Other studies are in accordance with this model,⁴¹⁶⁻⁴¹⁸ such as Shain and colleagues’ study of patient-matched PUMs and MUMs placing $G\alpha_q$ events first, rapidly followed by BSE and recurrent CNA events in ‘early trunks’ of the evolutionary trees (their proposed sequential order of events leading to MUM is show in Figure 15). Another recent study also finds consistent results, placing *BAP1* mutational events very early after tumor initiation (0.5–5 years after), and long before primary tumor diagnosis (9-33 years later), coinciding with the seeding of micrometastases.⁴¹⁸

TUMOR EVOLUTION

GENOMIC EVOLUTION

Only few studies of paired PUMs and MUMs in UM have been carried out so far.^{416,417} Rodrigues and colleagues⁴¹⁷ and other groups^{298,416} show that progression from PUM to MUM is remarkably stable, marked by very few additional SNVs, that some recurrent CNAs are acquired during metastatic progression (such as 1p, 6q, 8p, 16p losses, isodisomy 3, gain of 1q and 8q) compared to PUM, and that no additional major driver oncogenic event appears during this process. As expected, the main (and early) driver event of metastatic UM is *BAP1* inactivation,^{297,298,416} associated with large-scale methylomic repatterning.³³³

In a number of cases, however, infrequent to rare other pathogenic mutations (sometimes referred to as tertiary drivers) have been reported, probably arising at later stages in tumor progression. These include chromatin remodelers *EZH2* and *PRBM1*, but also *CDKN2A* (commonly inactivated in CM), *GNAQ* loss of heterozygosity, and rare occurrences of mutations in genes of the $G\alpha_q$ pathway such as *PTK2B*.^{298,416,417,419} A study integrating chromosomal aberrations with expression, mutational and methylation data in 182 PUM samples further described candidate genes on chromosomes 1p, 8q and 6q in UM progression, and a potential tumor suppressor role for *PHF10* on chr6q27, inactivated in 3 tumor samples and implicated in transcriptional regulation, adhesion and migration in UM cell lines.⁴²⁰ Recently, a single-cell analysis in 8 PUM and 3 MUM samples revealed multiple transcriptional states and genomic complexity at the subclonal level.⁴²¹ This was confirmed by another recent study in which authors reported, in a single-cell transcriptomic profiling analysis, intra-tumor heterogeneity in PUM, with specific sub-populations driving the metastatic process and notably evidenced a role for HES6 in a specific transcriptional state associated with poor prognosis and invasive functionality driving metastatic progression.⁴²²

IMMUNE MICROENVIRONMENT

Lastly, tumor immune microenvironment also plays an important role in tumor progression (and potential response to some treatments), where a minority of class 2 tumors (notably characterized by monosomy 3) have considerable immune infiltration in the primary tumor as in the metastasis, mainly composed of CD8+ T-cells and macrophages (especially of the M2 phenotype), which leads to the formation of blood vessels favoring metastatic dissemination, and is associated with poor prognosis in UM.⁴²³⁻⁴³⁰ Specifically, macrophage infiltration seems to be preferentially associated with 8q gain, explaining the class 2 inflammatory “subset D” in Figure 14; T-cell infiltration preferentially associates with monosomy 3 and *BAP1*-inactivated tumors, probably through the upregulation of genes implicated in immunosuppression resulting from *BAP1* loss, in almost 30% of monosomy 3 UMs.^{260,424,431,432} The immune infiltration in a subset of UM is therefore paradoxically characterized by an immunosuppressive inflammation.²²³ In a “mixed lymphocyte reaction”, UM cells can inhibit T-cell proliferation by expressing immune checkpoint blockade receptors and ligands such as PD1, PDL1, CTLA4, LAG3 and others.^{260,421,433,434} While a causal role for macrophage and T-

cell infiltration in conferring poor prognosis in UM (rather than solely associating to the class 2 subtype due to monosomy 3 and 8q gain in these tumors) has yet to be determined, characterization of immune response and inflammatory cells in MUM is critical as they may represent targets for immunotherapy-based treatments in UM.

TREATMENT

PRIMARY TUMOR

Local treatment of UM is usually very effective. Factors such as tumor size, location, retinal invasion or detachment and others are taken into consideration for management of PUM patients, which also depends on the country. Classical treatments include radiotherapy (mostly plaque brachytherapy or proton beam therapy) and enucleation (removal of the eyeball), which are equally efficient although radiotherapy has the advantage of being a globe-preserving therapy.^{223,435,436} Brachytherapy is preferred for small and accessible tumors, while larger and less accessible posterior-located tumors are better treated by proton beam therapy (associated with minimal scatter, lower collateral damage and minimal risk of local tumor recurrence) or enucleation, the latter being mostly used for advanced tumors, such as those with orbital invasion, large diameter and thickness.⁴³⁵ Radiotherapy sometimes precedes enucleation, for example for tumors with retinal detachment.^{437,438} Local resection is another method that, like radiotherapy, can preserve the eye globe; it is most often used for tumors that cannot be treated by radiotherapy, due to for example large tumor size and juxta-papillary location.^{439,440} Adjuvant chemotherapy (using fotemustine) has been attempted in UM patients at high risk of metastasis, but did not improve survival.⁴⁴¹

There is no specific treatment modality for primary UM tumors, as few studies have compared the treatment options in this rare disease, and follow-up of patients treated by different methods indicates no difference in survival. The preferred or most widely used treatment varies from one country to another, where proton beam therapy is now regarded as the preferred method when available, although brachytherapy is still most widely used.²²³

METASTASIS

On average, 50% UM patients will ultimately develop metastases, which remain without curative treatment and are associated with dismal prognosis. The overall survival after diagnosis of metastasis has consequently not improved over the past 30 years, not exceeding 12 months. Common chemotherapeutic drugs have been tested in clinical settings but have proved largely unsuccessful. While there is no standard first-line treatment in the management of MUM, numerous clinical trials are ongoing and novel therapeutic strategies are emerging as the biological and pathophysiological understanding of the disease is improving; a first treatment improving overall survival in MUM has been approved this year. We will present a brief overview of therapeutic strategies in UM, but systematic reviews are available for detailed reports of current therapeutic options, such as Rodriguez-Vidal *et al*⁴⁴² or Carvajal *et al*.⁴⁴³

CONVENTIONAL CHEMOTHERAPY

Until recently, and like for many other cancers, conventional chemotherapy was the major option for treatment of UM, despite its very limited efficacy and due to the absence of other treatment options. Most chemotherapy agents used in cutaneous melanoma were tested in metastatic UM, but response was significantly lower in the latter due to major chemoresistance, resulting in poor survival rates ranging from 4 to 17 months.⁴⁴² Dacarbazine, cisplatin, temozolomide, fotemustine, bendamustine, and multiple others were attempted, but resulted in poor response rate and no clinical improvement.⁴⁴³⁻⁴⁴⁷ More modern chemotherapeutic agents like paclitaxel, vincristine or docosahexaenoic acid did not yield better results.⁴⁴⁸ In recent years, progression-free interval of up to 5 months and overall survival not exceeding 13 months have been most consistently observed with the combination of dacarbazine and temozolomide,⁴⁴⁸ and encouraging results were seen with gemcitabine and treosulfan,⁴⁴⁹ but overall no systemic chemotherapeutic treatment gave meaningful clinical improvement.^{449,450}

HEPATIC THERAPIES

In almost 90% of cases, UM metastasizes to the liver, via dissemination through blood vessels after which hepatic metastases are supplied by hepatic artery branches, unlike normal liver. Therefore, liver-directed therapies (both intra-tumoral and regional) have been attempted, including surgical resection, intra-arterial chemotherapy and hepatic chemoembolization.²²³ Regarding intra-arterial immunotherapy, some clinical trials compared intravenous and intra-arterial or other types of hepatic perfusion of chemotherapeutic agents, such as fotemustine or melphalan, but did not improve overall survival.^{451,452} Hepatic chemoembolization combines hepatic artery embolization with infusion of chemotherapeutic agents, reducing systemic toxicity while achieving high pharmacological concentrations. Again however, clinical trials did not lead to improved survival.^{453,454} When possible, surgical resection of metastasis leads to longest average survival (12 months to 2 years),^{455,456} but studies are biased as patients must meet a number of criteria reflecting a lower progression rate to be eligible to such procedure.

TARGETED THERAPIES

Multiple biology-based therapies are currently being investigated owing to the progress made in understanding the biological pathways underlying UM, such as those downstream $G\alpha_q$ signaling pathway described earlier. Multiple clinical trials focused on MAPK inhibitors, exploring various drugs such as sorafenib, imatinib or sunitinib, either alone or in combination with chemotherapy, reaching a median overall survival of 6 months.⁴⁴² Combined inhibition of $G\alpha_q$ downstream effectors MEK and PI3K, PI3K and mTOR, and Mdm2 and PKC have shown promising results in UM cell lines, much more than by targeting of a single molecule, and many are moving towards assessment in clinical trials.^{292,457,458} However, these molecules did not perform as well in patients in clinical trial settings.^{252,459,460} Currently ongoing clinical trials targeting epigenetic processes (epidrugs) are testing Vorinostat (a histone deacetylase HDAC inhibitor) and PLX2853 (a small molecule BET inhibitor) in phase II,²²³ as additional studies investigating other downstream effectors of $G\alpha_q$, $G\alpha_q$ itself or epigenetic modifiers are showing promising results in pre-clinical settings.^{461,462} Targeting epigenetic modulators represents an attractive strategy considering the chromatin remodeling effects resulting from *BAP1* inactivation. Another interesting pathway to target is the YAP/TAZ pathway which was

recently thoroughly studied for a role in UM downstream $G\alpha_q$, but no inhibitors are available for testing to date.

IMMUNOTHERAPY

Several immune approaches, such as immune checkpoint blockade, have been attempted in UM but led to far less success than in cutaneous melanoma therapy so far. Agents targeting CTLA-4 (an immune checkpoint receptor expressed in UM), such as ipilimumab tested in multiple studies, yielded response rates <5% and median overall survival and progression free survivals below 6 months.^{463,464} Similar strategies were attempted with PD1 inhibitors, again resulting in low response rates (5–10%), considerably lower than in cutaneous melanoma, and associated with median survival of <1 year.⁴⁶⁵ However, a recent analysis of PUM and MUM samples at single-cell level revealed that PD1 and CTLA4 were not always highly expressed in tumor-infiltrating immune cells, and that LAG3 was predominant in CD8+ T-cells, providing new insights into strategies for immune checkpoint blockage in UM.⁴²¹ New immune strategies are currently emerging, such as T-cell therapy with or without combination with immune checkpoint inhibitor ipilimumab, currently in phase I/II trials.^{223,466}

Lastly, and importantly, T-cell redirection strategies have recently proven to be of high potential in the treatment of MUM,⁴⁶⁷ and a few months ago a novel immunotherapy drug, Tebentafusp, employing this strategy, has proven to significantly improve overall survival (21.7 months) in MUM patients in phase III clinical trials (NTC03070392).^{468,469} This drug (IMCgp100) is a bispecific fusion protein, targeting both the gp100 antigen presented by melanoma cells through a high-affinity, HLA-A2:01 restricted T-cell receptor binding domain, and T-cells on the other end (antibody portion targeting CD3) to redirect T-cells specifically to UM cells.⁴⁷⁰ Hopefully, this drug, in the process of being registered, continues to demonstrate efficacy in increasing overall survival of patients and can become the first standard treatment for metastatic UM patients.

EPIDEMIOLOGY

Although it is the most prevalent form of eye malignancy in adult, UM is considered as a relatively rare cancer with a mean age-adjusted incidence of 5.2 cases per million in the United States (US) (based on 4,999 UM cases from the Surveillance, Epidemiology, and End Results [SEER] program), which has remained stable in the last decades.⁴⁷¹ Most studies indicate an equal incidence of UM in men and women while recent studies suggest a slight predominance of UM in men.^{472,473} The median age at diagnosis is 62 years in Caucasian populations such as the US and Europe,^{363,473} with most cases arising between the ages of 50 to 70²²³ while occurrence in childhood is rare.⁴⁷⁴ The 5- and 15-year UM-related mortality of patients with malignant UM is 31% and 45%, respectively,⁴⁷⁵ metastasis occurs within 30% to 50% of all UM cases,⁴⁷⁶ almost invariably (90%) to the liver, and median survival with UM involving the liver does not exceed 13.5 months.⁴⁷³

In addition to sex, UM incidence also varies considerably according to ethnicity and country. In the US, the incidence is ~15 times higher in non-Hispanic whites (age-adjusted incidence per million of 6.02) than in Asians/Pacific Islanders and Black populations (0.38 and 0.31 respectively).⁴⁷³ In Asian countries such as South Korea and Japan, incidence is similar to Asians in the US.^{477,478} Strikingly, the European incidence of UM ranges from less than 1 to 10 cases per million per year depending on the population;²⁴⁹ incidence follows a South to North increasing gradient, with smallest incidences in southern Italy and Spain compared to northern European countries such as Norway, Ireland and Denmark where incidence is up to 4-times higher,^{223,479,480} overall ranging from <2 per million population to >8. European ancestry is a major risk factor for UM,^{476,481} in accordance with 98% of UM cases in the US arising in Caucasian populations.^{249,482}

Additional UM risk factors in these populations mainly include light eye color and fair skin,⁴⁸³ but also an inability to tan/increased tendency to sunburn and the presence of uveal (mostly choroidal) nevi.⁴⁸⁴ Uveal nevi are benign melanocytic tumors thought to arise in almost 5% of the US population of at least 40 years of age; nevi with large basal diameter (>2mm) and giant nevi (>10mm) confer higher risk than smaller nevi.⁴⁸⁴ Interestingly, a recent study has identified mutations of *GNAQ* or *GNA11* in almost all nevi studied,²⁷⁵ strengthening the

hypothesis that the malignant transformation pre-existing choroidal nevi may explain at least a portion of choroidal melanoma cases.^{484,485} It is estimated that 1/10 to 1/5 UMs develop from the malignant transformation of a choroidal nevus, which are thought to confer a 0.2% adjusted lifetime risk of UM.⁴⁸⁶ Lastly, dysplastic nevus syndrome and ocular/oculodermal melanocytosis have also been shown to confer UM risk.^{481,487,488} Oculodermal melanocytosis (ODM), or nevus of Ota, manifests as a dermal melanocytosis that produces a blue/brown hyperpigmentation of the eye (sclera, iris, choroid), eyelids and surrounding skin, mostly unilaterally. Although Ota nevi are benign melanocytic tumors, 1 in 400 patients with ODM are thought to develop UM over a lifetime⁴⁸⁹ and are also associated with increased risk of metastasis.⁴⁸⁷ Interestingly, *GNAQ* activating mutations are observed in ~15% of Ota nevi;⁴⁸⁸ a recent study reported an initial *GNAQ* mutation followed by subsequent *BAP1* pathogenic mutation in a cutaneous melanoma case derived from an Ota nevus, in a two-step progression model similar to UM.⁴⁹⁰ Ota nevi are due to excess melanocytes in the regions surrounding the eye outlined above, and arise from an incomplete migration of neural crest melanocytes;⁴⁸⁸ they are evocative of post-zygotic mosaicism of *GNAQ* activating mutations in melanocytes.⁴⁹¹

Other than the above-mentioned clinical and epidemiological characteristics that confer UM risk, a few environmental factors and some important genetic factors are also implicated in UM susceptibility.

ENVIRONMENTAL RISK FACTORS

Multiple environmental factors were reported to confer UM risk, such as welding,⁴⁹² which is often coupled with exposure to other chemicals such as asbestos,⁴⁹³ but also UV exposure, occupational cooking, chemicals such as pesticides and formaldehyde²²³ and exposure to sunlight. For a meta-analysis of all UM risk factors reported and their associated OR, please see Nayman *et al.*⁴⁹²

Sunlight exposure is of particular interest here, as it is one of the weakest environmental risk factors, with an OR=1.37 (95% CI 0.96 - 1.96). Iris nevi, however, are also risk factors for UM,⁴⁹⁴ consistent with a link between iris nevocarcinogenesis and sun exposure.⁴⁹² The fact that the association of iris nevi are a weaker UM risk factor than cutaneous nevi or atypical nevi⁴⁹² could be the reflection of the low frequency (~4%) of iris melanoma among all UM types. In

this specific UM subtype, there is much less controversy regarding the implication of UV exposure in explaining UM risk.

IRIS MELANOMA AND EXPOSURE TO ULTRA-VIOLET RADIATION

Factors that increase the chance of developing iris melanoma include light eye color and UV exposure, which more strongly affects the inferior half of the iris²⁵⁰ due to a better protection of the superior half from sunlight by the orbital rim and upper eyelid.⁴⁹⁵ Consistent with a higher role for environmental (UV) exposure in this UM subtype, iris melanomas often arise earlier than posterior UMs of the choroid or ciliary body, with a relatively higher frequency of iris melanoma in young individuals (children and adults <21 years of age) than in adults.^{250,496,497} Consistent with the implication of UVR exposure in the risk of iris melanoma, a recent whole-genome sequencing (WGS) study on 103 UM tumor samples coming from all sites of the uveal tract identified a high tumor mutation burden (TMB) and an UVR mutational signature (SBS7a, SBS7b and DBS1)²⁵ restricted to the iris tumors,²⁶⁵ further emphasizing the major biological and clinical differences between anterior (iris) UM and their posterior counterparts. Iris melanomas have the lowest incidence of metastasis, associated with a better prognosis²²³ which can be at least partly explained by the earlier diagnosis resulting from greater visibility than tumors with the posterior part of the eye.

GENETIC SUSCEPTIBILITY

FAMILIAL UM

Over the past decades and since the 1970s, multiple reports of the clinical observation of two or more occurrences of UM within close relatives (mostly first, but also second-degree relatives) at a higher frequency than would be expected by chance considering the rarity of the disease, led to the initial hypothesis of a hereditary cancer predisposition to UM.⁴⁹⁸⁻⁵⁰¹ Owing to the rarity of UM, the expected frequency of two or more UM cases occurring by chance in first degree relatives of a family would be expected to be as little as 0.0002,⁵⁰² yet around 1% of all UM cases occur in a familial context (familial UM, FUM, characterized by the diagnosis of UM in at least two family members).⁵⁰³ Based on clinical observations of 56

patients in 27 families (out of 4,500 patients with UM), an autosomal dominant mode of inheritance of a predisposing gene was postulated.⁴⁹⁹ In addition, multiple cases of co-occurrences of different tumor types within families, especially CM, prostate cancer and breast cancer, were also reported and supported the hypothesis of genetic risk factors to explain FUM.^{501,504} Diener-West, 2005 #12345 It is now thought that almost 12% of all UM cases have a high risk of hereditary cancer predisposition, characterized by a strong personal and/or family history of cancer.^{223,503} A study of a large prospective cohort of over 2,000 choroidal UM cases reported the existence of a personal history of second primary cancer in 10% of patients, of which the most common forms were prostate cancer, breast cancer and leukemia.⁵⁰⁵ Conversely, initial development of a UM was found to significantly increase the risk of developing additional tumors, including CM, thyroid and renal cancers but also UM of the other eye (bilateral).⁵⁰⁵ In this cohort, UM patients had a cumulative risk of 15% of developing second cancers over 10 years. Other studies have confirmed the association of some UM cases with second primary cancers, including CM, breast cancer, multiple myeloma, non-Hodgkin lymphoma, urinary tract and gastrointestinal tumors, liver and pancreatic cancers.^{503,506-509} UM patients, with or without familial occurrence, have an increased risk of developing a second primary cancer compared to the general population; this risk is estimated to be 4 times higher in FUM cases compared to the general population.⁴⁹⁹

Other than FUM occurrence and personal history of multiple cancers, additional clinical observations point towards a role for genetic risk factors explaining a portion of UMs. Singh and colleagues observed a number of bilateral UM cases (0.18% of 4,500 UM patients) largely exceeding the calculated expected occurrence, suggesting the involvement of germline predisposition and an underlying cancer predisposition syndrome.²⁵⁵ Of note, FUM is usually unilateral.⁵⁰⁶ Furthermore, while UM occurrence in children or young adults is rare, it was found that oculodermal melanocytosis (ODM) was 9 times more prevalent in young UM cases compared to adult UMs.⁵¹⁰ Perhaps linking FUM to clinical characteristics, the presence of atypical dysplastic nevi is frequently observed within FUM or co-occurrence of UM and CM.^{496,500,501,511}

The clinical characteristics mentioned above (familial cases, personal and/or familial history of other tumors, bilateral UM occurrence, ODM) were all evocative of the implication of genetic factors in inherited predisposition to UM.

Searching for genetic predisposition factors, germline deleterious mutations of *BAP1* gene were identified in 5 family members of one UM proband, as well as in cases of familial and sporadic mesothelioma, some of which had also previously been diagnosed with UM.^{307,310} The studies reported germline *BAP1* mutations associated with bi-allelic inactivation in the tumor, and were the first to describe a *BAP1*-related syndrome predisposing to UM, mesothelioma but also to some extent meningioma. As previously stated, *BAP1* is a major TSG driver gene in UM, with frequent inactivating somatic mutations (described in the “Genetics of UM” chapter), consistent with germline *BAP1* mutations predisposing to UM as the first of two hits driving tumor growth. *BAP1* has since then been associated with predisposition to multiple cancers as part of the *BAP1* tumor predisposition syndrome (*BAP1*-TPDS), which now includes UM, mesothelioma,³¹⁰ CM and multiple melanocytic tumors,^{311,312} renal cell carcinoma,³⁰⁹ basal cell carcinoma³¹³ and potentially multiple other cancers.³⁰⁸ *BAP1* germline deleterious mutations are thought to mediate cancer risk following an autosomal dominant pattern of inheritance, with incomplete penetrance.⁵¹² *BAP1* germline mutations as part of the *BAP1*-TPDS often predispose to highly aggressive tumors associated with higher propensity to metastasize and poor prognosis compared to their wild-type *BAP1* counterparts.⁵¹³ Furthermore, *BAP1* germline mutations may also explain rare forms of UM; they were recently identified in bilateral and multifocal UM cases, confirming the previously formulated hypothesis of an inherited predisposition to explain these rare occurrences.⁵¹⁴

To date, *BAP1* is the only known high penetrance UM susceptibility gene.^{83,515} Recent studies estimate the frequency of germline *BAP1* mutations in unselected UM cases at ~2–4%,³⁰⁴⁻³⁰⁶ slightly higher than in CM,⁵¹⁶ suggesting the implication of other genes in UM development. Furthermore, ~0.6–1% of all UMs occur in a familial context (FUM),^{503,506} where inherited risk is expected to account for a significant proportion of co-occurrences between family members. Yet, although germline *BAP1* mutation frequency increases to 18–22% in this context,⁵⁰⁶ around 80% of the familial risk of UM remains unexplained, suggesting the existence of other predisposing genes to explain these familial cases.

OTHER RARE GERMLINE MUTATIONS

The TCGA project allowed the search of additional pathogenic or likely pathogenic germline variants within 80 UM cases.^{83,260} In addition to *BAP1*, rare germline mutations of *FANCM*, *POT1*, *PRDM9*, *PTCH1* were identified and considered to be pathogenic variants in UM, while *FANCA*, *FANCD2* and *NBN* were categorized as likely pathogenic variants or of unknown significance (VUS).⁸³ However, according to an evidence-based framework developed by Strande and colleagues to measure the strength of identified gene disease associations,⁵¹⁵ only *BAP1* shows strong evidence of association with hereditary UM. *POT1* is the only gene among UM pathogenic variants listed above for which an independent case report corroborates a potential role in UM predisposition, where two truncating variants in *POT1* were found in individuals affected by both UM and CM, associated with longer telomeres.⁵¹⁷ These are the only *POT1* pathogenic variants reported in UM so far. By performing whole-exome sequencing, another study identified potentially pathogenic variants associated with hereditary UM (FUM) or high-risk of cancer (UM and familial or personal history of other cancers), within known hereditary cancer genes *PALB2*, *MLH1*, *MSH6*, *CHEK2*, *SMARCE1*, *ATM*, *BRCA1* and *CTNNA1*.⁵¹⁸ However, only *PALB2* and *MLH1* showed moderate association with UM hereditary predisposition, characterized by a bi-allelic inactivation in tumor. Pathogenic variants in all other genes except for *SMARCE1* were found at similar frequencies in UM patients and non-cancer controls. A single case report⁵¹⁹ had potentially implicated *MLH1* in a UM patient with a family history of Lynch syndrome (LS), who demonstrated microsatellite instability (MSI), loss of MLH1 mismatch repair protein expression in her UM tumor, and an uncommon ovarian metastasis. However, evidence is strongly lacking to integrate UM in the list of LS-associated tumors or to confirm the causative role of MLH1 in UM, and it has not been ruled out that in this case, UM may simply have occurred coincidentally in a LS patient. For *PALB2*, searching databases of germline mutations revealed one additional proband who initially developed breast cancer, in which a predisposing role for *PALB2* is well known; however, there was no evidence of loss of heterozygosity in tumor.⁵¹⁸ *SMARCE1* novel pathogenic variant was detected in FUM cases in a patient with co-occurrence of UM and endometrial carcinoma, but somatic analysis of mutational status was not available.⁵¹⁸ Finally, single UM case reports have described the presence of germline mutations in *CDKN2A*,⁵²⁰ *MSH6*⁵²¹ and *BRCA1*⁵²² and in all cases, UM was associated to other tumors (personally or

within the family) in which the respective genes are known to be implicated. MSH6 and MLH1 being implicated in DNA mismatch repair (MMR) pathway, a role for MMR mutations in UM has been suggested; however, in a recent study, out of six MMR genes searched within the TCGA, Catalogue of Somatic Mutations (COSMIC) and University of California, San Diego (UCSD) databases, mutations were found to be exceedingly rare in UM, and no mutations were reported for either *MLH1* or *MSH6*.⁵²³ Finally, *BRCA2* germline mutations were also detected in UM cases with history of breast or ovarian cancers,^{524,525} but functional assessment of the mutations were limited and reported mutations were not always verified for a deleterious effect.

Taken together, while *MLH1* and *PALB2* show moderate evidence of association with UM and other rare sporadic associations of UM with germline pathogenic variants of *BRCA2*, *MSH6*, *CDKN2A* and others have been reported, statistical and functional validations of these genes in UM predisposition are still lacking. Most of these genes have been identified through WES studies, and therefore multi-testing of thousands of genes may explain the occurrence of some of the pathogenic variants identified, highlighting the need for validation studies of these genes by single-gene testing in independent cohorts to determine the true relative risk of candidate predisposition genes. So far, these genes have not been reported in families with multiple occurrences of UM. *BAP1* remains the only predisposition gene significantly associated with UM risk (and even more so with FUM), and it is likely that other more common, yet lower penetrant genes or alleles, may further explain the familial occurrence of UM.

MODERATE TO LOW-RISK COMMON VARIANTS

Added to the fact that high penetrance *BAP1* germline mutations only explain a small portion of genetic risk in UM, the prevalence of UM in individuals of European ancestry suggests the existence of susceptibility alleles with low to moderate penetrance to explain this epidemiology. In 2017, a genome-wide association study (GWAS) conducted by our team in European UM cases and controls identified a candidate risk region on chromosome (chr) 5p15.33 within the *TERT/CLPTM1L* locus, marked by lead SNP rs421284 with odds ratio (OR) of 1.7 (confidence interval CI 1.43 - 2.05), which was further replicated in an independent series.⁵²⁶ A second locus on chromosome 15 on the *HERC2/OCA2* locus almost reached the statistical significance threshold (Figure 16). Unlike these two regions, other isolated SNPs with high p-value were inconsistent with linkage disequilibrium (LD) pattern of surrounding SNPs.

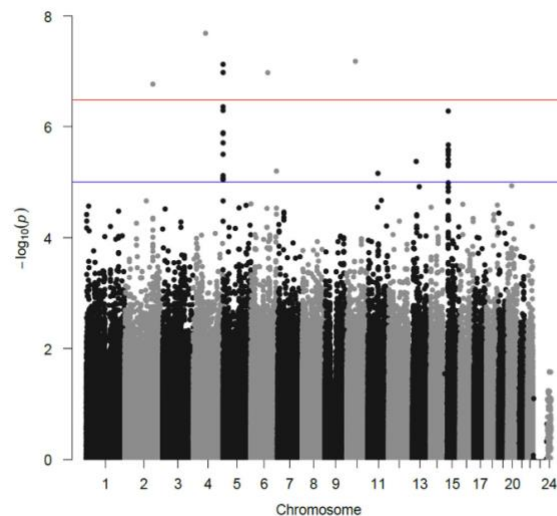


Figure 16. Manhattan plot of a GWAS conducted in UM.

259 UM cases and 401 controls were included in this discovery series. The association test P-values (y-axis, log10 scale) of individual SNPs (dots) are plotted against their chromosomal position (x-axis). Horizontal blue line represents tendency for association ($P < 1 \times 10^{-5}$) while SNPs above the red line reached statistical significance threshold ($P < 3.3 \times 10^{-7}$). Figure from Mobuchon *et al.*⁵²⁶

TERT/CLPTM1L LOCUS ON CHROMOSOME 5

In UM, the *TERT/CLPTM1L* risk locus is marked by multiple associated SNPs that follow a LD pattern consistent with degradation of LD as physical distance increases from the two lead

SNPs, rs421284 and rs452932, in high LD with one another ($r^2>0.9$).⁵²⁶ The two SNPs are common variants (minor allele frequency MAF=0.466 and 0.465 respectively in the general population based on the genome aggregation database gnomAD v2.1.1) that confer moderate UM risk with OR>1.7. They lie within intron 8 of the *CLPTM1L* gene while other risk SNPs span the entire length of the *CLPTM1L* (intronic) sequence. *CLPTM1L* encodes the cleft lip and palate transmembrane protein 1-like, while the most nearby gene sharing the same locus, *TERT*, encodes the telomerase reverse transcriptase (catalytic subunit of the telomerase enzyme). However, these SNPs do not account for the European prevalence of UM as the risk allele frequency is similar in populations of European and African ancestries.⁵²⁶

The 5p15.33 *TERT/CLPTM1L* locus has been extensively characterized as a multi-cancer risk locus, both in cancer-specific association studies and in meta-analyses.^{183,527-529} Well over 10 tumor types other than UM have been reported for cancer association within the >1 Mb region at 5p15.33, including breast (overall and different subtypes),¹⁵⁷ colorectal,¹⁶⁴ lung,¹⁶⁰ pancreatic,¹⁶¹ prostate,⁵³⁰ renal,¹⁶⁶ ovarian,¹⁵⁸ head and neck,⁵³¹ esophageal,⁵³² endometrial,⁵³³ testicular germ cell⁵³⁴ cancers as well as cutaneous melanoma^{171,535} and glioma.⁵³⁶

Characteristic of this pleiotropic locus is the presence of multiple independent signals and risk alleles, recently confirmed by a large-scale cross-cancer fine-mapping study of 14 cancer types that revealed ten independent associations in this region.⁵²⁹ These span the entire length of the *TERT* and *CLPTM1L* genes. Strikingly, in this study, pairwise cancer analysis of local genetic correlation revealed a strong positive correlation between melanoma and pancreatic cancer and a strong inverse correlation between prostate cancer vs. glioma and lung vs. pancreatic cancers respectively. This is consistent with the previous pan-cancer study conducted by Wang *et al*⁵²⁷ that had characterized at least six independent loci within 5p15.33 risk region. Specifically, “Region 2” in this study, initially marked by rs401681, is associated with risk of melanoma, pancreatic, lung, testicular and bladder cancers, and belongs to the same association signal as the one identified in UM. Similarly, in the most recent large-scale study by Chen *et al*, a region marked by rs465498 index SNP shared between pancreatic cancer, cutaneous melanoma and lung cancer in varying directionalities

is also in high LD with UM lead SNP rs452932, with the same direction of effect as cutaneous melanoma and pancreatic cancer (and inverse to lung cancer).

The pleiotropic effect of this multi-cancer risk locus marked by multiple independent association signals and in various directionalities suggests a region of shared inherited susceptibility to cancer, and points towards highly tissue- or cancer-specific regulation and underlying carcinogenic mechanisms, highlighting the need for local functional characterization of this risk locus in individual tumor types to understand the underlying pathophysiology.

PATHOPHYSIOLOGICAL MECHANISMS IN CANCER

The numerous independent signals at 5p15.33 and various directionalities make it a complex multi-cancer risk locus in which underlying functional mechanisms may be shared by some cancer types but not others. Additionally, it is highly likely that even independent “sub-regions” of risk loci, such as Region 2 of interest here which is a susceptibility region for multiple cancers,⁵²⁷ may consist in some tumor types in more than one independent signal while most risk may be explained by a single SNP in others. This point is illustrated by the study of Fang and colleagues¹⁸³ who identified a functional SNP within a *CLPTM1L* intron in “region 2”, mediating risk through the allele-specific regulation of *TERT* via ZNF148 binding. This SNP may explain a significant portion of cancer risk for testicular, pancreatic and lung cancers, yet the risk and protective alleles of this SNP differ among these cancer types and data further suggests that in the case of lung cancer, rs36115365 only partly explains cancer risk conferred by Region 2.

On one hand, the *CLPTM1L* protein and function have not extensively been characterized and are not fully understood. *CLPTM1L* is known to be over-expressed in cisplatin-resistant ovarian tumor cell lines,⁵³⁷ where it confers resistance to treatment and is associated with poor outcome in ovarian serous adenocarcinoma.⁵³⁸ *CLPTM1L* is also amplified (along with *TERT*) through copy number gains of non-small cell lung cancer tumors (NSCLC),⁵³⁹ and is the only overexpressed protein among the 5p15.33 genomic region resulting from duplication of chromosome 5p in cervical cancer cell lines.⁵⁴⁰ Furthermore, studies in lung cancer where 5p15.33 is a well-known susceptibility region have demonstrated a potentially oncogenic and anti-apoptotic role for this protein by showing that *CLPTM1L* regulates cell survival signaling

in lung cancer cells, where it regulates anti-apoptotic Bcl-xL protein, interacts with phosphoinositide 3-kinase (PI3K), participates in Ras transformation⁵⁴¹ and further promotes radioresistance in NSCLC.⁵⁴² In pancreatic cancer cells, CLPTM1L promotes growth, its overexpression enhances aneuploidy,⁵⁴³ and the protein has been shown to confer chemoresistance and cell survival through interaction with GRP78, cell surface relocalization and promotion of Akt signaling upon ER stress.⁵⁴⁴

On the other hand, the telomerase enzyme (made of a catalytic subunit encoded by *TERT* combined to an RNA template *TERC*) is well known for its role in the maintenance and elongation of telomeres protecting the end of chromosomes and maintaining genome integrity, but also as a major hallmark of cancer. Although *TERT* is normally repressed in most adult tissues, reactivation of telomerase (and of its rate-limiting enzyme *TERT*) following gradual (and normal) telomere attrition is one of the mechanisms that may be acquired by tumor cells to escape from crisis caused by telomere attrition which would otherwise lead to cell cycle arrest, senescence and apoptosis as part of a tumor suppressor pathway.⁵⁴⁵ Recurrent mutations of *TERT* are highly associated with risk of cutaneous melanoma,⁵⁴⁶ are also frequent in conjunctival melanomas⁵⁴⁷ but are exceptionally rare in UM.⁵⁴⁷⁻⁵⁴⁹

PIGMENTATION POLYMORPHISMS

The *OCA2* and *HERC2* genes play a major role in determination of eye color, as previously mentioned. The *OCA2/HERC2* “tendency” locus (i.e., not reaching empirical significant threshold) identified in the UM GWAS⁵²⁶ is also a major susceptibility region for cutaneous melanoma (CM),⁵⁵⁰ in addition to being a major determinant of pigmentation traits including both eye and hair color. Searching for shared genetic susceptibility between CM and UM, a targeted association study in UM of SNPs known to be risk factors for CM identified SNPs within two pigmentation loci, *HERC2/OCA2* locus on chr15, with most associated SNPs also present in UM GWAS (rs12913832, rs11074306, rs3930739) and *IRF4* on chr6 characterized by a single SNP, rs12203592.⁵⁵¹ *IRF4* SNP was also detected in the UM GWAS with high OR=1.88, but once again did not reach statistical significance. These SNPs are part of the six polymorphisms used to predict human eye color with the IrisPlex system (see “Determinants of eye color” chapter).²⁴² Of note, *OCA2/HERC2* variants were also shown to affect time to first cutaneous squamous cell carcinoma.⁵⁵²

Despite their striking differences in terms of oncogenesis and metastatic progression, UM and CM thus share common genetic predisposition factors, characterized by a common role in pigmentation. This is particularly interesting, as strong genetic predisposition in these two melanoma subtypes are highly unsimilar, with frequent *CDKN2A* germline mutations in familial CM (essentially absent or extremely rare in UM) and conversely, *BAP1* germline mutations often described in familial UM (sometimes seen in CM, but tumor types within the *BAP1*-TPDS mostly involve other melanocytic tumors).²²⁴ Similarly, recurrent driver mutations in CM and UM radically differ (*BRAF/NRAF/KIT/NF1* mutations in the former are absent in UM, and vice versa for *GNAQ/GNA11/BSE* events in UM). Yet, UM and CM share a significant portion of common genetic risk, characterized by polymorphisms involved in pigmentation, in line with the association of both diseases with individuals of fair skin. Implication of pigmentation traits in UM is indeed consistent with the well documented association of blue eyes and fair skin in UM;⁴⁸³ however, while the implication of *IRF4* and *OCA2/HERC2* polymorphisms is consistent with the epidemiology of UM, it is intriguing that they confer UM risk since light and UVR hardly penetrates the uveal melanocytes other than within the iris as stated previously and, as discussed in the “Genetics of UM” section, there is an absence of UVR-associated signature in posterior UM.²⁵⁹ Future work is thus required to elucidate the pathophysiological mechanisms of pigmentation risk loci in UM, and more generally, to functionally characterized all low-risk susceptibility loci in UM, including *TERT/CLPTM1L* multi-cancer risk region.

MBD4

While a major part of the work presented here is based on a GWAS approach to identify common susceptibility alleles in UM, a candidate-gene approach was undertaken in parallel to further explore UM predisposition factors. This final introductory chapter provides background information on *MBD4*, for which a germline mutation (associated with somatic bi-allelic inactivation in the tumor) was identified by our team in an outlier UM patient responding to anti-PD1 therapy,⁵⁵³ prompting us to further investigate this gene in UM predisposition.

BASE EXCISION REPAIR PATHWAY

The MBD4 protein is a DNA glycosylase, implicated in the initiating step of the base excision repair (BER) pathway. The BER pathway is a type of DNA repair pathway that recognizes small base lesions, initiated by the removal of the damaged base, which is then repaired to avoid mutagenesis, cytotoxic strand breaks and further genomic instability.

The BER pathway is a multi-enzyme DNA repair process that involves sequential reactions catalyzed by specific enzymes and implicating multi-protein complexes. BER specifically repairs small, non-bulky and non-helical-distorting base lesions from the genome.⁵⁵⁴ Although these can arise from exogenous sources such as anticancer agents, they often arise spontaneously, induced by common chemical damages such as oxidation, deamination and alkylation events, resulting in modified bases that can lead to DNA mutations through incorrect base-pairing, abasic sites or breaks in DNA if they remain unrepaired.⁵⁵⁵ Example of frequent DNA damages repaired by BER include (i) oxidation of guanine to form 8-oxoguanine (8-oxo-G), (ii) hydrolytic deamination of guanine to xanthine or adenine to hypoxanthine (low rate), and cytosine to uracil or 5-methylcytosine (5meC) to thymine (high rate), and (iii) methylation of guanosine to 7-methylguanosine. Repair of such damaged bases essentially comprises 5 sequential reactions catalyzed by the following groups of enzymes: glycosylase, endonuclease, lyase, polymerase and ligase (Figure 17).^{554,556}

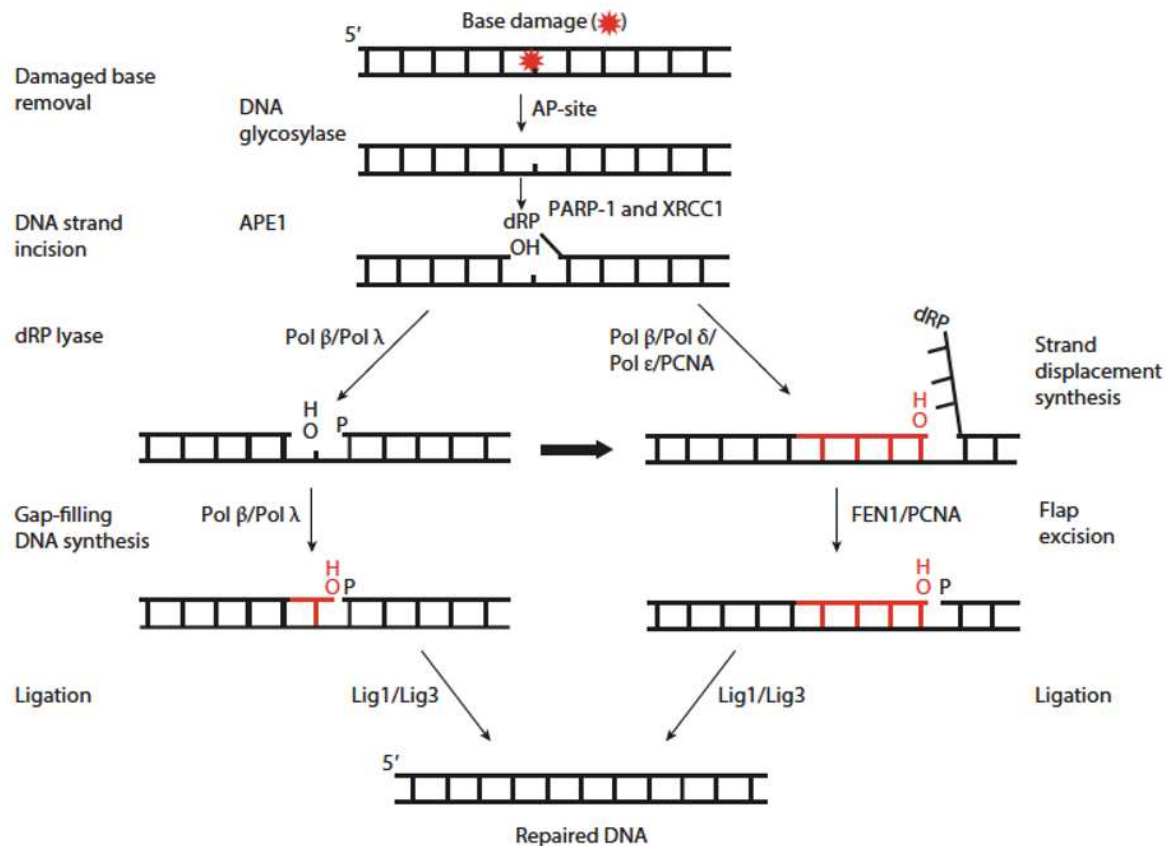


Figure 17. Canonical mammalian base excision repair (BER) pathway.

The initial step in BER pathway is the removal of the damaged base (red star) by a DNA glycosylase, although base loss can also occur spontaneously, resulting in an apurinic/aprimidinic (AP-) site. After cleavage of DNA directly 5' upstream the AP-site by AP-endonuclease 1 (APE1), 2 pathways can proceed: short-patch BER (left path) repairs single base damage via Polβ/Polλ gap-filling DNA synthesis, while long-patch BER (right path) involves strand displacement synthesis and flap excision requiring Polδ, Polε, PCNA (proliferating cell nuclear antigen) and FEN1 (flap endonuclease 1), resynthesizing 2-12 nucleotides. Ligation of newly synthesized bases is achieved by Ligases 1/3. Figure from Beard *et al.*⁵⁵⁴

BER is initiated by the activity of various DNA glycosylase enzymes (outlined below, MBD4 being one of them) that recognize specific substrates/types of lesions and remove the damaged base by distortion of the DNA double helix, flipping out of the mispaired base and cleavage of its 5' N-glycosylic bond,⁵⁵⁷ releasing the damaged base and leaving an apurinic/aprimidinic site (AP-site). Of note, AP-sites may also arise spontaneously, as one of the most common lesions in mammalian genome.⁵⁵⁸ While bifunctional glycosylases have both glycosylase and AP lyase activity, monofunctional glycosylases leave an intact phosphodiester backbone and thus require AP endonucleases (mostly APE1) to subsequently cleave the DNA backbone at the AP site, forming a single-strand break with a 3' hydroxyl (OH) on one end and a 5' deoxyribosephosphate (dRP) on the other.⁵⁵⁹ The pathway then splits into two paths that require different enzymes: short-patch BER (SP-BER) inserts a single

nucleotide, primarily catalyzed by polymerase (Pol) β (or Pol λ if absent),⁵⁶⁰ while long-patch BER (LP-BER) typically adds 2 to 12 new nucleotides and is mediated by Pol δ or Pol ϵ , with help of a processivity factor, proliferating cell nuclear antigen (PCNA), which helps displace the damaged strand. In SP-BER, polymerases also have a lyase domain that cleaves the 5'-dRP produced, and in LP-BER the polymerases perform a strand displacement synthesis that results in a 5' flap, which can then be cleaved by Flap endonuclease FEN1. Finally, the two paths converge at the last step, where ligation of the nick is catalyzed by DNA ligases (LigIII and I are preferred in SP- and LP-BER, respectively) along with cofactor X-ray repair cross-complementary (XRCC1), completing the BER pathway. Although the choice of SP or LP-BER is still under investigation, it is known to depend on several factors including availability of BER factors, cell cycle stage and physiological state, and the type of lesion; for example, it is thought that the pathway engages in LP-BER when the 5'-dRP group resulting from backbone cleavage does not allow Pol β lyase activity, such as at adenylated dRP or reduced and oxidized AP-sites.⁵⁶¹⁻⁵⁶³

DNA GLYCOSYLASES

DNA glycosylases are instrumental in the initiation of repair of damaged bases as they are responsible for their recognition, cleavage and removal. There are a total of 11 human glycosylases, divided into three types based on their function: monofunctional glycosylases only exhibit DNA glycosylase activity, requiring processing of the resulting AP-site by AP endonucleases, bifunctional enzymes also have AP-lyase activity allowing them to cleave the phosphodiester bond at the damaged base 3' end, creating a α,β -unsaturated aldehyde (via β -elimination reaction) that still requires APE1 processing, while Nei-like (NEIL) DNA glycosylases are also bifunctional but can cleave the phosphodiester bond either side of the removed base, in a β - δ elimination reaction that makes APE1 dispensable.⁵⁶⁴⁻⁵⁶⁶ Based on their functional and/or structural features, glycosylases can be divided into four groups (SF1-4): alpha-beta fold UDG superfamily (SF1), helix-hairpin-helix (HhH) superfamily (SF2), 3-methyl-purine glycosylase (MPG) superfamily (SF3) and NEIL hairpin-2-turn-hairpin superfamily (SF4).⁵⁶⁷

TDG AND MBD4 SUBSTRATE SPECIFICITIES

There are 6 monofunctional glycosylases, which participate in the removal of different types of base lesions depending on their specific substrates. SF1 glycosylases specifically target uracil (U) bases, which are formed upon cytosine deamination induced by oxidative stress. Common examples of SF1 monofunctional enzymes are uracil DNA glycosylases (UNG) and thymine DNA glycosylase (TDG). While monofunctional MBD4 (methyl-CpG binding domain protein 4) is structurally unrelated to TDG and belongs to the SF2 group, which harbors a very distinct structure with characteristic HhH motif, substrate specificity of TDG and MBD4 somehow overlaps, although MBD4 has a more specific substrate recognition. Indeed, while TDG was originally best known for its ability to recognize G:T lesions, it recognizes a broad range of lesions resulting from oxidation, alkylation or deamination of C, 5-methylcytosines (5meC), T and A.^{568,569} This includes recognition of G:U mismatches (where TDG actually exhibits higher catalytic activity than at G:T mismatches),⁵⁷⁰ oxidized/deaminated 5meC, 5-fluorouracil and 5-bromouracil.

MBD4 on the other hand, has substrate preference for G:T mismatches,⁵⁷⁰⁻⁵⁷² and more particularly for G:T (but also G:U) mismatches in a CpG (5'-cytosine-phosphate-guanine-3') context, where it catalyzes the removal of mispaired T and U. These G:T and G:U mismatches at CpG sites most often arise from the spontaneous hydrolytic deamination of 5meC to T and C to U respectively, and failure to remove these lesions before DNA replication results in C→T transition mutations at a high rate.⁵⁷¹⁻⁵⁷³ Importantly, deamination of 5meC is highly predominant at CpG sites and arise at a rate 3 times higher than that of deamination of C (which, in turn, is 50-fold preferred to purine substrates) highlighting the critical role of MBD4 in preventing DNA mutations arising from potentially highly mutagenic deamination of pyrimidine compounds, and particularly 5-meC at CpG sites. Therefore, although both TDG and MBD4 recognize G:T mismatches, MBD4 is specifically recruited at CpG sites and exhibits binding preference for 5meC over C, although it also recognizes other substrates (C and U) at these sites.^{569,574-576} 5meC is present at 1% of bases in mammals, making it the most abundant epigenetic modification of DNA.⁵⁷⁷ It is largely predominant at CpG sites where, as discussed later, methylation plays an important role in epigenetic control through temporal and spatial transcriptional repression, differentiation, genomic imprinting and others.^{575,577-579} MBD4 is

critical in the prevention of genomic instability and specifically C→T transitions, as one third of all intronic base pair mutations occur at CpG sites, where C→T transitions are the most frequent single base mutation type in human cancer.^{569,580,581}

MBD4 GENE AND PROTEIN STRUCTURE

The *MBD4* gene is located on chromosome 3q21.3, spanning a region of 9kb. Its coding sequence is 1,743bp long, spanning 8 exons and encoding a full-length protein of 580 amino acids (aa) (although several *MBD4* variants have been reported and are translated into shorter proteins of 574, 540, 572 and 262aa). MBD4 is unique in its structure compared to other DNA glycosylases, as it contains both a catalytic N-terminal methyl-CpG-binding domain (MBD) (aa 82-147) and a C-terminal DNA glycosylase domain (aa 426-580) (Figure 18). MBD4 is a core member of the MBD family of proteins, which all share a conserved MBD and harbor additional domains conferring specific catalytic functions and/or the ability to engage in specific interactions. MBD4 is the only MBD protein that harbors a glycosylase activity domain (Figure 18).

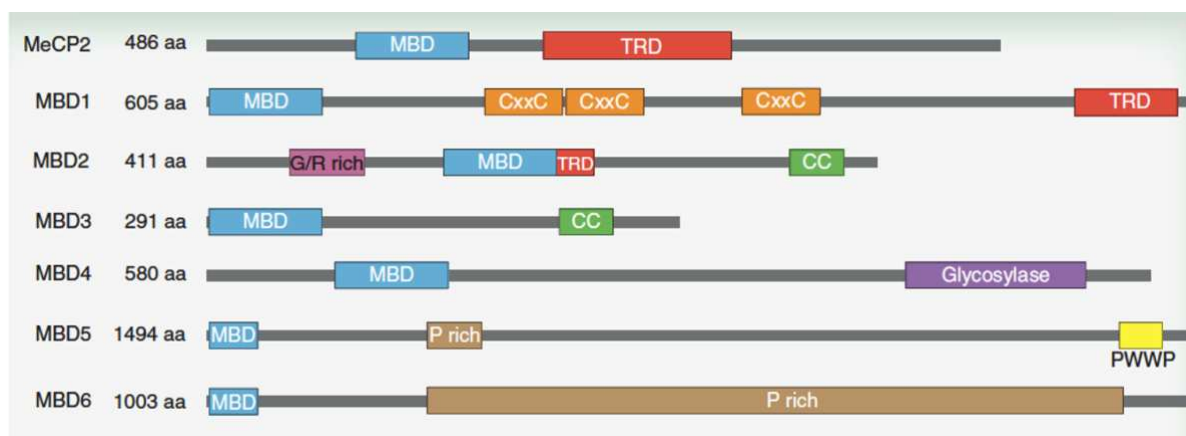


Figure 18. Methyl-binding domain (MBD) protein family.

The MBD family of protein consists of 7 members, methyl-CpG-binding protein 2 (MeCP2) and MBD1–6. MBD proteins share a conserved methyl-CpG-binding domain, and each have other specific domains, such as DNA-binding, catalytic or protein-protein interaction domains, reflecting the different roles of MBD proteins. TRD: transcriptional repression domain; CxxC: unmethylated-CpG-binding zinc finger domain; CC: C-terminal coiled coil domain; G/R rich: glycine-arginine repeat region; P rich: proline rich domain. Figure from Du *et al.*⁵⁷⁶

Crystallographic analyses of MBD4 domains have given insights on the mechanism of action of MBD4 to recognize 5meC at CpG sites and remove the required base. The glycosylase domain of MBD4, composed of 10 alpha helices that form a central cleft, binds DNA minor

groove at the site of the damaged base where it bends the DNA sugar-phosphate backbone at 57°, allowing to flip out the damaged base and dock it into its active-site cleft for catalysis. Specifically, three amino acids are critical for DNA binding and base flipping (Arg468, Thr469 and Ser470) while three others are required for T recognition and enzymatic activity.^{570,572,582-584} The fact that the glycosylase domain also provides a recognition pocket in which the opposing G nucleotide is stabilized through electrostatic interactions with active site residues further explains the strict recognition of G:T mismatches compared to A:T.⁵⁸⁴

However, the glycosylase domain itself is not able to recognize CpG sites, and it is therefore the MBD domain that confers substrate specificity to the enzyme.⁵⁶⁹ MBD domain of MBD4 shares sequence homology with other family members MeCP2 and MBD2⁵⁸⁵ and is made of 4 antiparallel beta-sheets, two of which are thought to interact with DNA major groove at the site of methylation,^{575,586} guiding MBD4 to genomic methylation-rich areas where spontaneous base deamination is most likely to occur.⁵⁸⁷ It is thought that the MBD binds meCpG sites near – but not directly on – the meCpG containing the lesion, as this is prevented by the binding of MBD4 glycosylase domain and the resulting bend in DNA minor groove.

Finally, and providing insights into the protein function, the long spacer domain between the two functional domains has been described to mediate interactions with other proteins such as E3 ubiquitin ligase UHRF1 and deubiquitinating enzyme USP7, which participate in the regulation of DNA maintenance methyltransferase (Dnmt1),⁵⁸⁸ and other epigenetic proteins including histone deacetylase I (HDACI).

BIOLOGICAL FUNCTIONS OF MBD4

Some of the functions of MBD4 have been elucidated decades ago, such as its implication in DNA repair and in the prevention of DNA mutagenesis, while other emerging roles have been proposed, although the exact biological mechanisms have yet to be elucidated.

SUPPRESSION OF MUTAGENESIS AND DNA REPAIR

As expected from its involvement in the initiating steps of the BER pathway, one of the first identified biological function of MBD4 due to its demonstrated thymine glycosylase activity⁵⁷³

was its role in protection against C→T transitions arising from spontaneous base decay (from deamination) and in DNA repair. Studies in mice 20 years ago demonstrated that *Mbd4*^{-/-} mice led to 3-times enhanced CpG mutation burden in the small intestine, and could further promote gastrointestinal tumorigenesis in mice also harboring a heterozygous mutation of adenomatous polyposis coli (*Apc*^{min/+}), where these mice showed high CpG>TpG mutations of the wild-type *Apc* allele.^{589,590} While *Mbd4* knockout did not induce tumorigenesis in mice by itself, these studies demonstrated the role of *MBD4* in suppressing mutation in mammals and further pointed towards a potential role for *MBD4* as a tumor suppressor, insufficient alone to initiate tumorigenesis but participating in acquiring a high mutation load that could potentially result in mutations of other cancer driver genes.

Corroborating the implication of *MBD4* in DNA repair, Bellacosa and colleagues reported the interaction of *MBD4* with *MLH1* (MutL homolog 1), a protein involved in the DNA mismatch repair (MMR) pathway. Defects in the *MLH1* gene are notably associated with hereditary nonpolyposis colon cancer displaying microsatellite instability (MSI).⁵⁷⁴ This work suggested a role for *MBD4* not only in BER initiation, but also in MMR pathway through formation of a complex with *MLH1*, consistent with the observation of frequent *Mbd4* mutations in MMR-deficient tumors exhibiting MSI in murine models.⁵⁹¹ However, *Mbd4*-deficient mice did not increase mutation burden nor accelerate tumorigenesis in MMR-deficient mice.⁵⁹² In line with *MBD4* mutations being associated with MSI context (and as discussed later in “*MBD4* deficiency in cancer” where *MBD4* somatic mutations are found in MSI-exhibiting tumors),⁵⁹³ *MBD4* contains two polyadenine stretches (A₆ and A₁₀) that are frequently affected by frameshift mutations and often result in a truncated protein lacking a functional glycosylase domain, explaining why *MBD4* is a target of MSI and why mutations of *MBD4* are often seen in MSI-related tumors.⁵⁹³⁻⁵⁹⁸

As a last evidence for a role of *MBD4* in DNA repair, *Mbd4* was shown to interact with DNA methyltransferase *Dnmt3b*, which led to reduced T:G mismatch repair efficiency upon loss of expression.⁵⁹⁹ A study in *Xenopus* embryos later suggested the co-recruitment of *MBD4*, *MLH1* and *DNMT1* at sites of DNA damage.⁶⁰⁰ Interactions were subsequently identified between human *MBD4* and *DNMT1*,⁶⁰¹ which coordinates with DNA damage repair pathways to protect from DNA mutagenesis.⁶⁰² However, as stated below, these interactions are more likely to be involved in transcriptional repression and/or apoptosis rather than DNA repair *per se*.

APOPTOSIS

An initial role for MBD4 in apoptosis was suggested when it was shown to interact with MLH1, which is functionally implicated in apoptosis as the MMR pathway recognizes mispairs resulting from damage induced by cytotoxic agents, such as O⁶-methylguanine, leading to cell cycle arrest and apoptosis.^{569,603} MLH1 defect has also been shown to confer resistance to apoptosis in tumor cells exposed to various DNA damaging drugs.^{604,605} However, although in murine models, reduced expression of MMR proteins and reduced apoptotic response was observed in embryonic fibroblasts deficient for *Mbd4*,^{603,606} double mutants *Mbd4*^{-/-} and *Mlh1*^{-/-} exposed to cytotoxic agents did not demonstrate an additive effect,⁵⁹⁰ suggesting that these proteins act in the same apoptotic pathway essentially through MMR proteins. A role for MBD4 in apoptosis became more apparent when Sreaton and colleagues demonstrated the direct nuclear interaction between MBD4 and Fas-associated death domain (FADD) protein,⁶⁰⁷ an adaptor protein that usually functions at the cell surface as a death receptor adaptor promoting apoptotic cascades required for activation of caspases. In their work, they demonstrated another (nuclear) function for FADD through interaction with MBD4 (and with MLH1), which regulated Fas ligand-, DNA damage- and cell-detachment induced-apoptosis.⁶⁰⁷ Similarly, the previously mentioned study in *Xenopus* embryos demonstrated that depletion of maintenance methyltransferase DNMT1 results in an activation of p53-dependent apoptotic response mediated by MBD4 and MLH1, and overexpression of xMBD4 and xMLH1 in embryos induced programmed cell death.⁶⁰⁰ Conversely, depletion of MBD4 and MLH1 increased survival of DNMT1-depleted embryos. Finally, Laget and colleagues further demonstrated a role for MBD4 in directing DNMT at sites of oxidation-induced DNA damage, where MBD4 is thought to help determine whether to initiate DNA repair and cell survival or induce cell death pathways.⁶⁰¹

TRANSCRIPTIONAL REPRESSION AND HISTONE DEACETYLATION

A first indication of an implication of MBD4 in epigenetic silencing is the observation that MBD4 strongly associates with heterochromatin, the tightly packed form of chromatin in which genes are untranscribed and silenced.^{585,588,600} Linking its functions in apoptotic regulation and gene silencing, MBD4 has been shown to recruit its interacting partners MLH1

and DNMT1 to heterochromatin sites in multiple studies,^{574,600,603} part of an MMR-dependent apoptotic process. The documented interactions between MBD4, MLH1, DNMT1 and FADD indicate a more global role for MBD4 in genome surveillance linked to apoptosis, DNA repair and the control of gene expression. Laget and colleagues, who demonstrated upregulation of MBD4 at sites of oxidative damage, also evidenced MBD4/DNMT1 binding and synergistic transcriptional repression at methylated CpG islands of *CDKN1A/p21* and *MSH4*,⁶⁰¹ indicating that MBD4 may participate in the repression or silencing of genes at methylated CpG (meCpG) sites. Indeed, DNA hypermethylation is known to be linked to heterochromatin, transcriptional repression and histone hypoacetylation as part of epigenetic gene silencing.^{608,609} In line with these findings, Konda and colleagues further reported an activity for MBD4 in transcriptional repression via binding to histone deacetylase HDAC1 and transcriptional repressor Sin3A, and evidenced MBD4 binding to hypermethylated promoters of *p16^{INKα}* and *MLH1* genes via meCpG sites.⁶⁰⁹

Meng and colleagues further demonstrated the implication of MBD4 in epigenetic silencing. By searching for MBD4 interactors that may contribute to genome stability, they identified UHRF1 (Ubiquitin-like with PHD and RING finger domains 1) and USP7 (ubiquitin specific peptidase 7) as specific interactor proteins of MBD4.⁵⁸⁸ UHRF1 is known to participate in heterochromatin formation and replication upon cell proliferation, and to localize to chromocenters, which cluster at heterochromatin regions.^{610,611} USP7 regulates UHRF1 via its deubiquitinase activity, and forms a complex with DNMT1 and UHRF1 in proliferating cells, relocating to nuclear foci and regulating the stability of DNMT1. MBD4 was found to specifically interact with USP7 and UHRF1 and to actively recruit USP7 to chromocenters/heterochromatin foci, further implicating MBD4 in the regulation of DNMT1 and suggesting that it may play a more global role in targeting specific methylation patterns and/or maintaining genome-wide methylation pattern at chromocenters.⁵⁸⁸

Taken together, these data strongly support a role for MBD4 in transcriptional repression via recognition of hypermethylated regions of the genome at meCpG sites, a critical epigenetic mark that is non-randomly methylated in about 70–80% of all CpG sites across the genome⁵⁷⁸ and that also enables the regulation of key processes such as genomic imprinting,

development and genomic stability. In this respect, MBD4 participates in the maintenance of genetic and epigenetic integrity of CpG sites by regulation of gene expression.⁵⁷⁵ The ability of MBD4 to excise T from deaminated 5-meC (leading to replacement of a new C and restoration of the 5-meC site) may also be a critical MBD4-dependent mechanism of maintenance of genomic methylation states.

DNA DEMETHYLATION?

Lastly, role for MBD4 in active DNA demethylation has been suggested, but remains widely controversial.^{575,576} In a Zebrafish model, 5-meC was shown to be removed via Mbd4 coupling with activated induced deaminase (AID, a deaminase converting 5-meC to T), which was promoted and enhanced by Gadd45 protein.⁶¹² They further found that overexpression of Mbd4 and AID/Apobec enzymes (same deaminase family) resulted in potent DNA demethylation in embryos, which only occurred when both the glycosylase and deaminase were co-expressed, suggesting a coupled mechanism of AID-driven 5-meC demethylation (to T) followed by T excision induced by MBD4. However, results failed to be reproduced in a subsequent study that stated that no effect from MBD4 and AID could be linked to methylation in Zebrafish embryo.⁶¹³ Another study claimed that MBD4 was implicated in hormone-induced active demethylation of *CYP27B1* promoter via PKC-mediated phosphorylation causing a shift in MBD4 substrate specificity, but the paper was subsequently retracted.^{614,615} Yet, Sabag and colleagues identified a demethylation pathway in mouse embryonic cells including Mbd4 along with Aid, Gadd45a and Tet1 (hydroxymethylation enzyme catalyzing 5meC to 5-hydroxy-meC, 5-hmC) and suggest its use in somatic cell reprogramming.⁶¹⁶ It should be noted however that there is still debate regarding whether AID/APOBEC enzymes can efficiently lead to deamination of 5-hmC, as its rate is expected to be slow, this activity is not observed in vitro,^{617,618} and deficiency of these enzymes do not lead to significant developmental effects as would be expected for major enzymes involved in active demethylation.⁶¹⁹ Furthermore, mice lacking functional MBD4 do not present with developmental or methylation defects.⁵⁷⁵

MBD4 DEFICIENCY IN CANCER

Considering its roles in suppression of mutation and DNA repair and the CpG>TpG hypermethylation pattern seen in mice upon *Mbd4* knockout,^{589,590} it was rapidly suggested that MBD4 may act as a tumor suppressor gene in humans. Besides, a potential role for MBD4 in MMR-deficient context was suggested by mice studies showing that although tumorigenesis could not be initiated by *Mbd4* knockout alone, they did result in increased tumor mutation burden (TMB) in cancer susceptible *Apc*^{min/+} mice, including mutations of cancer driver genes such as *Apc* itself, leading to accelerated gastrointestinal tumorigenesis.⁵⁹⁰ In humans, somatic deleterious mutations of *MBD4* mutations have essentially been described (in limited cohorts) in tumors associated with MSI (mostly through previously mentioned frameshift mutations in polyadenine stretches), mainly sporadic colorectal cancer (CRC) associated with MMR defects but also but also rare occurrences in MSI-exhibiting gastrointestinal, pancreatic and endometrial carcinomas,^{593-598,620,621} where monoallelic MBD4 mutations are proposed to cause dominant negative impairment of DNA repair,⁶²² often result in MBD4 lacking a functional glycosylase domain through truncating mutations, and may lead to tumor progression via increased transition mutation frequency.^{575,594,595,622}

In this respect, MBD4 deleterious mutations (mostly in the heterozygous form) were associated with tumors in a MMR-deficient context, but (i) inactivation of *Mbd4* in MMR-deficient mice does not alter spontaneous mutation frequency, tumor spectrum, tumor onset or MSI compared to single mutant *Mlh1* or *Msh2* mice,⁵⁹² suggesting that *MBD4* heterozygous mutations in MSI-exhibiting tumors may simply be induced by the MMR-deficient context without actively affecting tumorigenesis; and (ii) the question of whether MBD4 inactivation by itself could be implicated in tumorigenesis remained without definitive answer.

It was hypothesized a long time ago that germline *MBD4* mutations could play a role in cancer predisposition,⁵⁷³ however it was not until recently that rare occurrences in a few tumor types (outside MSI context) confirmed the protective role of MBD4 against cancer more generally. In 2018, our team reported a germline *MBD4* mutation in an outlier UM patient, exhibiting

bi-allelic inactivation in her tumor, and who initially responded to anti-PD1 treatment while UM patients typically do not respond to immunotherapy.⁵⁵³ This prompted us to further investigate a potential role for *MBD4* in UM predisposition.

MBD4 GERMLINE MUTATION IN AN OUTLIER UM PATIENT

In a study led by Rodrigues within our team, one outlier metastatic UM patient, UVM_IC, was identified among a series of 42 metastatic UM patients treated at the Curie Institute with pembrolizumab (PD1 immune checkpoint inhibitor), achieving exceptional tumor response.⁵⁵³ WES of her primary tumor and liver metastasis revealed a hypermutated profile (>250 SNVs, an almost 20-fold increase compared to an in-house series), where a vast majority (>90%) of SNVs occurred in a CpG>TpG mutational context, compared to <30% in other tumors) (Figure 19a&b). Because these mutations were most probably linked to 5meC deamination, a search for mutations within *MBD4* or *TDG* allowed to identify a germline deleterious *MBD4* frameshift deletion, associated with *MBD4* complete inactivation (loss of heterozygosity, LOH) in the tumor (Figure 19c) due to loss of the second allele by monosomy 3. Furthermore, mining of the UM (n=80) and pan-cancer TCGA tumor series exhibiting high SNV mutational load (>200 SNVs per tumor) and with high proportion of CpG>TpG mutational context (compared to all other C>T contexts) revealed two additional tumors with deleterious *MBD4* germline mutations, one other UM (UVM_1) and one glioblastoma (GBM_4) (Figure 19d). LOH was also observed in these patients, and UVM_1 also had tumor monosomy 3.⁵⁵³

These results demonstrated the potential implication of rare *MBD4* germline deleterious mutations in cancer predisposition, leading to a dramatic increase in tumor mutation burden. Furthermore, this case report provided evidence that *MBD4*-deficient tumors may be sensitive to immune checkpoint inhibitors, most often unsuccessful in UM tumors. Around the same time, Sanders and colleagues reported germline bi-allelic *MBD4* inactivation in 3 patients with acute myeloid leukemia (AML), characterized by a CpG>TpG methylation damage signature, driving a common path of clonal evolution notably characterized by pathogenic mutations in *DNMT3A* and other critical cancer genes.⁶²³ A few reports of germline *MBD4* mutations started emerging in other cancers as well, suggesting a larger tumor spectrum for *MBD4* cancer predisposition (detailed in Article 1). Taken together, these

findings prompted us to determine whether *MBD4* was associated with UM predisposition, by investigating the frequency of germline *MBD4* deleterious mutations in UM compared to their occurrence in the general population.

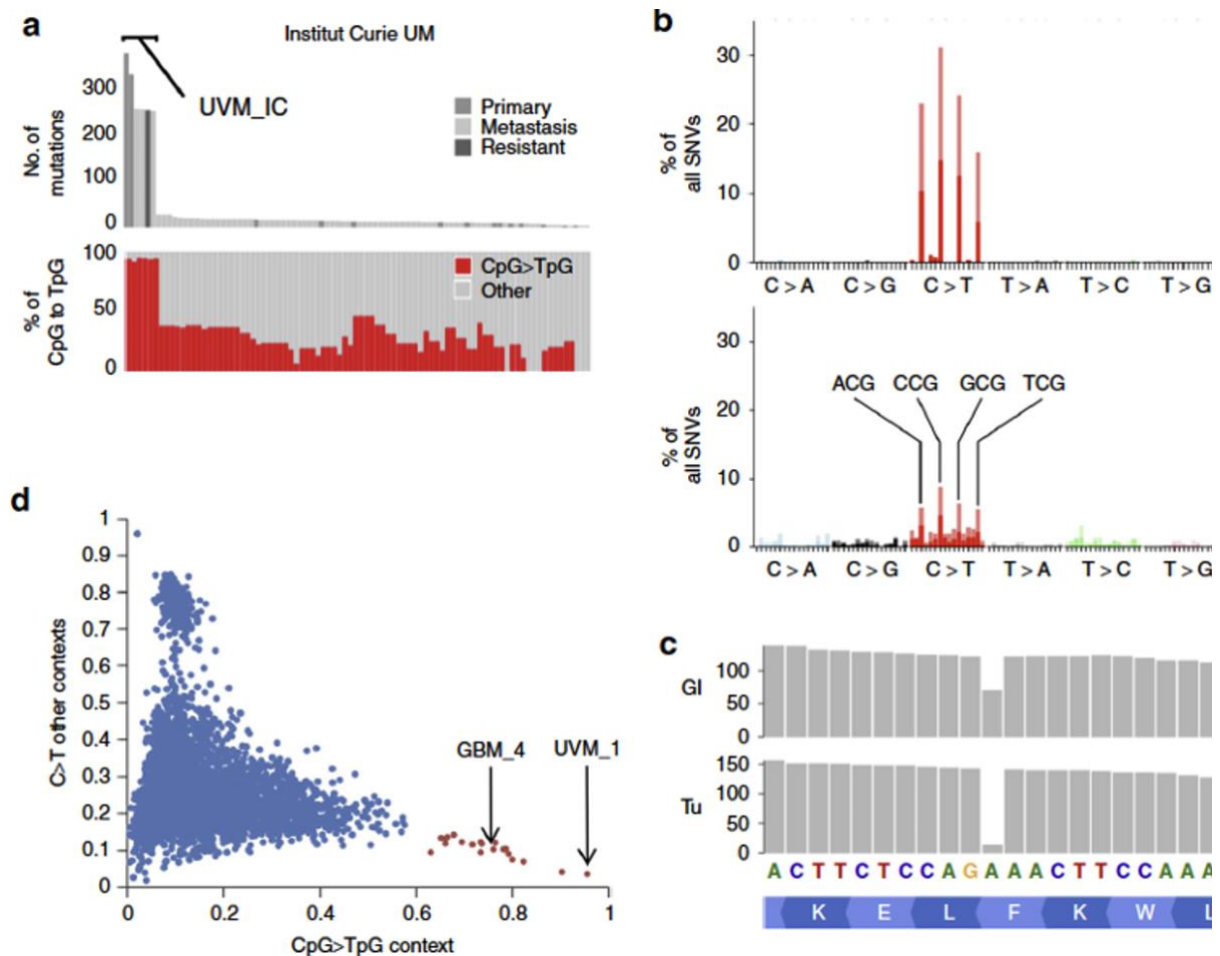


Figure 19. *MBD4* germline mutation in an outlier UM patient.

a) Tumor mutation load (top) and proportion of CpG>TpG mutations compared to all other types (bottom) in an outlier patient (UVM_IC) in a UM series of 23 patients including 14 primary and 71 metastatic tumor samples). **b)** Mutational signature of UVM_IC (top) and the rest of the series (below), where x-axis indicates the 96 possible tri-nucleotide substitutions (3' and 5' base context around the substitution) and y-axis indicates the relative proportion from all SNVs. **c)** *MBD4* mutation in UVM_IC shows loss of heterozygosity (LOH) in the tumor (Tu) compared to germline (Gl) DNA. **d)** Plot of proportions of C>T mutational contexts (CpG>TpG context on x-axis, all other contexts on y-axis) for TCGA tumors harboring more than 200 single nucleotide variants. The 20 tumors with highest CpG>TpG relative context appear in red. Two tumors harbored germline *MBD4* mutations: one UM patient (UVM_1) and one glioblastoma patient (GBM_4). Figure from Rodrigues *et al.*⁵⁵³

ARTICLES

INTRODUCTION TO ARTICLE 1

GERMLINE *MBD4* MUTATIONS AND PREDISPOSITION TO UVEAL MELANOMA

Although the initial aim of this PhD project was to unravel low to moderate UM genetic risk variants UM by genome-wide association study (GWAS), the case report⁵⁵³ from the study led by Rodrigues within our team of a metastatic UM patient presenting dramatic immune responses against her metastases upon anti-PD1 treatment and harboring a germline *MBD4* deleterious mutation, gave us an unprecedented opportunity to further unravel genetic predisposition to UM using a candidate-gene approach, concomitantly with the GWAS approach. The results described in the case report were encouraging as the outlier patient initially demonstrated a dramatic response to immunotherapy, most likely due to the hypermutated CpG>TpG phenotype of her tumors resulting from *MBD4* inactivation, highlighting the importance of recognizing patients with *MBD4* germline mutations. Furthermore, this study published in 2018 (describing two UM patients and one glioma with *MBD4* germline mutations, tumor bi-allelic inactivation and associated CpG>TpG hypermutated phenotype) was the first of a series of case reports of germline *MBD4* mutations associated with loss-of-heterozygosity in the tumor and dramatic increase in the tumor mutation burden, which started emerging in different cancers such as in another UM patient⁶²⁴ and in three cases of acute myeloid leukemia (AML),⁶²³ suggesting that *MBD4* mutations may represent a cancer predisposition condition for multiple types of tumors.

This context prompted us to investigate whether *MBD4* is a UM predisposition gene, as no epidemiologic study investigating its germline mutational frequency had been performed. As we were in the process of accruing a large consecutive cohort of UM patients (>1,000 cases) treated at the Curie Institute who had given their consent for studies on their germline DNA in the context of a large GWAS study in UM, we performed *MBD4* targeted sequencing on germline DNA from UM patients of this cohort to detect potential other deleterious mutations

and determine the mutation frequency and the relative risk conferred by *MBD4* in this consecutive series.

In this first article, published in the *Journal of the National Cancer Institute*, we mine UM public databases for additional *MBD4* germline mutations, report occurrences in other tumors, investigate *MBD4* predisposition in UM in a large consecutive series of 1,093 patients and in a tumor series of 192 UM samples with monosomy 3, and functionally characterize the tumors and phenotype of patients harboring deleterious *MBD4* mutations.

ARTICLE 1: GERMLINE MBD4 MUTATIONS AND PREDISPOSITION TO UVEAL MELANOMA

Abstract

Background

Uveal melanoma (UM) arises from malignant transformation of melanocytes in the uveal tract of the eye. This rare tumor has a poor outcome with frequent chemo-resistant liver metastases. *BAP1* is the only known predisposing gene for UM. UMs are generally characterized by low tumor mutation burden, but some UMs display a high level of CpG>TpG mutations associated with *MBD4* inactivation. Here, we explored the incidence of germline *MBD4* variants in a consecutive series of 1093 primary UM case patients and a series of 192 UM tumors with monosomy 3 (M3).

Methods

We performed *MBD4* targeted sequencing on pooled germline (n = 1093) and tumor (n = 192) DNA samples of UM patients. *MBD4* variants (n = 28) were validated by Sanger sequencing. We performed whole-exome sequencing on available tumor samples harboring *MBD4* variants (n = 9). Variants of unknown pathogenicity were further functionally assessed.

Results

We identified 8 deleterious *MBD4* mutations in the consecutive UM series, a 9.15-fold (95% confidence interval = 4.24-fold to 19.73-fold) increased incidence compared with the general population (Fisher exact test, $P = 2.00 \times 10^{-5}$, 2-sided), and 4 additional deleterious *MBD4* mutations in the M3 cohort, including 3 germline and 1 somatic mutations. Tumors carrying deleterious *MBD4* mutations were all associated with high tumor mutation burden and a CpG>TpG hypermutator phenotype.

Conclusions

We demonstrate that *MBD4* is a new predisposing gene for UM associated with hypermutated M3 tumors. The tumor spectrum of this predisposing condition will likely expand with the addition of *MBD4* to diagnostic panels. Tumors arising in such a context should be recognized because they may respond to immunotherapy.

Germline MBD4 Mutations and Predisposition to Uveal Melanoma

Anne-Céline Derrien, MSc ^{1,†} Manuel Rodrigues, MD, PhD,^{1,2, †} Alexandre Eeckhoutte, BSc ¹
Stéphane Dayot, BSc ¹ Alexandre Houy, BSc ¹ Lenha Mobuchon, PhD ¹ Sophie Gardrat, MD,^{1,3}
Delphine Lequin, MD,³ Stelly Ballet, MSc,³ Gaëlle Pierron, PhD ³ Samar Alsafadi, PharmD, ^{1,4}
Odette Mariani, PhD ⁵ Ahmed El-Marjou, PhD ⁶ Alexandre Matet, MD, PhD ^{7,8}
Chrystelle Colas, MD, PhD,⁹ Nathalie Cassoux, MD, PhD,^{7,8} Marc-Henri Stern, MD, PhD ^{1,9,*}

¹Inserm U830, DNA Repair and Uveal Melanoma (D.R.U.M.), Equipe Labellisée Par la Ligue Nationale Contre le Cancer, Paris, France; ²Department of Medical Oncology, Institut Curie, PSL Research University, Paris, France; ³Department of Biopathology, Institut Curie, PSL Research University, Paris, France; ⁴Translational Research Department, Institut Curie, PSL Research University, Paris, France; ⁵Biological Resource Center, Institut Curie, PSL Research University, Paris, France; ⁶Institut Curie, PSL Research University, UMR144, Recombinant Protein Facility, Paris, France; ⁷Department of Ocular Oncology, Institut Curie, Paris, France; ⁸Faculty of Medicine, University of Paris Descartes, Paris, France and ⁹Department of Genetics, Institut Curie, Paris, France

*Correspondence to: Marc-Henri Stern, MD, PhD, 26 rue d'Ulm, 75248 Paris cedex 05, France (e-mail: marc-henri.stern@curie.fr).

†Authors contributed equally to this work.

Abstract

Background: Uveal melanoma (UM) arises from malignant transformation of melanocytes in the uveal tract of the eye. This rare tumor has a poor outcome with frequent chemo-resistant liver metastases. BAP1 is the only known predisposing gene for UM. UMs are generally characterized by low tumor mutation burden, but some UMs display a high level of CpG>TpG mutations associated with MBD4 inactivation. Here, we explored the incidence of germline MBD4 variants in a consecutive series of 1093 primary UM case patients and a series of 192 UM tumors with monosomy 3 (M3). **Methods:** We performed MBD4 targeted sequencing on pooled germline (n = 1093) and tumor (n = 192) DNA samples of UM patients. MBD4 variants (n = 28) were validated by Sanger sequencing. We performed whole-exome sequencing on available tumor samples harboring MBD4 variants (n = 9). Variants of unknown pathogenicity were further functionally assessed. **Results:** We identified 8 deleterious MBD4 mutations in the consecutive UM series, a 9.15-fold (95% confidence interval = 4.24-fold to 19.73-fold) increased incidence compared with the general population (Fisher exact test, $P = 2.00 \times 10^{-5}$, 2-sided), and 4 additional deleterious MBD4 mutations in the M3 cohort, including 3 germline and 1 somatic mutations. Tumors carrying deleterious MBD4 mutations were all associated with high tumor mutation burden and a CpG>TpG hypermutator phenotype. **Conclusions:** We demonstrate that MBD4 is a new predisposing gene for UM associated with hypermutated M3 tumors. The tumor spectrum of this predisposing condition will likely expand with the addition of MBD4 to diagnostic panels. Tumors arising in such a context should be recognized because they may respond to immunotherapy.

Uveal melanoma (UM) is the most frequent primary intraocular tumor in adults with an overall mean incidence of 5.2 per million per year in the United States (1). Metastases arise in more than 30% of case patients, almost invariably in the liver, with a dismal prognosis because of the absence of effective treatment (median survival of 10 months) (2,3). UMs with high risk of developing metastases are characterized by loss of chromosome 3 and by BAP1 (encoded in 3p21) inactivation resulting from loss-of-function (LoF) mutations and loss of the remaining wild-type copy on chromosome 3 (4). Rare familial UMs are associated with germline mutations of BAP1 (Mendelian Inheritance in Man [MIM]: 614327) (5,6), which is the only known highly

penetrant UM predisposition gene. UM mainly affects individuals of European ancestry and is associated with fair skin and light iris color. However, the low tumor mutation burden (TMB) and lack of an ultraviolet-associated mutational signature argue against a role for ultraviolet radiation in UM oncogenesis (7).

Recently, the characterization of a metastatic UM patient with an exceptional response to anti-Programmed cell death protein 1 (anti-PD-1) therapy led us to identify a CpG>TpG mutator phenotype linked to germline protein truncating variants (PTV) in MBD4 (Methyl-CpG Binding Domain Protein 4) and somatic loss of the wild-type allele in tumors in 2 patients with UM and 1 with glioma (8). Another UM patient responding to

Received: December 20, 2019; Revised: March 19, 2020; Accepted: March 26, 2020

© The Author(s) 2020. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

immune checkpoint inhibitors was subsequently reported with a germline *MBD4* PTV (9). *MBD4* encodes a glycosylase involved in the base excision repair of DNA damage arising from spontaneous deamination of 5-methylcytosine to thymine (10,11), which is consistent with the mCpG>TpG transitions [mutational signature SBS1 (12)] observed in *MBD4*-inactivated tumors (12). *MBD4*, located on chromosome 3, is thought to act as a tumor suppressor gene, following Knudson's 2-hit model with loss of the wild-type allele by monosomy 3 (M3) in UMs (8). Altogether, these germline deleterious *MBD4* variants in UM prompted us to investigate the role of *MBD4* as a predisposing gene for UM. Here, we performed *MBD4* targeted-sequencing in germline DNA of a large consecutive series of 1093 UM patients and in tumor DNA of a second cohort of 192 UM patients with M3 and investigated the TMB and mutational signature in patients harboring *MBD4* mutations.

Methods

Study Patients

The 1099 individuals with UM were diagnosed at Institut Curie, France, from 2013 to 2018. The sex proportion of female to male was 52.2 to $47.8 \pm 3.0\%$ (95% confidence interval [CI]) and the median age at diagnosis was 64 years old (Q1-Q3 quartile interval = 54-73). All patients provided written informed consent to perform germline genetic analyses and somatic genetic analyses of tumor samples. Six patients were subsequently removed from the study: the UM diagnosis was not confirmed for 5 patients, and the sixth patient had undergone a bone marrow transplantation and his blood sample corresponded to his donor's (Supplementary Figure 1A, available online). The study was conducted in accordance with the declaration of Helsinki and was approved by the ethical committee and institutional review board of the Institut Curie. Germline DNA was extracted from the blood of all patients (DNeasy Blood and Tissue kit, Qiagen). When available, tumor genomic status as part of the prognostication assessment was extracted from the medical record and classified as M3, including isodisomy 3 or disomy 3 (D3). Tumor samples were collected from primary eye tumors. A second series of 192 UM tumor samples with M3 was also accrued at Institut Curie, of which 120 patients were independent from the consecutive germline cohort.

MBD4 Targeted Sequencing

Germline DNA of 1099 UM patients from the UM consecutive series (before removal of the 6 aforementioned patients) and DNA of a series of 192 M3 UM tumors were screened for *MBD4* variants by pooled *MBD4* targeted sequencing. Details on the sequencing strategy and bioinformatics pipeline are described in the Supplementary Methods (available online). Deconvolution of the identified pooled DNA samples with an *MBD4* variant was carried out by Sanger sequencing.

Identified *MBD4* variants (Supplementary Table 1, available online) were defined following the recommendations of the Human Genome Variation Society (<http://varnomen.hgvs.org/>) and numbered based on the *MBD4* (MIM: 603574) cDNA and protein sequences (GenBank accession numbers NM_003925.2 and NP_003916.1, respectively).

Glycosylase Activity Assay

Wild-type and mutant *MBD4* were expressed to assess their enzymatic activity by in vitro *MBD4* glycosylase assay as previously described (13,14) (Supplementary Methods, available online).

Whole-Exome Sequencing (WES) and Mutation Calling

WES was performed on tumor samples from *MBD4* variant carriers who consented to germline studies (Supplementary Table 2, available online). Variant calling and TMB analysis are described in Supplementary Methods (available online).

Statistical Analysis

Statistical analysis was performed using R software v3.6. Fisher exact test was used to calculate P values between our cohort and the general population. Random subsampling 1 000 000 times of 2186 alleles from the GnomAD cohort was also used as statistical "matching" strategy to calculate P values between our cohort and the general population. A Mann-Whitney U test was used to compare the median $\pm$ median absolute deviation for the mutation burden and CpG>TpG proportion between *MBD4*-deficient (*MBD4*_{def}) and *MBD4*-proficient patients. To compare age of UM onset, a Wilcoxon Rank-Sum test was used. For survival analysis using Kaplan-Meier curves, statistical analysis was carried out using the log-rank test. All statistical tests were 2-sided, except for 1-sided Wilcoxon Rank-Sum test for early age of onset in *MBD4*_{def} patients. Confidence intervals were carried out at 95% confidence level. Confidence intervals for relative risk (RR) measurement was calculated as previously described (15). A P value less than .05 was considered to be statistically significant.

Results

Mining *MBD4* Germline Variants in Public UM Cohorts

To evaluate the potential predisposing role of *MBD4* in UM, we mined all available public UM cohorts for germline *MBD4* variants. We identified 1 case harboring the germline deleterious *MBD4* PTV c.1443delT (p.Leu482Trpfs*9) in a first cohort containing 37 UM patients (phs001421.v1.p1) (16). A second cohort of 98 UM patients (phs000823.v1.p1) included a second case with an *MBD4* c.1020delA (p.Asp341Thrfs*13) PTV. Collectively, 5 *MBD4* germline deleterious variants were found in 268 analyzed UM patients (1.9%) (8,9,16–18). In contrast, such variants are exceedingly rare in an unselected population (88 of approximately 125 000 individuals in GnomAD v2.1.1). Out of these 5 UM case patients with germline *MBD4* variants, 4 had available tumor profiles that all showed M3 and somatic *BAP1* inactivation (8,9,16), presumably because of the localization of both *MBD4* and *BAP1* on chromosome 3.

Identification of *MBD4* Germline Variants in the In-House Consecutive UM Series

To assess the actual prevalence of *MBD4* germline deleterious variants in UM, we next explored an in-house cohort of 1093 (approximately one-half the annual incidence in United States) consecutive patients diagnosed with UM at Institut Curie between 2013 and 2018. Targeted next-generation sequencing in

pooled patient DNA followed by Sanger sequencing revealed germline MBD4 PTVs in 7 patients (Table 1): 2 splice site variants (c.1562-1G>T [p.Asp521Profs*4] in 2 patients and a c.335+1G>A [p.Arg83Profs*5] variant), 2 frameshift deletion variants (c.1443delT [p.Leu482Trpfs*9] and c.1384delG [p.Ala462Leufs*29]), and 1 stop-gain near the end of the last exon of MBD4 (c.1706G>A [p.Trp569*] in 2 patients).

We also identified and characterized 9 rare germline MBD4 variants (frequency <1% in the general population): 7 missense variants in 13 patients and 2 intronic variants in 2 patients (Figure 1A). Out of these, 3 were predicted with possible splicing consequences: 2 missense variants (c.1652A>G [p.Asn551Ser] and c.1400A>G [p.Asn467Ser]), and 1 intronic variant (c.1277-18T>A) (Supplementary Table 1, available online).

Functional Assessment of MBD4 Variants

Exon-trapping assays performed on the 3 aforementioned variants demonstrated the use of an alternative acceptor site with c.1277-18T>A, albeit to a lesser extent than the canonical splice site, whereas c.1652A>G and c.1400A>G did not show any measurable effect (Supplementary Figure 2, available online).

All missense variants within the MBD4 glycosylase domain [aa425-580] were assessed by in vitro glycosylase assay together with p. Trp569* because of its localization near the end of the protein (Figure 1; Table 1). The assay confirmed that p. Trp569* results in a catalytically inactive protein and further demonstrated p. Arg468Trp to be a LoF variant (Figure 1). Consistent with this finding, the key role of Arg468 was previously established in binding at the G/T mismatch site, maintaining the T base in a position required for catalysis, and interacting with the orphan G base through hydrogen bonding (22).

We next characterized by WES the 3 available tumors from patients carrying germline MBD4 LoF variants (UM75: p. Trp569*, UM605: p. Ala462Leufs*29, and UM656: p. Leu482Trpfs*9) (8). We confirmed that the 3 MBD4 germline LoF mutations were associated with somatic loss of the wild-type allele in the tumors by either M3 (UM75 and UM605) or isodisomy 3 (UM656) (Figure 1C; Supplementary Table 1, available online), consistent with the 3q21.3 location of MBD4.

MBD4 Mutations in the M3 Tumor Series and Tumor Signature

To further evaluate the incidence of MBD4 alterations in UM, and assuming that MBD4_{def} UMs are associated with M3, we accrued a series of 192 UM tumor samples with M3 (of which 120 case patients were independent from the present consecutive UM series) and screened for MBD4 mutations using the aforementioned strategy (Supplementary Figure 1B, available online). We identified 6 additional MBD4 variants, including 4 LoF mutations (UMT62: c.1688T>A [p.Leu563*], UMT45: c.1562-1G>T [p.Asp521Profs*4], UMT61: c.1002delTTTG [p.Lys335Phefs*18], UMT162: c.541C>T [p.Arg181*]) and 2 missense variants (UMT88: c.1402C>T [p.Arg468Trp] and UMT105: c.1073T>C [p.Ile358Thr]) (Table 1; Figure 1; Supplementary Table 1, available online). Of these, 4 patients (UMT45, UMT61, UMT162, and UMT88) had available germline DNA and all consented to germline studies. Characterization by WES of their tumor samples showed that the 3 tested LoF mutations were germline variants, the missense p. Arg468Trp was somatic, and all were associated with Loss Of Heterozygosity (LOH) of the wild-type allele in tumors (Figure 1, A and B; Supplementary Table 1, available online).

MBD4 inactivation has been associated with a high TMB and a CpG>TpG mutational pattern (8). We confirmed the high TMB in all 3 available MBD4_{def} UMs from the germline consecutive cohort with 275, 122, and 181 variants per exome in UM75, UM605, and UM656, respectively, compared with 16 ± 4.0 (median $\pm$ median absolute deviation) variants in MBD4-proficient UMs (18) (Figure 1C; Supplementary Table 2, available online). CpG>TpG transitions represented 96.4%, 85.7%, and 92.8% of all single nucleotide variants (SNVs), respectively, compared with $24.3 \pm 7.6\%$ in MBD4-proficient UMs (18) (Figure 1, C and D; Supplementary Table 2, available online). In line with the glycosylase assay, TMB results and somatic chromosome 3 LOH further confirmed the deleterious effect of p. Trp569* in UM75 (Figure 1). Similarly, within the M3 UM tumor series, the 3 available UMs carrying a germline LoF MBD4 variant exhibited a high TMB (269, 288, and 86 variants per exome in UMT162, UMT45, and UMT61, respectively) and a predominance of CpG>TpG transitions (85.6%, 94.4%, and 63.9%, respectively) among all SNVs (Figure 1, C and D; Supplementary Table 2, available online). The tumor sample of patient UMT88, carrying a somatic p. Arg468Trp variant identical to that found as a germline variant in UM293, also carried a high TMB (243 variants) and the CpG>TpG mutational pattern (92.5%) (Figure 1, C and D; Supplementary Table 2, available online), thereby confirming the deleterious effect of this missense variant previously demonstrated in the glycosylase assay (Figure 1). Taken together, these 7 patients with MBD4 deleterious mutations had a 15-fold increase in number of variants per exome (MBD4_{def}: 243 ± 66.7 variants vs MBD4_{pro}: 16 ± 4.0 , Mann-Whitney $P = 8.72 \times 10^{-5}$) and a statistically significantly higher CpG>TpG median proportion among SNVs (MBD4_{def}: $92.5 \pm 5.7\%$ vs MBD4_{pro}: $24.3 \pm 7.5\%$, $P = 9.82 \times 10^{-7}$) (18).

In addition, we characterized the available tumor samples from 2 patients harboring missense variants that were not predicted to be deleterious (UM102: c.139G>A [p.Gly47Arg] and UM350: c.1652A>G [p.Asn551Ser]; Figure 1; Supplementary Table 1; Supplementary Figure 2, available online). The low TMB (33 and 40 variants per exome, respectively) and absence of CpG>TpG signature confirmed their neutral effect (Supplementary Table 2; Supplementary Figure 3, available online).

Incidence of Germline MBD4 Deleterious Mutations in UM

Taken together, we thus identified 8 LoF germline variants in MBD4 among the 1093 consecutive UM case patients, including p. Arg468Trp with deleterious effect on MBD4 glycosylase activity (Table 1). These account for a statistically significant 9.15-fold increase in deleterious variant frequency compared with the general population, even when restricting ourselves to truncating and splicing MBD4 LoFs as defined by GnomAD (7 LoFs of 2186 observed alleles in UM, representing a variant allele frequency [VAF] of 0.0032 vs 88 LoFs out of a median of 251 450 alleles in the GnomAD v2.1 general population; VAF = 3.50×10^{-4} ; Fisher exact test $P = 2.00 \times 10^{-5}$). To circumvent the imbalanced dataset, a “matching” subsampling approach was used, giving a similar P value (1.60×10^{-5}). Therefore, we demonstrate that the prevalence of MBD4 germline deleterious variants in UM is approximately 0.7%, close to that of BAP1 germline mutation in UM (1.6%) (23), and that MBD4 mutations strongly predispose to UM with an RR of 9.15 (95% CI = 4.24 to 19.73). A comparison between MBD4 germline LoF frequency in

Table 1. MBD4 germline deleterious variants in UM and in other malignancies^a

Patient series	Patient	Variant	dbSNP	Mutation type	Glycosylase assay	GnomAD allele frequency (NFE ^b)		
						Allele count	Obs. allele number	Frequency
UM ^e germline consecutive series	UM75	p.Trp569*	rs939751619 ^c	stop_gain	Inactive ^d	2	129 130	1.55 × 10 ⁻⁵
	UM1033							
	UM49	p.Asp521Profs*4	rs778697654 ^c	splice_acceptor	ND	5	113 766	4.39 × 10 ⁻⁵
	UM1088							
	UM656	p.Leu482Trpfs*9	rs769076971 ^c	frameshift_deletion	ND	3	113 752	2.64 × 10 ⁻⁵
	UM293	p.Arg468Trp	rs1380952147	nonsynonymous_SNV	Inactive ^d	0	113 630	0.00
UM M3 tumor series	UM605	p.Ala462Leufs*29	—	frameshift_deletion	ND	—	—	—
	UM436	p.Arg83Profs*5	rs552296498 ^c	splice_donor	ND	3	129 158	2.32 × 10 ⁻⁵
	UMT45	p.Asp521Profs*4	rs778697654 ^c	splice_acceptor	ND	5	113 766	4.39 × 10 ⁻⁵
	UMT61	p.Lys335Phefs*18	rs1443006605	frameshift_deletion	ND	0	113 650	0,00
	UMT162	p.Arg181*	rs1270271346	stop_gain	ND	2	128 972	1.55 × 10 ⁻⁵
	UM (9)	p.Leu563*	rs200758755	stop_gain	ND	8	113 702	7.04 × 10 ⁻⁵
UM (public data)	TCGA_UVM_1 (8)	p.Asp521Profs*4	rs778697654 ^c	splice_acceptor	ND	5	113 766	4.39 × 10 ⁻⁵
	UM _{phs001421.v1.p1} (16)	p.Leu482Trpfs*9	rs769076971 ^c	frameshift_deletion	ND	5	113 752	4.40 × 10 ⁻⁵
	UVM_IC (8)							
	UM _{phs000823.v1.p1}	p.Asp341Thrfs*13	—	frameshift_deletion	ND	—	—	—
	AML _{EMC} -AML-1 (13)	p.His567del	rs775848563	inframe_deletion	ND	—	—	—
	AML _{WEHI} -AML-1/2 (13)	p.Asp521Profs*4	rs778697654 ^c	splice_acceptor	ND	5	113 766	4.39 × 10 ⁻⁵
Other malignancies	AML _{WEHI} -AML-1/2 (13)	p.Glu314Argfs*13	rs558765093 ^c	frameshift_insertion	ND	—	—	—
	Spiradenocarcinoma (19)							
	TCGA_GBM_4 (8)	p.Arg83Profs*5	rs552296498 ^c	splice_donor	ND	3	129 158	2.32 × 10 ⁻⁵
	Colorectal polyposis (20)	p.Gln73*	rs148098584	stop_gain	ND	0	113 750	0.00
	Pilocytic astrocytoma (21)	NA ^g	NA	NA	ND	NA	NA	NA
	Gastric adenocarcinoma (21)	NA	NA	NA	ND	NA	NA	NA
	Pancreatic adenoK (21)	NA	NA	NA	ND	NA	NA	NA
	Pancreatic endocrine tumor (21)	NA	NA	NA	ND	NA	NA	NA

^aadenoK = adenocarcinoma; AML = acute myeloid leukemia; GBM = glioblastoma; M3 = monosomy 3; NA = not available; ND = not determined; NFE = non-Finnish European; UM or UVM = uveal melanoma; — = no value given because of the absence of the variant in dbSNP and/or in the GnomAD NFE population.

^bNFE population of the Genome Aggregation Database (GnomAD v2.1.1).

^cVariant found in more than 1 nonrelated patient.

^dInactive: absence of glycosylase activity of the recombinant protein carrying the variant.

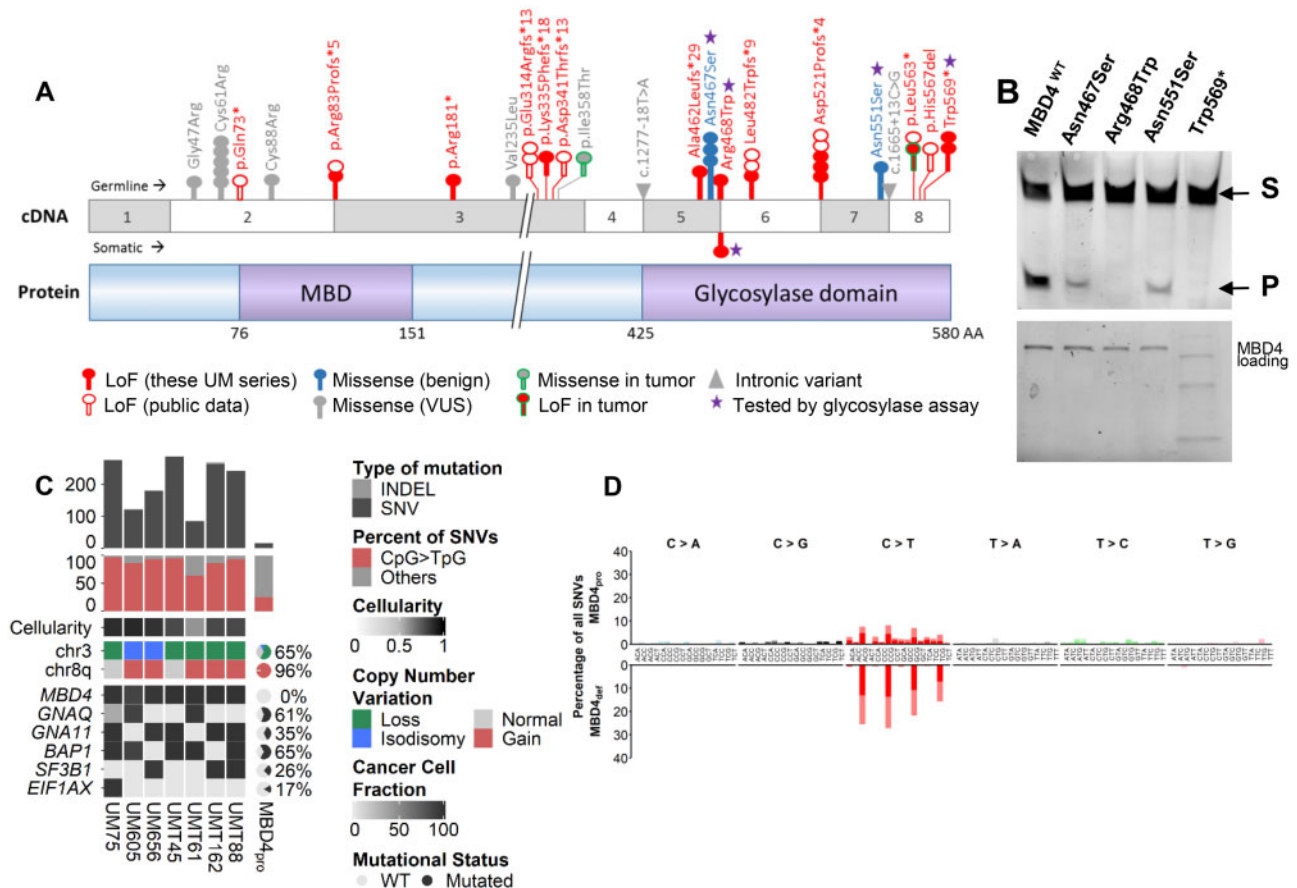


Figure 1. Functional consequences and phenotype associated with germline and somatic MBD4 deleterious variants. **A)** Schematic representation of MBD4 cDNA (top) and protein (bottom) sequences. Functional methyl-binding domain (MBD) and glycosylase domain are indicated. The position of all MBD4 variants identified in the 2 uveal melanoma (UM) series (consecutive germline UM series and tumor monosomy 3 [M3] series) is highlighted, with germline and somatic variants above and below the cDNA sequence, respectively, and the 2 variants from the tumor M3 cohort with unknown somatic or germline origin circled in green. These MBD4 variants include loss-of-function (LoF, in red), missense (either benign, in blue-filled circles, or of unknown biological significance [VUS] in gray-filled circles) and intronic (gray triangles) variants. Each circle represents 1 patient harboring the variant. Other MBD4 germline deleterious variants mined on public data are also shown (empty red circles). **B) Top:** Glycosylase activity assay of recombinant wild-type MBD4 (MBD4^{WT}) and mutant proteins resulting from missense variants and 1 stop gain variant (purple star in 1A) residing in the MBD4 glycosylase domain. Substrate = S; cleaved product = P. **Bottom:** loading blot for MBD4 wild-type and mutant recombinant proteins corresponding to the glycosylase assay. **C)** Tumor characteristics of MBD4-deficient (MBD4_{def}) patients compared with that of MBD4-proficient UM patients (MBD4_{pro}) (18). MBD4_{def} patients include UM75, UM605, and UM656 from the consecutive germline series and UMT45, UMT61, UMT162, and UMT88 from the M3 UM tumor series. All patients harbor germline MBD4 variants, except for UMT88 with a somatic MBD4 variant. **Top:** tumor mutation burden estimated by number of variants (single nucleotide variants [SNVs] in dark gray, and insertions-deletions [INDELs] in light gray) in the exome; **middle:** proportion of CpG>TpG transitions (red) relative to all SNVs (gray); **bottom:** copy number alterations in chromosomes 3 and 8q, and mutational status of MBD4, GNAQ, GNA11, BAP1, SF3B1, and EIF1AX, represented as percentage for the MBD4_{pro} series (18). The clonality or subclonality of these key mutational events is indicated by their cancer cell fraction in black-gray gradation, taking into account the variant allele frequency (VAF), copy number change, and cellularity. A plot of the VAF distribution of all variants in the 7 exomes is available in Supplementary Figure 4 (available online). For each exome in the MBD4_{def} group, tumor cellularity is indicated by black-gray shading (and quantified in Supplementary Table 2, available online). **D)** Mutational patterns of the MBD4_{def} (top) and MBD4_{pro} (bottom) groups based on the relative proportion (y-axis) of each of the 96 types of trinucleotide substitution (x-axis). Dark or bright colors correspond to sense or antisense strands. Individual mutational pattern for all tumor exomes assessed are available in Supplementary Figure 3 (available online).

this UM consecutive series and in different subsets of the GnomAD population (in the general and European populations) is further presented in Table 2.

Within the UM tumor cohort with M3, we identified a total of 5 MBD4 LoF variants, including at least 3 of germline origin and 1 somatic, out of 192 UM patients. These 3 germline LoF variants by themselves account for a VAF of 0.016, more than twice that found in the germline consecutive UM cohort. This was expected given the recurrence of approximately 50% of chromosome 3 loss event among all UM patients (24). This finding confirms that MBD4 deficiency in UM is mainly associated with M3 and that MBD4 germline mutations specifically predispose to hypermutated high-risk M3 UMs.

Defining the MBD4 Predisposition Syndrome

To further characterize this new cancer predisposition, we investigated the medical records of MBD4 mutation carriers. In contrast with the high RR of 9.15 (and therefore an approximately 9-fold higher risk of developing a UM) conferred by MBD4 LoF germline mutations, none of these individuals had familial or bilateral UM. With a lifetime risk of UM estimated at 7.69×10^{-5} in the general population (25), an RR of 9 would result in a lifetime risk of UM of 6.92×10^{-4} . Such incidence is still too low to observe familial aggregation, which is consistent with our finding. Assuming that all MBD4_{def} UMs are associated with M3, we compared their medical records with patients from this cohort with available tumor

Table 2. Frequency of *MBD4* germline deleterious variants in the UM series compared with various populations of the GnomAD database^a

Study population		No. of LoF variants	Allele count ^c	Frequency	RR ^d (95% CI ^e)	Fisher test (P value)
UM consecutive series		7 ^f	2186	0.00320	—	—
GnomAD v2.1.1	NFE ^b	47	113 736	0.00041	7.75 (3.51 to 17.12)	6.86 × 10 ^{−5}
	Population ^b	88	251 450	0.00035	9.15 (4.24 to 19.73)	2.00 × 10 ^{−5}
GnomAD v2.1.1 (controls only)	NFE	13	42 768	0.00030	10.53 (4.20 to 26.38)	2.82 × 10 ^{−5}
	General population	33	109 404	0.00030	10.62 (4.70 to 23.97)	1.16 × 10 ^{−5}
GnomAD v2.1.1 (noncancer only)	NFE	41	102 730	0.00040	8.02 (3.60 to 17.86)	5.89 × 10 ^{−5}
	General population	82	236 912	0.00035	9.25 (4.28 to 19.99)	1.90 × 10 ^{−5}
GnomAD v3	NFE	20	64 571	0.00031	10.34 (4.38 to 24.42)	2.00 × 10 ^{−5}
	General population	39	143 286	0.00027	11.76 (5.27 to 26.27)	5.50 × 10 ^{−5}

^aCI = confidence interval; LoF = loss-of-function (deleterious) variants; NFE = non-Finnish European; RR = relative risk; UM = uveal melanoma; — = no value given here because the relative risk, confidence interval, and statistic tests are presented between the UM consecutive series and each GnomAD subpopulation in the rows below.

^bNFE population subset of the Genome Aggregation Database (GnomAD v2.1.1).

^cFor all GnomAD populations described, refers to the median number of allele count.

^dRR here is calculated by dividing the LoF frequency in the UM consecutive series by the LoF frequency in the corresponding GnomAD population subset.

^eConfidence interval of the relative risk is calculated as previously described (15).

^fSeven LoF variants correspond to the 8 deleterious *MBD4* variants identified in this study, with removal of the missense deleterious variant p. Arg468Trp so as to restrict the analysis to LoF variants as defined by GnomAD for accurate comparison.

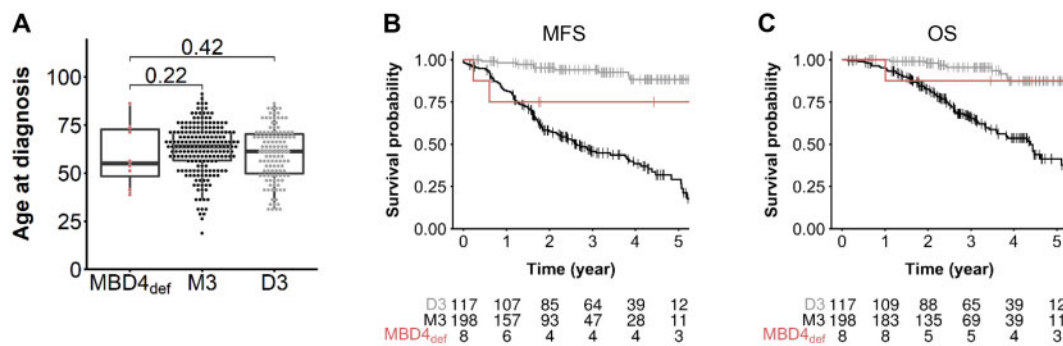


Figure 2. Uveal melanoma (UM) clinical characteristics in an *MBD4*-deficient (*MBD4*_{def}) context. **A)** Age of UM onset of *MBD4*_{def} patients (n = 8) in the germline consecutive UM series compared with disomy 3 (D3, n = 117) and monosomy 3 (M3, n = 198) *MBD4*-proficient (*MBD4*_{pro}) UMs. Wilcoxon test, 1-sided (testing early UM onset in *MBD4*_{def} patients): *MBD4*_{def} vs M3: P = .22, *MBD4*_{def} vs D3: P = .42; no age difference found between D3 and M3 groups, Wilcoxon test, 2-sided P = .087; – not shown). **B and C)** Metastasis-free survival (MFS, B) and overall survival (OS, C) of *MBD4*_{def} UM patients (n = 8) and *MBD4*_{pro} UM patients with M3 or D3. Time zero refers to time at primary UM diagnosis. MFS was defined as the interval between the date of primary UM diagnosis and the date of distant metastasis (first imaging) or death from any cause. The number of patients in each group at each time point (year) is indicated. Survival distributions were estimated by the Kaplan-Meier method and compared using the log-rank test: log-rank test, 2-sided, M3 vs D3: P = 1.98×10^{-9} (OS), P = 1.11×10^{-16} (MFS); M3 vs *MBD4*_{def}: P = .11 (OS), P = .06 (MFS); D3 vs *MBD4*_{def}: P = .62 (OS), P = .10 (MFS).

genomic status, that is, 198 *MBD4* wild-type M3 and 117 *MBD4* wild-type D3 UMs. Surprisingly, no early-onset UM was found in *MBD4* carriers compared with noncarriers regardless of their chromosome 3 status (*MBD4*_{def}: median age and Q1-Q3 quartile interval = 55.5, 95% CI = 48.4 to 72.8, N = 8; D3: 61.3, 95% CI = 49.8 to 70.3, N = 117; M3: 63.4, 95% CI = 56.6 to 71.4, N = 198; Wilcoxon test, 1-sided: *MBD4*_{def} vs M3: P = .22, *MBD4*_{def} vs D3: P = .42; no age difference found between D3 and M3 groups, Wilcoxon test, 2-sided P = .087; Figure 2A). Although the size of the *MBD4*_{def} series prevents any definitive conclusion, we observed no difference in metastatic-free survival or overall survival between *MBD4*_{def} and M3 UM patients in contrast with the better outcome in D3 compared with M3 UMs (Figure 2, B and C). Only 1 *MBD4* carrier (UM49) had another cancer, a thyroid papillary carcinoma unrelated to *MBD4* (low TMB without CpG>TpG signature; data not shown).

Discussion

This study demonstrates that *MBD4* is a predisposing gene for UM, conferring an RR of 9.15 for this dismal disease. We further

demonstrated that *MBD4* deficiency specifically predisposes to high-risk M3 UM. One surprising observation for this new cancer-predisposing condition is the absence of early-onset UM. Interestingly, the same is observed in germline *BAP1*-mutant carriers, even with the high penetrance in that context (6,23). A potential explanation for this paradox is that neither *MBD4* nor *BAP1* predisposing genes can act before a first step in the malignant transformation. This first step, presumably the Gαq-initiating event consisting of mutually exclusive activating mutations in *GNAQ*, *GNA11*, *PLCB4*, or *CYSLTR2* (26–29), would be the main determinant of age of onset (Figure 3). The second step in the malignant transformation, composed of mutations in *BAP1*, *SF3B1*, or *EIF1AX* (“BSE” events), leads to a punctuated evolution of UM (16), which would marginally influence age of onset (Figure 3). Whether *MBD4* deficiency favors malignant transformation by increasing driver mutations by modifying the methylation landscape or a distinct mechanism has yet to be determined.

Importantly, germline *MBD4* mutations were recently reported in other types of malignancy, including a polyposis-associated colorectal adenocarcinoma (20), a spiradenocarcinoma

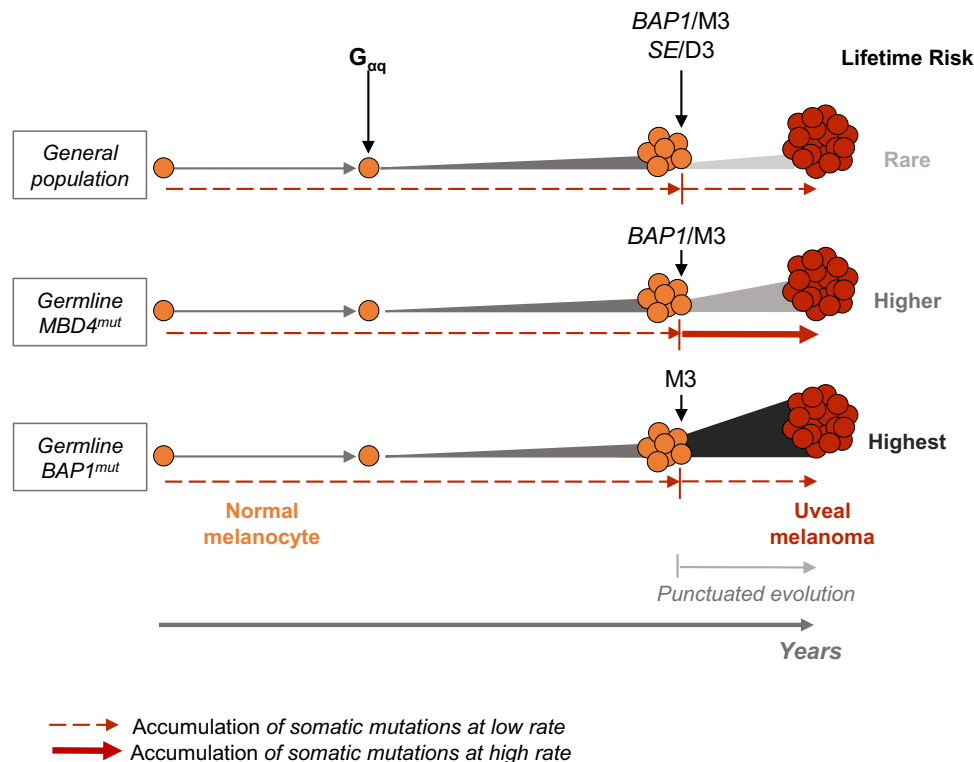


Figure 3. Working model for uveal melanoma (UM) malignant transformation process throughout time in different genetic backgrounds. Germline $MBD4^{mut}/BAP1^{mut}$ population with $MBD4/BAP1$ germline mutation. Following the $G_{\alpha q}$ activating mutation, secondary mutational events in each genetic background are indicated, along with their association with either disomy 3 (D3) or monosomy 3 (M3). SE = SF3B1 or EIF1AX mutation. Relative lifetime risk of UM is represented by the expansion size and color from normal melanocytes to UM. Dashed and full red arrows indicate the rate of accumulation of somatic mutations throughout time (low and high, respectively).

(19), a glioblastoma (8), a pilocytic astrocytoma, a gastric adenocarcinoma, a pancreatic adenocarcinoma, and a pancreatic endocrine tumor (21) (Table 1). Furthermore, although the above case patients were all heterozygous in the germline, biallelic germline deleterious $MBD4$ mutations were reported in 3 individuals who developed acute myeloid leukemias, 2 of which had additional colonic polyposis (13). It is therefore likely that the tumor spectrum associated with $MBD4$ germline mutations will expand when this gene becomes more systematically explored in clinical diagnosis. It is already clear that this spectrum mostly includes relatively rare tumors and some biological tumor features may underlie their association with $MBD4$. To be noticed, both leukemias and UMs associated with $MBD4$ inactivation share a consistent inactivation of the $BAP1$ - $ASXL$ complex (8,13).

It should be noticed that we found no other $MBD4$ -related tumors in our UM series. However, the follow-up and cohort size of this prospective series are limited, and future studies will better characterize the medical history of $MBD4$ carriers. Larger cohorts will also more precisely define $MBD4$ mutation frequency in UM patients and the RR conferred by these mutations. Another limitation to the study is the bias for a European population in our cohort, which reflects the higher incidence of the disease in this population (30).

Interestingly, 5 recurrent $MBD4$ germline deleterious mutations were identified when taking together the LoF variants from our UM cohort and those found in public databases and reports of other cancer types: c.1706G>A [p.Trp569*] (2 patients), c.1562G>T [p.Asp521Profs*4] (4 patients), c.1443delT [p.Leu482Trpfs*9] (3 patients), c.335+1G>A [p.Arg83Profs*5] (2 patients), and c.939insA [p.Glu314Argfs*13] (2 patients) (Table 1),

suggesting founder mutations. Furthermore, the observation of different tumor types associated with the same $MBD4$ germline mutation suggests a more global role of $MBD4$ in cancer predisposition. The peculiar UM proneness in $MBD4$ -mutant carriers (13 out of 23 carriers; Table 1) remains unexplained, but the fact that the frequent M3 in UM inactivates wild-type copies of both $BAP1$ and $MBD4$ suppressor genes may at least in part explain the frequent inactivation of $MBD4$ in UM.

In summary, we described here a novel autosomal-dominant syndrome that is caused by germline mutations of $MBD4$, characterized by a high RR of developing hypermutated UM and possibly other malignancies. Tumors arising in such a context are associated with a CpG>TpG mutator phenotype and have clinical relevance because they may respond to immune-checkpoint inhibitors.

Funding

Supported by funding from the European Commission under the Horizon 2020 program and innovation program under the Marie Skłodowska-Curie grant agreement No 666003 (A-CD), the Horizon 2020 program UM Cure (LM; Project number: 667787), the INCa/ITMO/AVIESAN PhD fellowship program "Formation à la recherche translationnelle" (MR), the Ligue Nationale Contre le Cancer (AE), the Institut National de la Santé et de la Recherche Médicale (INSERM), the Institut Curie, the Ligue Nationale Contre le Cancer (Labellisation), the Programme de Recherche Translationnelle en Cancérologie (PRT-K19-51) INCa-DGOS, and the Site de Recherche Intégrée sur le

Cancer (SiRIC) de l'Institut Curie. The Institut Curie ICGex NGS platform is funded by the EQUIPEX "investissements d'avenir" program (ANR-10-EQPX- 03) and ANR10-INBS-09-08 from the Agence Nationale de la Recherche.

Notes

Role of the funder: The funders had no role in the design of the study; the collection, analysis, and interpretation of the data; the writing of the manuscript; and the decision to submit the manuscript for publication.

Disclosures: The authors have no conflict of interest to declare except for the following: M.R. received a grant support from Bristol-Myers Squibb and Merck.

Author contributions: A.-C.D. and M.R. conceived the study, performed experiments, interpreted the data and wrote the manuscript. A.E. and A.H. performed bioinformatics analyses. L.M., S.A., A.E.-M. and G.P. interpreted the data and provided critical advice. O.M., A.M., S.G. and N.C. provided patients specimens and critical advice. S.D., D.L. and G.P. performed biological analyses. C.C. provided clinical data and critical advice. M.-H.S. conceived and guided the study, interpreted the data and wrote the manuscript. All authors reviewed and approved the final manuscript.

Acknowledgment: The authors thank Joshua J. Waterfall for his insightful comments.

References

1. Aronow ME, Topham AK, Singh AD. Uveal melanoma: 5-year update on incidence, treatment, and survival (SEER 1973-2013). *Ocul Oncol Pathol*. 2018;4(3):145-151.
2. Khoja L, Atenafu EG, Suci S, et al. Meta-analysis in metastatic uveal melanoma to determine progression free and overall survival benchmarks: an international rare cancers initiative (IRCI) ocular melanoma study. *Ann Oncol*. 2019;30(8):1370-1380.
3. Carvajal RD, Schwartz GK, Tezel T, et al. Metastatic disease from uveal melanoma: treatment options and future prospects. *Br J Ophthalmol*. 2017;101(1):38-44.
4. Harbour JW, Onken MD, Roberson EDO, et al. Frequent mutation of BAP1 in metastasizing uveal melanomas. *Science*. 2010;330(6009):1410-1413.
5. Wiesner T, Murali R, Fried I, et al. A distinct subset of atypical Spitz tumors is characterized by BRAF mutation and loss of BAP1 expression. *Am J Surg Pathol*. 2012;36(6):818-830.
6. Walpole S, Pritchard AL, Cebulla CM, et al. Comprehensive study of the clinical phenotype of germline BAP1 variant-carrying families worldwide. *J Natl Cancer Inst*. 2018;110(12):1328-1341.
7. Furney SJ, Pedersen M, Gentien D, et al. SF3B1 mutations are associated with alternative splicing in uveal melanoma. *Cancer Discov*. 2013;3(10):1122-1129.
8. Rodrigues M, Mobuchon L, Houy A, et al. Outlier response to anti-PD1 in uveal melanoma reveals germline MBD4 mutations in hypermutated tumors. *Nat Commun*. 2018;9(1):1866.
9. Johansson PA, Stark A, Palmer JM, et al. Prolonged stable disease in a uveal melanoma patient with germline MBD4 nonsense mutation treated with pembrolizumab and ipilimumab. *Immunogenetics*. 2019;71(5-6):433-436.
10. Hendrich B, Hardeband U, Ng HH, et al. The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. *Nature*. 1999;401(6750):301-304.
11. Yoon JH, Iwai S, O'Connor TR, et al. Human thymine DNA glycosylase (TDG) and methyl-CpG-binding protein 4 (MBD4) excise thymine glycol (Tg) from a Tg: G mismatch. *Nucleic Acids Res*. 2003;31(18):5399-5404.
12. Alexandrov LB, Kim J, Haradhvala NJ, et al. PCAWG Mutational Signatures Working Group. The repertoire of mutational signatures in human cancer. *Nature*. 2020;578(7793):94-101.
13. Sanders MA, Chew E, Flensburg C, et al. MBD4 guards against methylation damage and germ line deficiency predisposes to clonal hematopoiesis and early-onset AML. *Blood*. 2018;132(14):1526-1534.
14. Hashimoto H, Liu Y, Upadhyay AK, et al. Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res*. 2012;40(11):4841-4849.
15. Rothman KJ. *Epidemiology, an Introduction*. 2nd ed. New York, NY: Oxford University Press; 2002.
16. Field MG, Durante MA, Anbunathan H, et al. Punctuated evolution of canonical genomic aberrations in uveal melanoma. *Nat Commun*. 2018;9(1):116.
17. Robertson AG, Shih J, Yau C, et al. Integrative analysis identifies four molecular and clinical subsets in uveal melanoma. *Cancer Cell*. 2017;32(2):204-220.e15.
18. Rodrigues M, Mobuchon L, Houy A, et al. Evolutionary routes in metastatic uveal melanomas depend on MBD4 alterations. *Clin Cancer Res*. 2019;25(18):5513-5524.
19. Davies HR, Hodgson K, Schwalbe E, et al. Epigenetic dysregulation underpins tumorigenesis in a cutaneous tumor syndrome. *bioRxiv* 2019; doi: 10.1101/687459.
20. Tanakaya K, Kumamoto K, Tada Y, et al. A germline MBD4 mutation was identified in a patient with colorectal oligopolyposis and early-onset cancer: a case report. *Oncol Rep*. 2019;42(3):1133-1140.
21. Waszak SM, Tiao G, Zhu B, et al. Germline determinants of the somatic mutation landscape in 2,642 cancer genomes. *bioRxiv* 2017; doi: 10.1101/208330.
22. Morera S, Grin I, Vigouroux A, et al. Biochemical and structural characterization of the glycosylase domain of MBD4 bound to thymine and 5-hydroxymethyluracil-containing DNA. *Nucleic Acids Res*. 2012;40(19):9917-9926.
23. Gupta MP, Lane AM, DeAngelis MM, et al. Clinical characteristics of uveal melanoma in patients with germline BAP1 mutations. *JAMA Ophthalmol*. 2015;133(8):881-887.
24. Prescher G, Bornfeld N, Hirsch H, et al. Prognostic implications of monosomy 3 in uveal melanoma. *Lancet*. 1996;347(9010):1222-1225.
25. Singh AD, De Potter P, Fijal BA, et al. Lifetime prevalence of uveal melanoma in white patients with ocular (dermal) melanocytosis. *Ophthalmology*. 1998;105(1):195-198.
26. Johansson P, Aoude LG, Wadt K, et al. Deep sequencing of uveal melanoma identifies a recurrent mutation in PLCB4. *Oncotarget*. 2016;7(4):4624-4631.
27. Van Raamsdonk CD, Bezrookove V, Green G, et al. Frequent somatic mutations of GNAQ in uveal melanoma and blue naevi. *Nature*. 2009;457(7229):599-602.
28. Van Raamsdonk CD, Griewank KG, Crosby MB, et al. Mutations in GNA11 in uveal melanoma. *N Engl J Med*. 2010;363(23):2191-2199.
29. Moore AR, Ceraudo E, Sher JJ, et al. Recurrent activating mutations of G-protein-coupled receptor CYSLTR2 in uveal melanoma. *Nat Genet*. 2016;48(6):675-680.
30. Mobuchon L, Battistella A, Bardel C, et al. A GWAS in uveal melanoma identifies risk polymorphisms in the CLPTM1L locus. *NPJ Genom Med*. 2017;2(1):1-7.

SUPPLEMENTARY MATERIALS

Supplementary Methods

***MBD4* targeted-sequencing**

Germline DNA of UM patients 1 to 1,099 from the UM consecutive series (prior to the removal of the 6 abovementioned patients) and tumor DNA of patients UMT1 to UMT192 from the M3 UM tumor series were plated in 138 pools (consecutive series) of 8 samples (except for 1 pool of 9 including a positive control, 1 pool of 7 and 2 pools of 6) and 48 pools (tumor series) of 4 samples, in equimolar amounts. Pooled DNA samples were used to amplify the 8 coding exons of *MBD4* by 2 multiplex-PCRs yielding fragments between 200 and 485bp (Supplementary Table 3), according to manufacturer's instructions (Phusion High-Fidelity DNA Polymerase, New England Biolabs). Libraries and index adaptors for exon-sequencing were prepared using the AmpliSeq Library PLUS kit (Illumina). Final NGS libraries were sequenced with paired-end primers generating two 300bp reads on an Illumina MiSeq instrument, generating an output of ~25 million sequencing reads. Full workflows from *MBD4* targeted-sequencing to identification of deleterious variants for both the germline consecutive series and the tumor M3 series are described in Supplementary Figures 1A and 1B, respectively. Deconvolution of the identified pooled DNA samples containing one patient with an *MBD4* variant was carried out by Sanger sequencing. DNA sequences were visualized under FinchTV chromatogram viewer and the identified variants were confirmed in both sequencing directions. Zoomed-in images of chromatograms for the 28 variants are displayed in Supplementary Figure 5.

***MBD4* variant calling and filtering**

For data processing and analysis of *MBD4*-targeted sequencing, FastQC was used to control the quality of sequencing data. Sequenced reads were aligned to the chromosome 3 of the human

genome (hg19 assembly) with BWA MEM (version 0.7.15). The primers were soft-clipped with BAMclipper (1). Base quality score recalibration was applied to the BAM files according to GATK Best Practices (2) (version 4.0.11.0). Four tools were used to detect the variants in the pooled samples in single sample mode: Freebayes (3) (version v1.2.0-2-g29c4002), HaplotypeCaller (version 4.0.11.0), Mutect2 (version 4.0.11.0) and Bcftools mpileup (4) (version 1.9). The ploidy arguments were set to NumberOfSample*2 for germline calls and NumberOfSample*4 for somatic calls. The union of all the variants detected was annotated with ANNOVAR (5) according to different databases: ensGene, avsnp150 (6), cosmic84 (7), popfreq_all_20150413 and dbnsfp33a. UTR variants were filtered out. Variants with a position depth (DP) inferior to 500 or with a variant allele frequency (VAF) inferior to 1.5% (<2% for tumor variants) were also taken out. Finally, variants with a frequency in the general population > 1% and/or those with a germline frequency > 1% in the tested population (>2% for tumor variants) were filtered out. Intronic variants >30bp away from the nearest exons were also removed. *In silico* tools SIFT (v5.2.2) (8) and PolyPhen-2 (v2.2.2) (9) were used to predict the deleterious effect of the identified variants (Supplementary Table 1).

Quality control of the targeted-sequencing pipeline

To check the sensitivity of the pipeline, the frequency of the common SNPs (Minor Allele Frequency, MAF>1%) was estimated by the cumulative VAF found in positive pools for each variant. All common SNPs present in the targeted sequence were found at the expected frequencies (compared to the Non-Finnish European population subset of GnomAD) (Supplementary Table 4).

Evaluation of splice-site mutations by exon-trapping

The online tool Human Splicing Finder was used to predict cryptic acceptor and/or donor sites among *MBD4* missense and intronic variants located <30bp away from the nearest exon. For each

candidate variant, SNP-centered amplicons of ~250bp were generated by PCR amplification of genomic DNA from UM patients harboring the *MBD4* variants and from HEK293T cells, using specific primer sets (Supplementary Table 3). Resulting amplicons were cloned with In-fusion HD cloning kit (Clontech) into the *Bam*H1 site of a pET01 ExonTrap vector (Mobictec) containing functional donor and acceptor sites. Plasmid DNA was extracted and sequences were verified by Sanger sequencing. HEK293T cells were then transfected with minigene constructs containing the candidate splice site mutations using Lipofectamine 2000 reagent (Invitrogen) following the manufacturer's instructions. After 48 hours, total RNA was extracted from cells and used as a template for cDNA synthesis with the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). Synthesized cDNA was then amplified by RT-PCR using a universal forward primer and reverse-specific primers or vice-versa, depending on the splice site (donor or acceptor) being tested (Supplementary Table 3). Fragments were analyzed on a 2.5% agarose gel (Supplementary Figure 2).

Generation of *MBD4* mutant vectors by site-directed mutagenesis

To assess the enzymatic activity of wild-type and mutant MBD4 proteins, full length *MBD4* cDNA (NCBI Reference Sequence NM_003925.2) coding for MBD4 protein isoform 1 (longest variant with a length of 580 amino acids, NP_003916.1), was obtained by genomic PCR amplification from a lymphoblastoid cell line followed by reverse-transcription using standard procedures. *MBD4* cDNA was then cloned into a pET28b bacterial expression vector (Merck) carrying an N-terminal 6-His-tag using the In-Fusion directional recombination cloning kit (primers described in Supplementary Table 3) following the manufacturer's protocol (Clontech Laboratories). Successful cloning was verified by Sanger sequencing of the full *MBD4* sequence. To generate the MBD4 mutants containing the missense variants (p.Asn551Ser, p.Arg468Trp and p.Asn467Ser) and the stop gain variant (p.Trp569*) within the glycosylase domain, the QuikChange XL Site-Directed

Mutagenesis kit (Agilent) was used according to the manufacturer's protocol, with 4 distinct primer pairs (Supplementary Table 3).

Expression and purification of human recombinant MBD4

Wild-type and mutant pET28b-MBD4 vectors were expressed in One Shot® BL21 Star (DE3) *E. Coli* bacteria cells (Thermo Scientific), along with pRare vector (Thermo Scientific). They were grown in 2x YT medium supplemented with kanamycin (50ug/mL) and chloramphenicol (34ug/mL) and incubated for 16 hours at 37°C with shaking. Cells were diluted to OD_{600nm}=0.1 and cultured in 2x YT medium until it reached an OD_{600nm} of 0.6 to 0.8. Cell cultures were then induced with 1mM IPTG in 2x YT medium with appropriate antibiotics and incubated for 16 hours at 20°C with shaking. Cells were then pelleted by centrifuging 15min at 4,000rpm at 4°C and re-suspended in lysis buffer (PBS1X, 350mM NaCl, 20mM imidazole pH 7.4, 10mM MgCl₂, 0.5% Triton X-100, 1mg/mL lysozyme, 0.2μL/mL benzonase, 1X protease inhibitor) before incubation at 4°C for 1 hour. To collect the insoluble fraction, containing the MBD4 proteins expressed as inclusion bodies, cell lysates were centrifuged for 1 hour at 20,000rpm at 4°C and the insoluble pellet was re-suspended in wash buffer containing 350mM NaCl, 8M urea and 20mM imidazole. The 6His-MBD4 protein was isolated from the crude lysate in with a histidine-tagged purification resin (Ni Sepharose 6 Fast Flow, GE Healthcare) loaded onto a resin Ni Sepharose 6 Fast Flow column (GE Healthcare) and eluted with cold elution buffer containing 8M urea, 350mM NaCl and 250mM imidazole in 1X PBS. Directly after elution, the MBD4 wild-type and mutant proteins were refolded overnight at 4°C in a refolding buffer (0.2M Tris-HCl pH7.4, 10mM EDTA, 0.6M L-Arginine HCl, 20% glycerol, 50mM NaCl and 1mM DTT). Finally, soluble MBD4 was dialyzed in a 300X volume buffer containing 20mM Tris-HCl, 20% glycerol, 150mM NaCl, 1mM EDTA and 1mM DTT. Proteins were verified by SDS-PAGE using 4-20% polyacrylamide Mini-PROTEAN® TGX Stain-Free Precast Gels (Bio Rad) with Tris-Glycine-SDS running buffer at 180V for 40 minutes, and

visualized using the ChemiDoc imaging system (Bio Rad). Bradford protein assay (Bio Rad) was performed to quantify proteins.

Glycosylase activity assay

In vitro MBD4 glycosylase assay testing wild-type and mutant MBD4 proteins was performed as previously described (10,11) using the following 32-bp FAM-labeled DNA probes:

FAM-5'-TCGGATGTTGTGGGTCAG(C/T)GCATGATAGTGTA-3';

5'-TACACTATCATGCGCTGACCCACAACATCCGA-3'. Double-stranded matched and mismatched oligonucleotides were hybridized as previously described (10,11). 0.5µM of purified human recombinant MBD4 (wild-type or mutant) was added to 0.5µM FAM-labeled 32bp oligonucleotides and enzymatic activity was assessed by denaturing polyacrylamide gel electrophoresis. Single-stranded, FAM-labeled products were visualized using the ChemiDoc imaging system (Bio Rad). A blot of the glycosylase assay using both perfectly-matched and mismatched oligonucleotides is presented in Supplementary Figure 6.

For the loading control presented in Figure 1B, the same amount (0.5µM) of each recombinant MBD4 protein was loaded onto a 4-20% polyacrylamide Mini-PROTEAN® TGX Stain-Free Precast Gel (Bio Rad). Following SDS-PAGE, the gel was imaged with ChemiDoc Stain-Free mode.

Whole-Exome Sequencing and mutation calling

Samples of the 9 UM patients harboring *MBD4* variants with both germline and tumor samples available (UM75, UM102, UM350, UM605, UM656, UMT45, UMT61, UMT88 and UMT162) were histologically reviewed by a pathologist prior to DNA extraction. DNA was extracted by the *Centre*

de Ressources Biologiques (Institut Curie tumor biobank), purified on Zymo-Spin IC (Zymo Research), and quantified by Qubit (Thermo Fisher Scientific). 500ng to 1ug DNA was used to prepare 100bp paired-end multiplexed WES libraries following the Sureselect Agilent-XT2 protocol (Agilent technologies). Libraries were sequenced on the NovaSeq 6000 System (Illumina) and coverage depth was set up a priori at 30X for germline and 100X for somatic DNA. After removing duplicates, WES data underwent variant calling for SNVs and indels using the combination of two variant callers: HaplotypeCaller (12) and SAMtools mpileup. Union of variants detected with these 2 algorithms were annotated using ANNOVAR, with the following databases: ensGene, avsnp150 (6), popfreq_all and dbnsfp33a. Somatic variants with less than 10 reads of DP in germline and/or less than 10 reads of somatic DP and/or at least 1 read of germline AD and/or less than 5 reads of somatic AD and/or a population frequency higher than 1% (popfreq_all>0.01) and/or with significant strand bias (p-value from Fisher's exact test less than 0.05) were filtered out. Finally, all somatic mutations called by this procedure were controlled manually using the Integrative Genomics Viewer (IGV) by at least two authors. Because alterations of *BAP1*, *EIF1AX* and *SRSF2* in UM may be difficult to call, these genes were entirely checked on IGV in all samples. A list of all somatic variants called using this pipeline in the 9 UM tumors, and the individual tumor mutational profiles are displayed in Supplementary Table 5 and Supplementary Figure 3, respectively. Results for tumor mutation burden and for mutational status of *MBD4*, *GNAQ*, *GNA11*, *BAP1*, *EIF1AX* and *SF3B1* are displayed in Figure 1C and Supplementary Table 2. Copy number and tumor cellularity were obtained using Facets (13), with 20 as minimal mapping quality, 13 as minimal base quality, and between 25 and 1000 read depth to output a position. Cancer Cell Fraction (CCF) of all mutations was obtained using PyClone (14) with binomial model and default parameters. A plot of the VAF distribution for each of the 7 exomes of patients with deleterious *MBD4* mutations is presented in Supplementary Figure 4.

Survival Analysis

Among the 1,093 UM patients in the UM consecutive series, 323 had medical records with available tumor chromosome 3 status: the 8 *MBD4-deficient* patients, 198 *MBD4*-wild-type M3 patients and 117 *MBD4*-wild-type D3 patients. All of the above patients underwent a treatment of the primary disease. Liver ultrasound was performed prior to treatment. Local treatment consisted of enucleation for large tumors and proton beam radiotherapy or iodine 125 brachytherapy for small-to-medium-sized tumors. Patients were then seen every 6 months with complete eye examination, ultrasound bio-microscopy and liver ultrasonography or MRI. Suspicion of liver metastasis was systematically confirmed by liver biopsy. Tumor genomic profiles and follow-up events (distant recurrences, death from uveal melanoma or from any other cause) were prospectively collected. The French Death Registry was consulted for patients lost to follow-up. Patients with metastatic disease were treated by an oncologist at our institution. Survival analyses (metastasis-free survival, MFS, and overall survival, OS) were carried out on all the above patients (M3, D3, and *MBD4*-deficient groups). Time t0 corresponds to the treatment of the primary UM tumor (less than 1 month after UM diagnosis). MFS was defined as the interval between the date of diagnosis of primary UM and the date of distant metastasis (first imaging) or death from any cause, whatever comes first. Survival distributions were estimated by the Kaplan-Meier method and compared using the log-rank test.

Data availability

Sequencing data have been deposited in and are available from the European Genome-phenome Archive database under number EGAS00001003941.

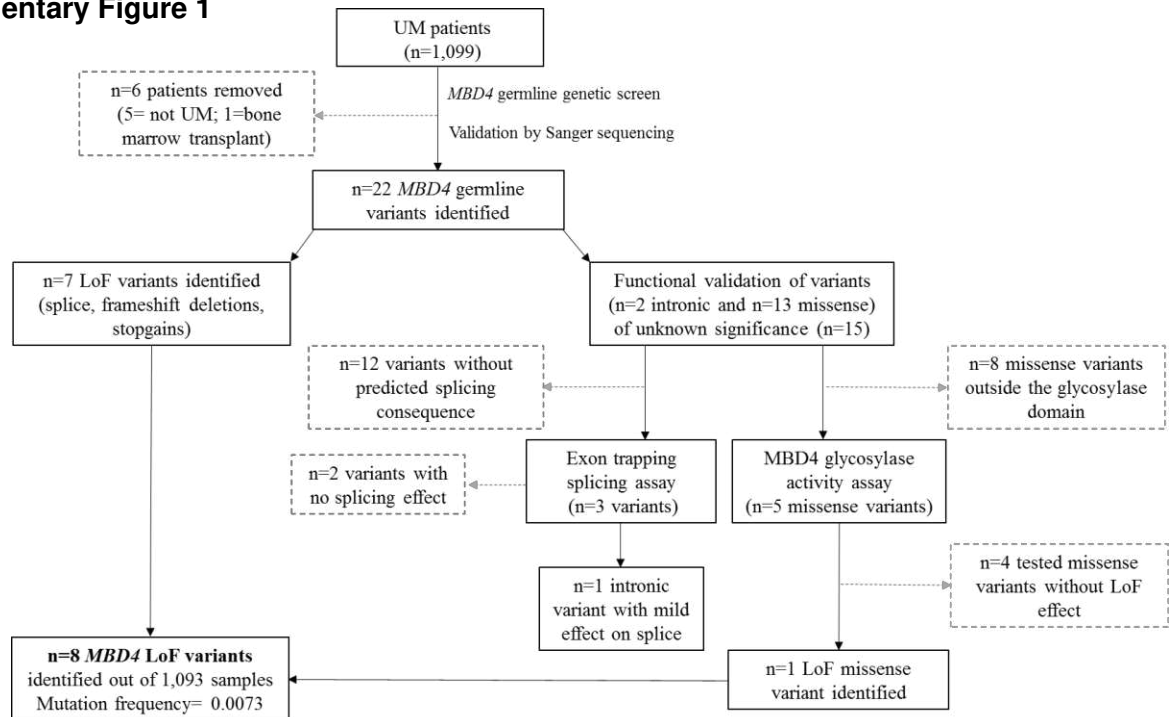
Supplementary References

1. Au CH, Ho DN, Kwong A, *et al.* BAMClipper: removing primers from alignments to minimize false-negative mutations in amplicon next-generation sequencing. *Sci Rep* 2017;7(1):1567.
2. DePristo MA, Banks E, Poplin R, *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011;43(5):491-8.
3. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. *arXiv* 2012;1207.3907.
4. Li H, Handsaker B, Wysoker A, *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;25(16):2078-9.
5. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;38(16):e164.
6. Sherry ST, Ward M, Sirotkin K. dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome Res* 1999;9(8):677-9.
7. Forbes SA, Beare D, Gunasekaran P, *et al.* COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res* 2015;43(Database issue):D805-11.
8. Sim NL, Kumar P, Hu J, *et al.* SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res* 2012;40(Web Server issue):W452-7.
9. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet* 2013;Chapter 7:Unit7 20.
10. Sanders MA, Chew E, Flensburg C, *et al.* MBD4 guards against methylation damage and germ line deficiency predisposes to clonal hematopoiesis and early-onset AML. *Blood* 2018;132(14):1526-1534.
11. Hashimoto H, Liu Y, Upadhyay AK, *et al.* Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res* 2012;40(11):4841-9.
12. Lescai F, Marasco E, Bacchelli C, *et al.* Identification and validation of loss of function variants in clinical contexts. *Mol Genet Genomic Med* 2014;2(1):58-63.
13. Shen R, Seshan VE. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res* 2016;44(16):e131.
14. Roth A, Khattra J, Yap D, *et al.* PyClone: statistical inference of clonal population structure in cancer. *Nat Methods* 2014;11(4):396-8.

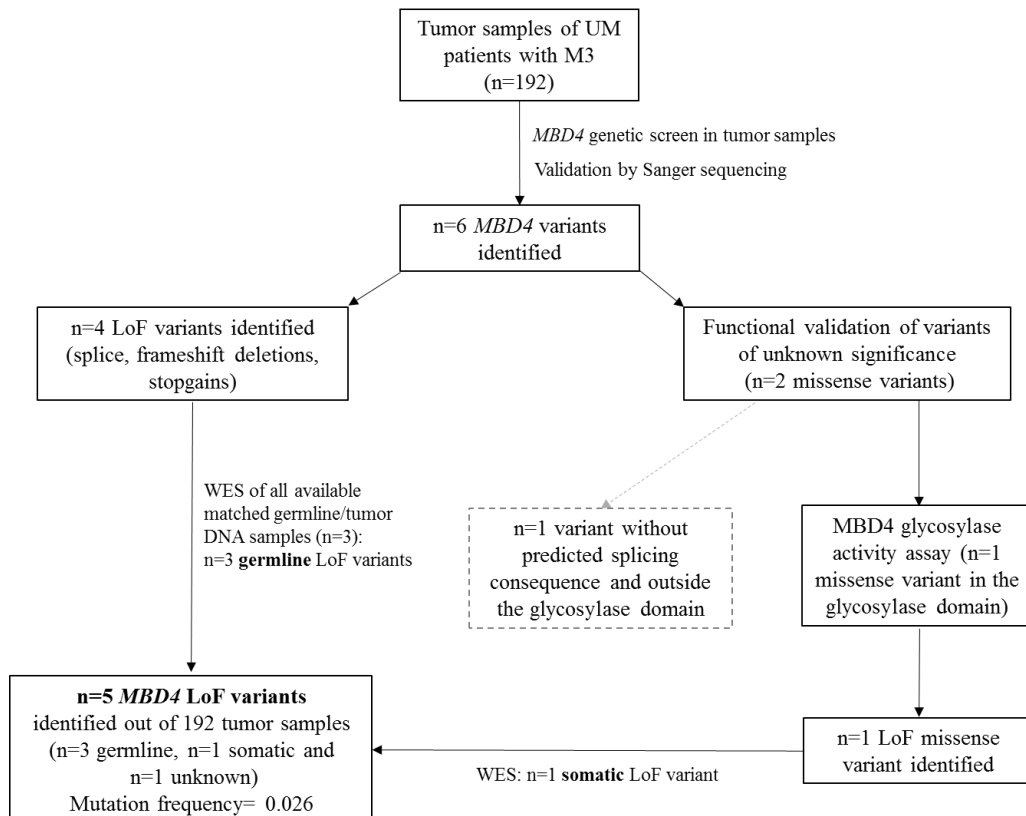
Supplementary Figures

Supplementary Figure 1

A.

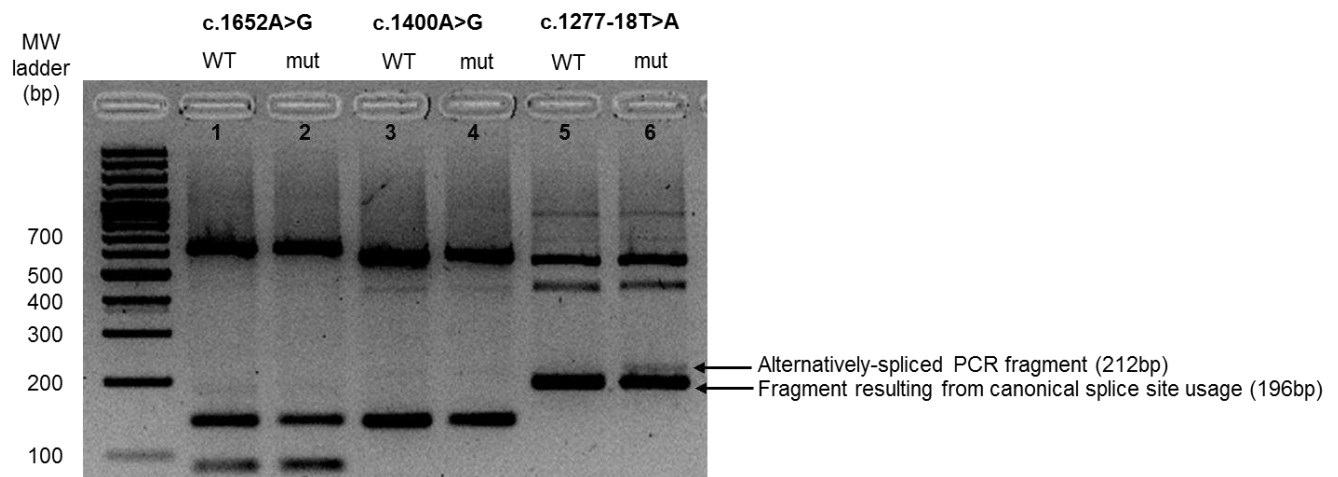


B.



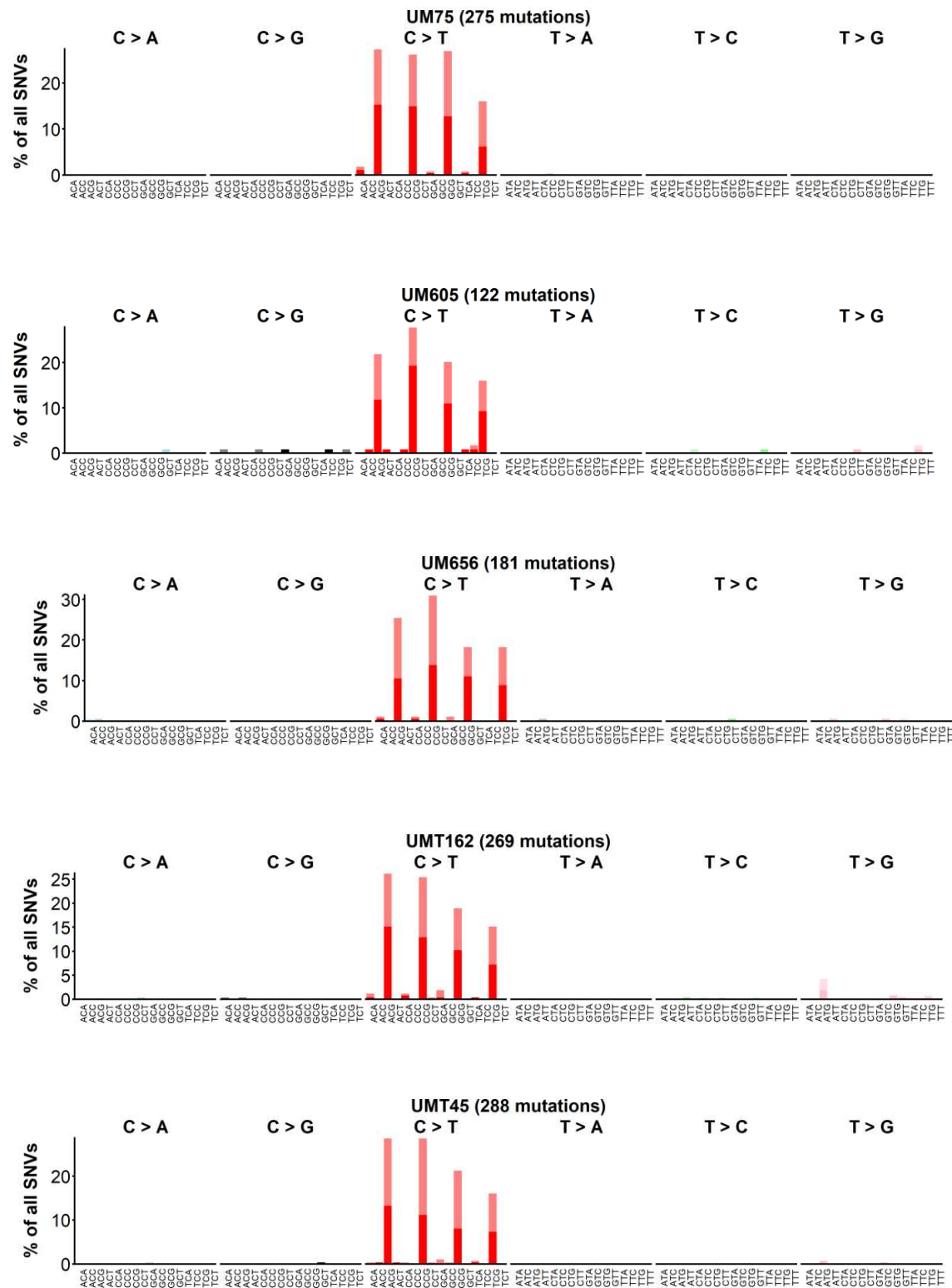
Supplementary Figure 1. Full workflow for the identification of loss-of-function *MBD4* variants among (A) our in-house cohort of 1,099 consecutive uveal melanoma patients and (B) our monosomy 3 tumor series of 192 uveal melanoma patients. Boxes in dotted lines indicate the patients that were excluded or variants that were either not tested or shown not to have a deleterious effect. UM: uveal melanoma; LoF: loss-of-function; WES: whole-exome sequencing.

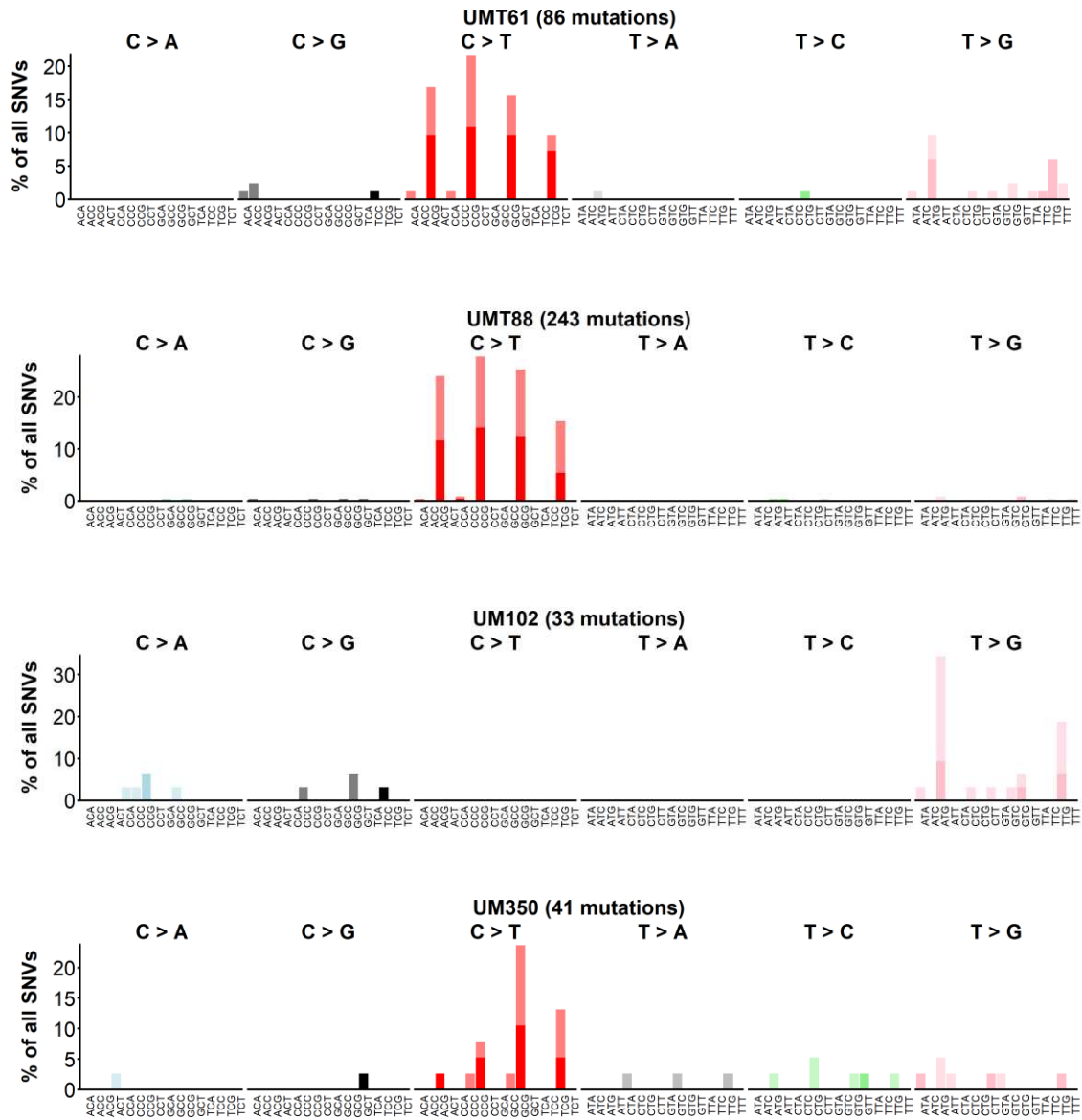
Supplementary Figure 2



Supplementary Figure 2. Exon-trapping assay assessing the splicing activity of variants c.1652A>G, c.1400A>G and c.1277-18T>A predicted to have a potential splice effect by Splice Site Finder¹⁸. Lanes 1-2: c.1652A>G, lanes 3-4: c.1400A>G, and lanes 5-6: c.1277-18T>A. Details regarding the construction of the three minigene vectors are described in Methods and the primers used for RT-PCR amplification are listed in Supplementary Table 3. Bands represent the migration of the various fragments obtained by RT-PCR amplification centered on the 3 SNPs being tested ("WT"= wild-type, "mut"= mutant) on a 2.5% agarose gel. Shifts in band size between WT and mutant SNP indicate the use of an alternative splice site. Molecular weight (MW) ladder indicates DNA size in base pairs (bp).

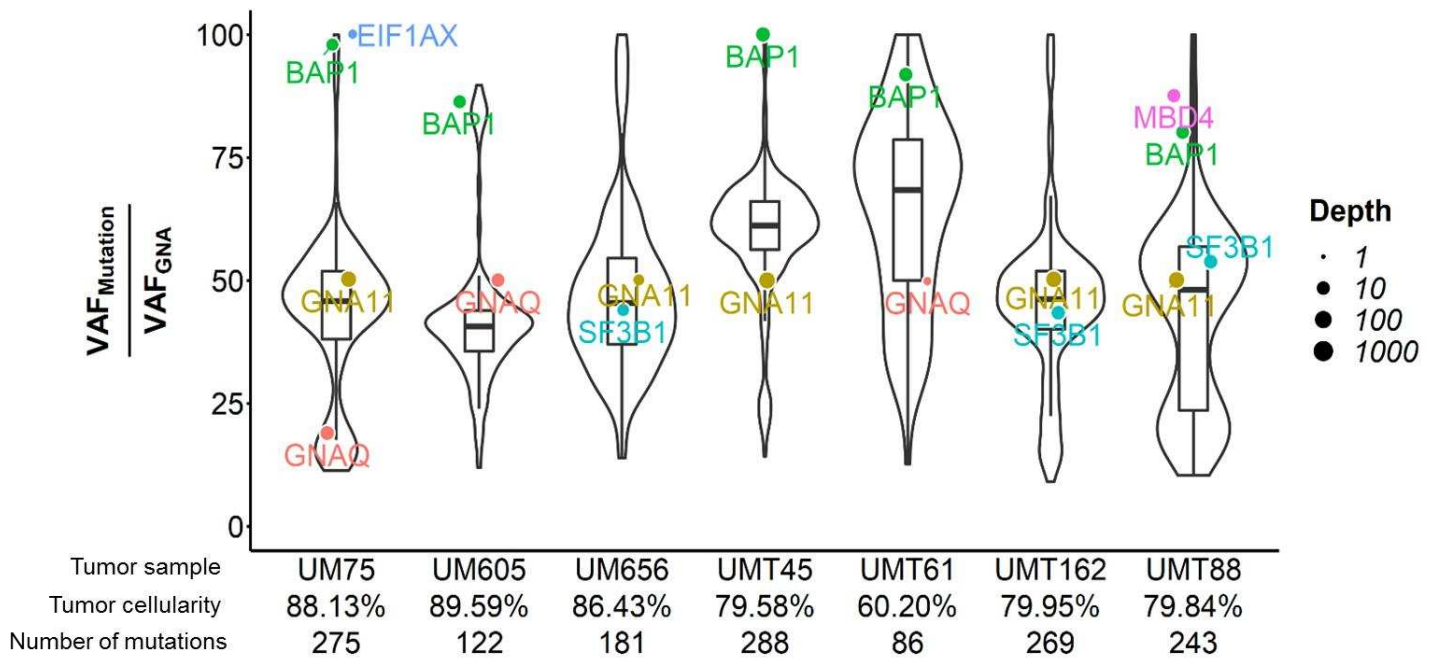
Supplementary Figure 3





Supplementary Figure 3: Individual mutational patterns of tumors of UM patients with an *MBD4* mutation (7 with a deleterious mutation and 2 with benign mutations, UM102 and UM350), based on the relative proportion (y-axis) of each of the 96 types of trinucleotide substitution (x-axis). Dark/bright colors correspond to sense/anti-sense strands.

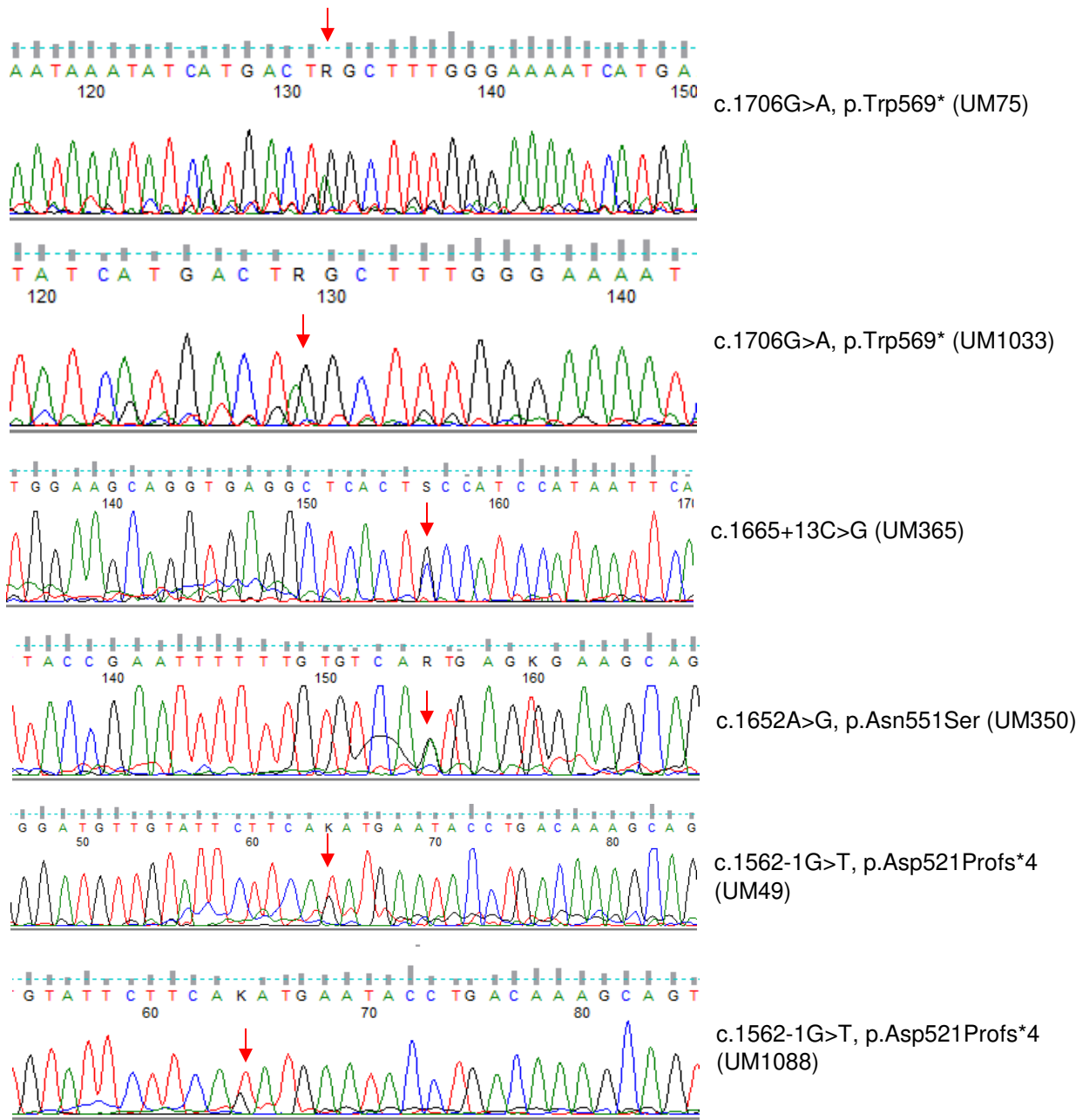
Supplementary Figure 4

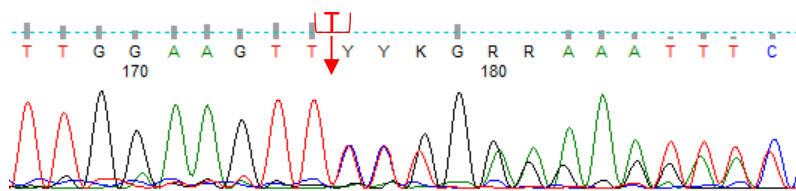


Supplementary Figure 4. Variant Allele Frequency (VAF) distribution of all somatic variants in the tumors of the 7 *MBD4*-deficient UM patients with available exome data. The box plot represents the median VAF and Q1 – Q3 quartile interval. The key driver events (mutations in *BAP1*, *GNA11*, *GNAQ*, *EIF1AX*, *SF3B1* and/or *MBD4*) are indicated, along with the WES coverage depth at the variant position, represented by the circle size. The VAF of each mutation is normalized to the VAF of the *GNAQ/GNA11* event set at 50% to account for tumor cellularity. Width of the violin plot represents the number of variants at a given VAF. Tumor cellularity (%) and total number of mutations in each tumor are indicated. A list of all somatic variants detected in the 7 tumors is presented in Supplementary Table 5.

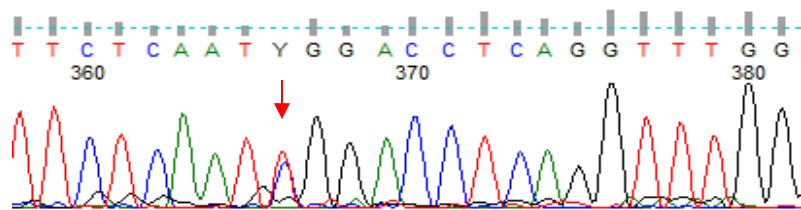
Supplementary Figure 5

Consecutive germline UM series:

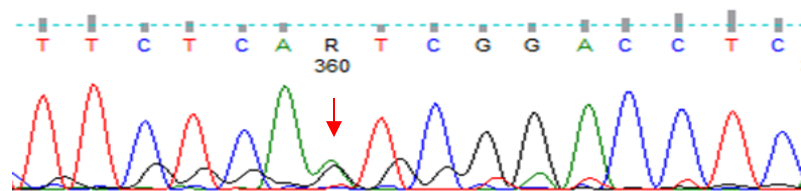




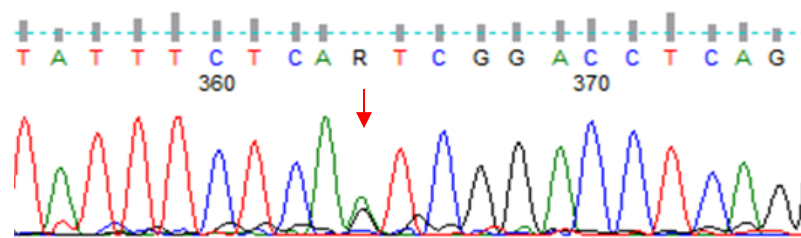
c.1443delT, p.Leu482Trpfs*9 (UM656)



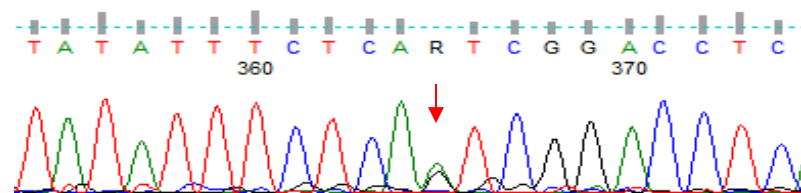
c.1402C>T, p.Arg468Trp (UM293)



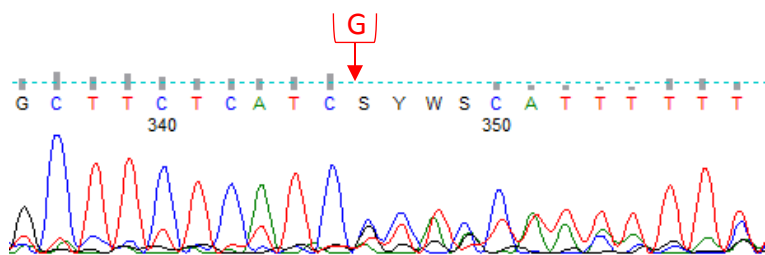
c.1400A>G, p.Asn467Ser (UM867)



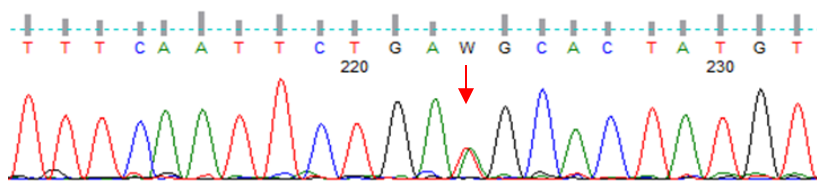
c.1400A>G, p.Asn467Ser (UM75)



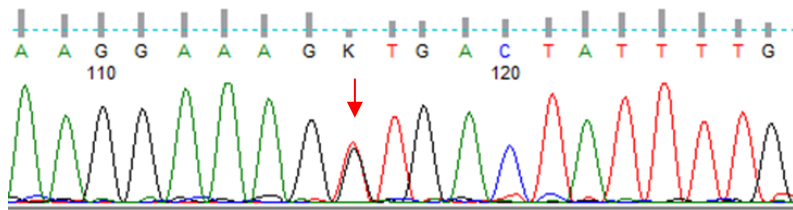
c.1400A>G, p.Asn467Ser (UM547)



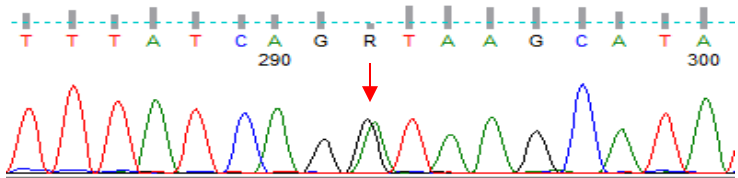
c.1384delG, p.Ala462Leufs*29 (UM605)



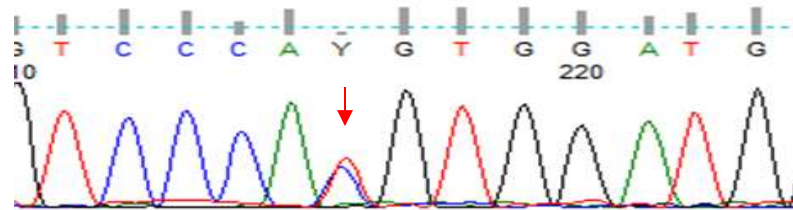
c.1277-18T>A (UM343)



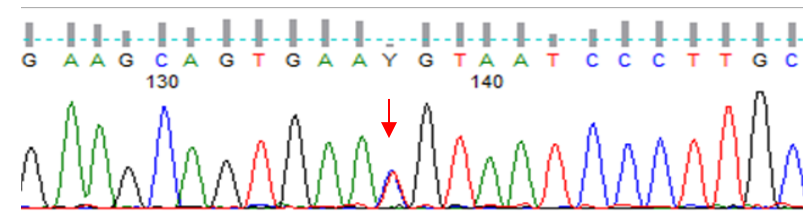
c.703G>T, p.Val235Leu (UM616)



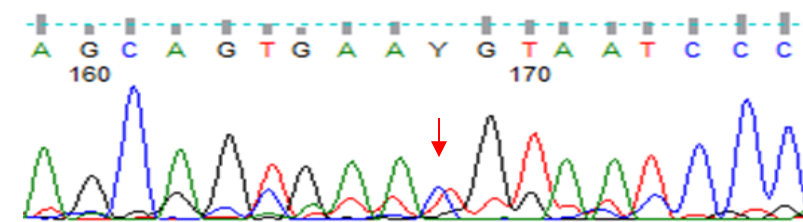
c.335+1G>A, p.Arg83Profs*5 (UM436)



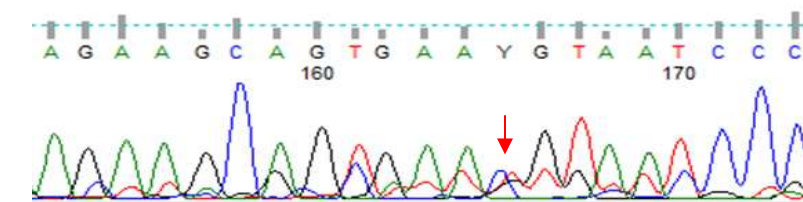
c.262T>C, p.Cys88Arg (UM1055)



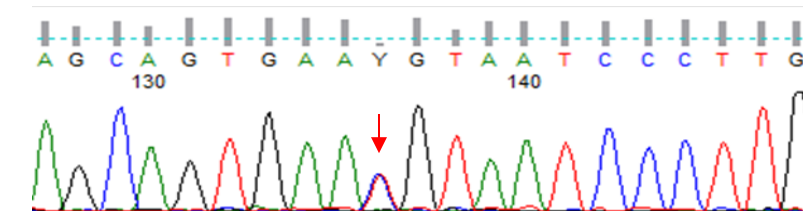
c.181T>C, p.Cys61Arg (UM402)



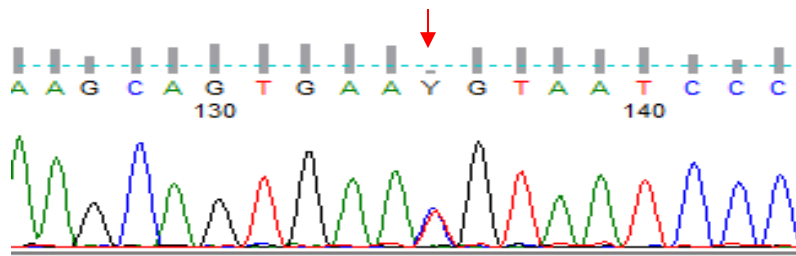
c.181T>C, p.Cys61Arg (UM835)



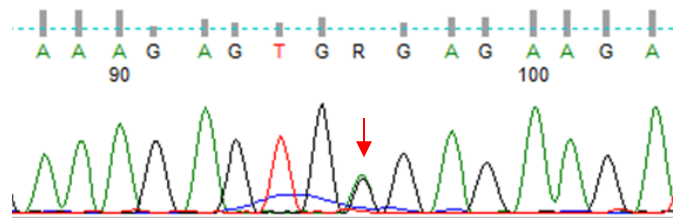
c.181T>C, p.Cys61Arg (UM872)



c.181T>C, p.Cys61Arg (UM552)

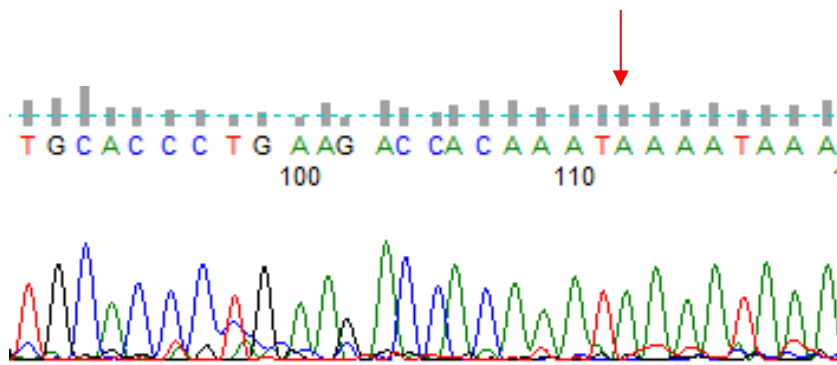


c.181T>C, p.Cys61Arg (UM42)

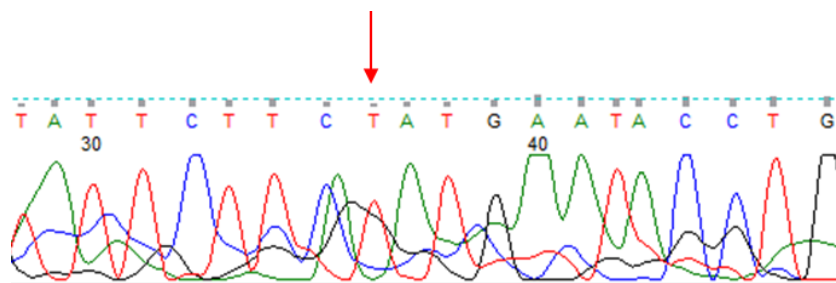


c.181T>C, p.Cys61Arg (UM102)

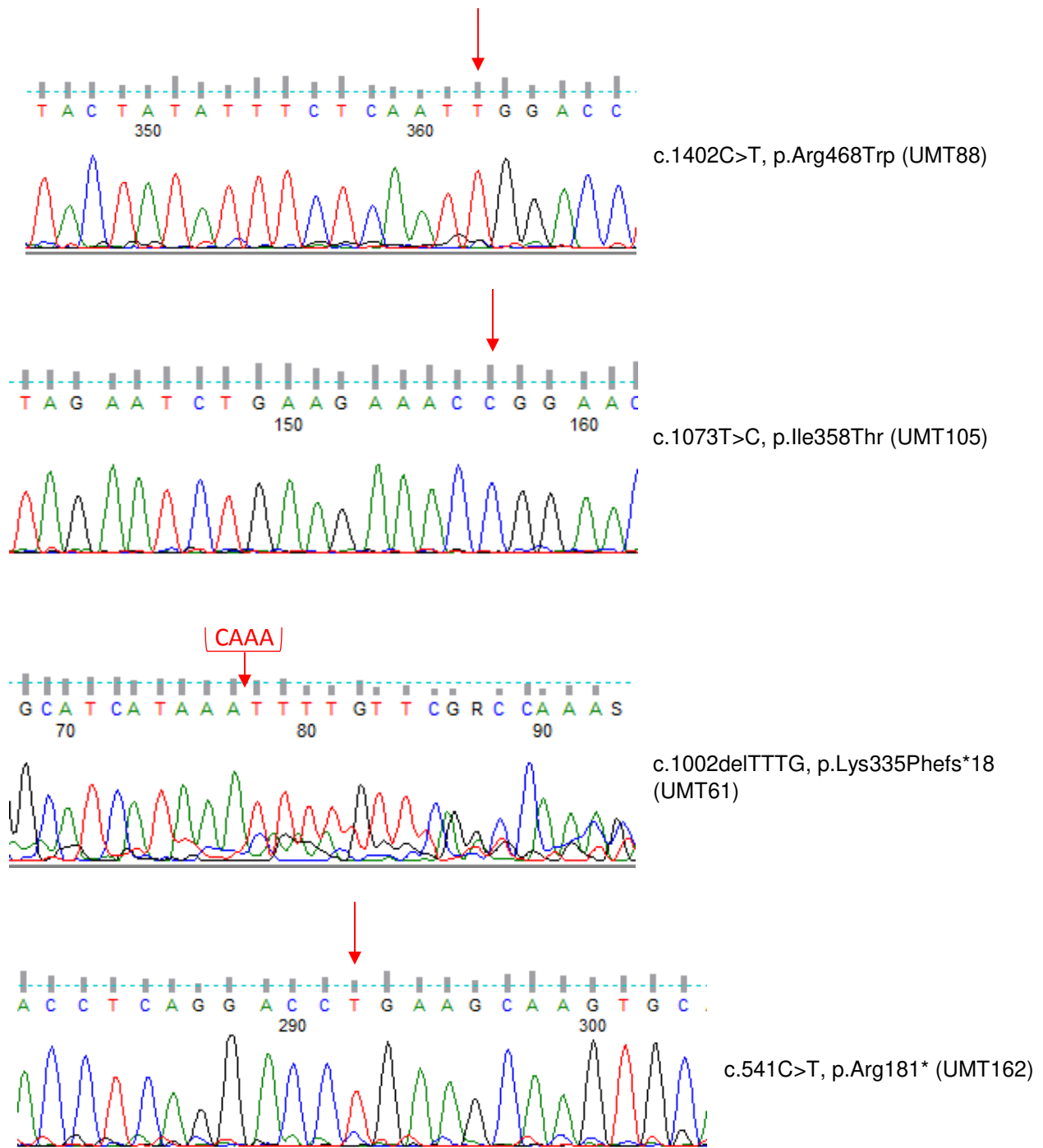
Monosomy 3 tumor UM series:



c.1688T>A, p.Leu563* (UMT62)

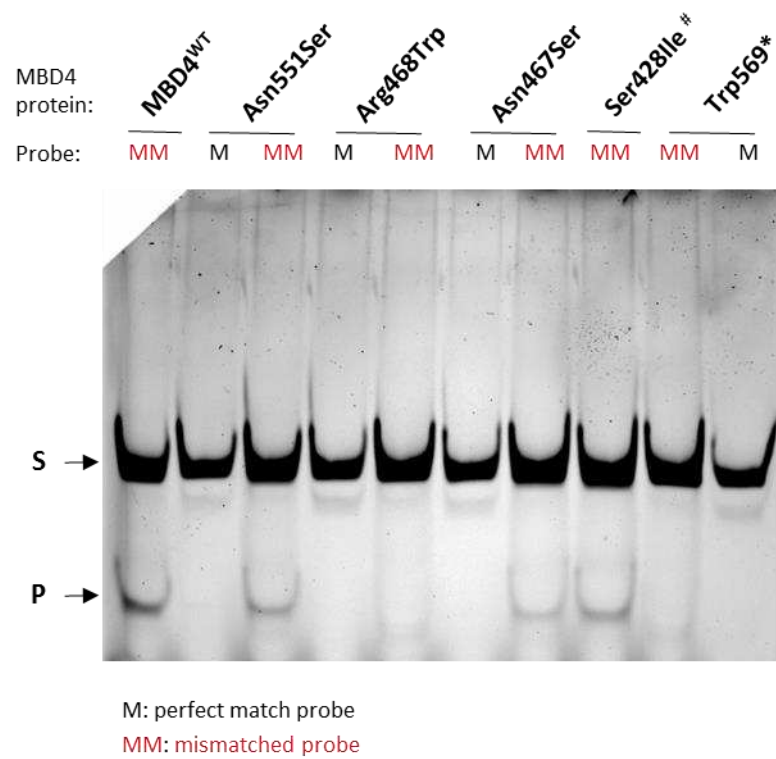


c.1562-1G>T, p.Asp521Profs*4 (UMT45)



Supplementary Figure 5. Chromatograms of all *MBD4* variants validated by Sanger sequencing. These include 22 variants from the consecutive germline UM series and 6 variants from the M3 UM tumor series. Red arrows indicate the variant position. The corresponding variant nomenclature and UM patient harboring it are indicated.

Supplementary Figure 6



[#] the Ser428Ile variant is not part of the current study.

Supplementary Figure 6. Glycosylase activity assay of recombinant wild-type MBD4 (MBD4^{WT}) and mutant proteins, using the perfectly-matched (M, in black) and mismatched (MM, in red) probes. Substrate = S; cleaved Product P.

Supplementary Table 1. Full description of the 28 *MBD4* variants validated by Sanger sequencing, including 22 from the germline consecutive UM series and 6 from the monosomy 3 UM tumor series.

Patient	Position on chromosome 3	Variant Annotation (codon)	Variant Annotation (protein)	dbSNP	Mutation type			Poly-Phen2 prediction	SIFT prediction	Glycosylase assay	LOH (WES or Sanger)(%) [‡]	Tumor Mutation N burden (#variants)	Splice prediction (Splice Site Finder, SSF)	SSF score (/100) vs. canonical donor/acceptor	Splicing effect (exontrap)	GnomAD [§] v2.1.1 variant allele frequency (general population)			GnomAD v2.1.1 variant allele frequency (Non-Finnish Europeans)		
																variant allele count	total allele count	frequency	variant allele count	total allele count	frequency
UM75	129150381	c.1706G>A	p.Trp569*	rs939751619	stop_gain	LoF*	GL [‡]	-	-	Inactive [§]	Yes (99.0%)	275	-	-	ND	2	282518	7,08E-06	2	129130	1,55E-05
UM1033											ND	ND									
UMT62											Yes (Sanger)	ND									
UM365	129151333	c.1665+13C>G	c.1665+13C>G	rs764602863	intronic	intronic	GL	Benign	Tolerated	ND	ND	ND	None	None	ND	8	251454	3,18E-05	8	113752	7,03E-05
UM350	129151359	c.1652A>G	p.Asn551Ser	rs577234840	nonsynonymous_SNV	missense	GL	Benign	Deleterious	Benign	Yes (99.8%)	40	Cryptic donor/acceptor site	70.5 vs 90.6	No	7	282894	2,47E-05	6	129206	4,64E-05
UM49	129151450	c.1562-1G>T	p.Asp521Profs*4	rs778697654	splice_acceptor	LoF	GL	-	-	ND	ND	ND	-	-	ND	10	251472	3,98E-05	5	113766	7,03E-05
UM1088											Yes (100%)	288									
UMT45											Yes 97.7%	180									
UM656	129152059	c.1443delT	p.Leu482Trpfs*9	rs769076971	frameshift_deletion	LoF	GL	-	-	ND	Yes 97.7%	180	-	-	ND	3	251476	1,19E-05	3	113752	2,64E-05
UM293	129152702	c.1402C>T	p.Arg468Trp	rs1380952147	nonsynonymous_SNV	LoF (missense)	GL	Probably damaging	Deleterious	Inactive	ND	ND	None	None	ND	1	251308	3,98E-06	0	113630	0.00
UMT88											Yes (100%)	243									
UM867											ND	ND									
UM75	129152704	c.1400A>G	p.Asn467Ser	rs78782061	nonsynonymous_SNV	missense	GL	Possibly damaging	Tolerated	Benign	Yes (0.0%)**	275	Cryptic acceptor site	79.9 vs 71.1	No	479	282746	1,69E-03	349	129084	2,64E-05
UM547											ND	ND									
UM605											Yes (89.8%)	122									
UM343	129152845	c.1277-18T>A	c.1277-18T>A	rs1434697310	intronic	intronic	GL	-	-	ND	ND	ND	Cryptic acceptor site	76.7 vs 77.8	Mild	1	251458	3,98E-06	1	113742	8,79E-06
UMT105	129155414	c.1073T>C	p.Ile358Thr	rs2307298	nonsynonymous_SNV	missense	ND	Benign	Tolerated	ND	Yes (Sanger)	ND	None	None	ND	2041	282712	7,22E-03	1425	129070	1,10E-02
UMT61	129155482	c.1002delTTTG	p.Lys335Phefs*18	rs1443006605	frameshift_deletion	LoF	GL	-	-	-	Yes (97.2%)	85	-	-	ND	2	251306	7,96E-06	0	113650	0.00
UM616	129155784	c.703G>T	p.Val235Leu	-	nonsynonymous_SNV	missense	GL	Benign	Deleterious	ND	ND	ND	None	None	ND	-	-	-	-	-	-
UMT162	129155946	c.541C>T	p.Arg181*	rs1270271346	stop_gain	LoF	GL	-	-	ND	Yes (99.9%)	269	-	-	ND	2	282478	7,08E-06	2	128972	1,55E-05
UM436	129156562	c.335+1G>A	p.Arg83Profs*5	rs552296498	splice_donor	LoF	GL	-	-	ND	ND	ND	-	-	ND	3	282844	1,06E-05	3	129158	2,32E-05
UM1055	129156636	c.262T>C	p.Cys88Arg	rs373768718	nonsynonymous_SNV	missense	GL	Benign	Tolerated	ND	ND	ND	None	None	ND	1	251470	3,98E-06	1	113754	8,79E-06
UM402	129156717	c.181T>C	p.Cys61Arg	rs2307296	nonsynonymous_SNV	missense	GL	Benign	Tolerated	ND	ND	ND	None	None	ND	447	282876	1,58E-03	394	129184	2,64E-05
UM835																					
UM872																					
UM552																					
UM42	129156759	c.139G>A	p.Gly47Arg	rs755035506	nonsynonymous_SNV	missense	GL	Benign	Tolerated	ND	Yes (11.3%)**	33	None	None	ND	8	282770	2,83E-05	7	129132	5,42E-05
UM102																					

* loss of function (deleterious) mutation

‡ germline mutation

§ somatic mutation

§ absence of glycosylase activity of the recombinant protein carrying the variant

|| not determined

‡ LOH: loss of heterozygosity status (yes/no), and LOH quantification (%) corresponding to the Cancer Cell Fraction (CCF) carrying the variant as inferred from PyClone

Genome Aggregation Database v2.1, accessed 22/11/19

** benign variant in LOH but lost in tumor (retention of the wild-type allele)

- no value given due to: absence of rs for the variant (dbSNP); absence of Poly-phen2, SIFT or splice (SSF) prediction for loss-of-function or intronic variants; absence of the variant in the GnomAD database.

Supplementary Table 2: Tumor characteristics of all samples with *MBD4* variants on which Whole-Exome Sequencing was performed. This includes samples from 5 patients in the germline consecutive UM series (Germline), 3 in the tumor monosomy 3 cohort (Tumor), and 1 patient in both series (Both). The total number of mutations per exome represents all single-nucleotide variants and insertions-deletions. Percent of CpG>TpG transitions is calculated relative to the total number of SNVs.

Patient	UM Cohort	<i>MBD4</i> variant	<i>MBD4</i> mutation type	Number of mutations	Number of CpG>TpG SNVs*	Number of INDELs [‡]	Proportion of CpG>TpG SNVs*	Tumor Cellularity [§]	HyperMutated status
UM75	Germline	c.1706G>A, p.Trp569*	Germline	275	265	0	96.4%	88.1%	Yes
UM605	Germline	c.1384delG, p.Ala462Leufs*29	Germline	122	102	3	85.7%	89.6%	Yes
UM656	Germline	c.1443delT, p.Leu482Trpfs*9	Germline	181	168	0	92.8%	86.4%	Yes
UMT162	Tumor	c.541C>T, p.Arg181*	Germline	269	226	5	85.6%	79.9%	Yes
UMT45	Tumor	c.1562-1G>T, p.Asp521Profs*4	Germline	288	271	1	94.4%	79.6%	Yes
UMT61	Tumor	c.1002delTTTG, p.Lys335Phefs*18	Germline	86	53	3	63.9%	60.2%	Yes
UMT88	Both ^H	c.1402C>T, p.Arg468Trp	Somatic	243	223	2	92.5%	79.8%	Yes
UM102	Germline	c.139G>A, p.Gly47Arg	Germline	33	0	1	0.0%	94.0%	No
UM350	Germline	c.1652A>G, p.Asn551Ser	Germline	40	18	3	48.6%	26.5%	No

* single nucleotide variants

‡ insertions - deletions

^H patient common to both UM cohorts, but harboring a variant of somatic origin

[§] tumor content is inferred from Whole-Exome Sequencing using Facets.

Supplementary Table 3. Primer sequences used for *MBD4* germline screen and functional validation of variants. FW: forward primer; RV: reverse primer; hMBD4: human MBD4 protein

Purpose		FW primer sequence (5' - 3')	RV primer sequence (5' -3')
MBD4 Multiplex PCR 1 (exons 1, 3, 6 and 7)	Exon 1	CCGTGAGCTGAAGAGGTTTC	AGAAAGGCCACACACTGTC
	Exon 3 (part 1)	AAAATTTGATCCTGAACTCAATG	GTTGCAGGAGAGCAGAGGAC
	Exon 6	TCTGAAAGTGTTGCTGGTTC	AGTGGGAGACTGTGGTTTGG
	Exon 7	CACACATTTTGGGAGGGTG	GGTGGACTTATTTTGCCTCAG
MBD4 Multiplex PCR 2 (exons 2, 3, 4, 5 and 8)	Exon 2	GGTTCCTGCATTGTCATGG	GCTATGCTCCCACTACCTGC
	Exon 3 (part 2)	GGCAGCAATACAAGATGCAG	GACCCTCAGTGTGACCAGTG
	Exon 3 (part 3)	CATCATCAACACCCTCATCTTC	CAGATACCTATGGCAACATTTGG
	Exons 4-5	ATAGTGCTTGGCATGCTTTG	ATGGACTTTGAACCCAGGC
	Exon 8	TGGTATCGTAATGTACTGTCCCC	CTCTATGGCTGGAAGGTGG
Generation of mutant vectors by directed mutagenesis	p.Asn551Ser	CCGAATTTTTTGTGTCAGTGAGTGGAAGC AGGTGCACC	GGTGCACCTGCTTCCACTCACTGACACAAA AAATTCGG
	p.Arg468Trp	CATCGCTACTATATTTCTCAATTGGACCTC AGGC AAAATGGC	GCCATTTTGCCTGAGGTCCAATTGAGAAATA TAGTAGCGATG
	p.Asn467Ser	CATCGCTACTATATTTCTCAGTCGGACCT CAGGCAAAATGGC	GCCATTTTGCCTGAGGTCCGACTGAGAAAT ATAGTAGCGATG
	p.Trp569*	GACCACAAATTAAATAAATATCATGACTAG CTTTGGGAAAATCATG	CATGATTTTCCCAAGCTAGTCATGATATTT ATTTAATTTGTGGTC
Production of recombinant hMBD4 (cloning of <i>MBD4</i> cDNA into expression vector)		CGCGCGGCAGCCATATGATGGGCACGAC TGGG	GTCATGCTAGCCATATGTTAAGATAGACTTA ATTTTTCATGAT
Cloning of variants with predicted splice effect in pET01 Exontrap vector (minigene constructs)	c.1652A>G	GCAGCCCGGGGGATCGAATACCTGACAA AGCAGTGG	TAGAACTAGTGGATCACTGATCAAAAACCCC AAAACCCAC
	c.1400A>G	GCAGCCCGGGGGATCATGGACACCTCCT CGGTACAC	TAGAACTAGTGGATCAGGGTGAAGGGGGGAA TGCC
	c.1277-18T>A	GCAGCCCGGGGGATCTTTCCCAATCAGA ACAGCAA	TAGAACTAGTGGATCAGGTCCGATTGAGAA ATATAGTAGC
RT-PCR primers for fragment analysis of minigene constructs	Universal primers	5'FAM-GAGGGATCCGCTTCCTGCCCC-3'	TCCACCCAGCTCCAGTTG
	c.1652A>G	GCAGCCCGGGGGATCGAATACCTGACAA AGCAGTGG	TAGAACTAGTGGATCACTGATCAAAAACCCC AAAACCCAC
	c.1400A>G	GCAGCCCGGGGGATCATGGACACCTCCT CGGTACAC	TAGAACTAGTGGATCAGGGTGAAGGGGGGAA TGCC
	c.1277-18T>A	GCAGCCCGGGGGATCTTTCCCAATCAGA ACAGCAA	TAGAACTAGTGGATCAGGTCCGATTGAGAA ATATAGTAGC

Supplementary Table 4. Common variants detected with the targeted-sequencing pipeline fo*MBD4* variant calling, as a quality control of the sensitivity of the pooled approach. All exonic and intronic (<30bp away from the nearest exon) variants and the number of pools in which they are found are listed. The calculated variant allele frequency in the total pooled population results from the estimated variant allele count and total allele count in the UM consecutive series (1,093 individuals). For comparison, the variant allele frequency in the GnomAD European population is provided, as representative of the expected frequency for each variant.

Position on chromosome 3	Ref* allele	Alt* allele	dbSNP	Mutation type	Number of pools positive for variant allele	Cumulative VAF [§] in pool population ^H	Estimated variant allele count [§]	Total allele count	VAF in total pool population	GnomAD v2.1 Non-Finnish European population (NFE)			
										Variant allele count	Total allele count	VAF in NFE population	Fisher exact test
129,151,927	G	A	rs140697	intron_variant	124	1553.6	248	2186	0.1134	12916	129144	0.1000	0.0406
129,152,089	G	A	rs140696	synonymous_variant	123	1342.6	214	2186	0.0979	12250	129112	0.0949	0.6324
129,156,536	A	G	rs140692	intron_variant	123	1499.5	239	2186	0.1093	12241	129032	0.0949	0.0249
129,155,670	C	T	rs10342	missense_variant	112	1334.2	212	2186	0.0970	10731	129128	0.0831	0.0213
129,155,451	C	T	rs140693	missense_variant	7	53.3	9	2186	0.0041	496	129086	0.0038	0.7282
129,155,463	A	G	rs2307289	missense_variant	2	15.7	3	2186	0.0014	300	129070	0.0023	0.4992

* reference and alternative alleles
è VAF: variant allele frequency
H cumulative VAF corresponds to the sum of the VAF in each pool positive for the variant allele
§ calculated from the cumulative VAF and the expected frequency of one allele count in a given pool.

CONCLUSION TO ARTICLE 1

GERMLINE *MBD4* MUTATIONS AND PREDISPOSITION TO UVEAL MELANOMA

The main observation from this work is that *MBD4* is a UM predisposition gene that confers an almost 10-fold increased risk of developing UM, with a frequency of germline deleterious mutation (0.7%) lower but comparable to that of germline deleterious *BAP1* mutations in UM.^{304,305,625} Strikingly, other rare occurrences of *MBD4* germline mutations were described in relatively rare tumors, including a spiradenocarcinoma, a glioblastoma, a polyposis-associated colorectal adenocarcinoma, a pancreatic adenocarcinoma and an astrocytoma.^{417,553,623,626-628} Future work will be needed to characterize the cancer predisposition syndrome conferred by *MBD4* in these tumors and unravel potentially common oncogenic mechanisms. In the 3 acute myeloid leukemia (AML) patients,⁶²³ germline deleterious mutations of *MBD4* were bi-allelic and associated with an early-onset of cancer (as it is also the case for the colorectal adenocarcinoma). In UM, *MBD4* predisposition did not significantly affect age of onset, and no associated tumors were found, suggesting that *MBD4* predisposition in UM somehow differs from that of *BAP1*, probably also due to *BAP1* much higher disease penetrance. Similar to *BAP1* however, the specific association of *MBD4* predisposition with high-risk UM subgroup (associated with monosomy 3) is thought to be at least partly explained by the chromosome 3 location of *MBD4*. Like *BAP1*, malignant transformation may only occur during monosomy 3 event, following Knudson's two-hit model; however, this does not explain why *MBD4* germline mutations were exclusively found in monosomy 3 UMs, nor why so far, they are more observed in UM patients compared to other tumors.

Another important consequence of this work is its clinical relevance, as *MBD4* is now being added to oncogenic panels for diagnostic of multiple cancers, and a few *MBD4*-deficient UM patients show outlier response to immune checkpoint inhibitors,^{553,624} unlike most *MBD4*-proficient cases most likely due to their low tumor mutation burden. However, (i) *MBD4* mutations are essentially rare, and therefore the number of patients that may respond to

immunotherapy based on their hypermutator phenotype will likely be limited; (ii) furthermore, the initial outlier UM patient with *MBD4* deficiency developed secondary resistance mechanisms to her treatment, which could be due to the tumor heterogeneity conferred by *MBD4*-deficient context, highlighting the need for additional treatment options in these cases.

Since *MBD4* inactivation causes a hypermutator phenotype, future studies will be interesting to assess how similar or how different the oncogenic mechanisms in UM are compared to *MBD4* wild-type UM cases with high metastatic risk, since there is no apparent change in overall survival or metastatic free progression in *MBD4*-deficient cases compared to other monosomy 3 UMs, (although the *MBD4*-deficient group in this study was largely limited by its small size). Following-up on this work in the team, a study from Rodrigues *et al* looking at tumor evolution from primary to metastatic UM in *MBD4*-proficient and *MBD4*-deficient contexts revealed potentially driver CpG>TpG mutations of *SMARCA4*, *TP53* or *GNAS* in *MBD4*-deficient tumors,⁴¹⁷ mutations of *SMARCA4* and *TP53* having also been detected in TCGA UM and glioblastoma cases in an *MBD4*-deficient context, suggesting potentially shared mechanisms. While similar at the CNA level, Rodrigues showed that *MBD4*-deficient tumors were much more heterogeneous and instable at the SNV level than their *MBD4*-proficient counterparts, with only 1% SNVs in common between paired primary and metastatic UM, and showed a continuous process of genetic heterogeneity through acquisition of SNVs at a constant rate.⁴¹⁷ Besides, *MBD4* deficiency may mediate its oncogenic mechanism through epigenetic alterations given its role in the recognition of methylated CpG, and could potentially participate in the deregulation of methylomic patterns favorable to cancer progression, in a similar fashion to *BAP1*.³³³ Of note, while most *MBD4*-deficient patients in UM also carried a *BAP1* mutation, AML tumors with inactivated *MBD4* also had driver mutations in chromatin remodeler DNA methyltransferase *DNMT3A* and in *ASXL1*,⁶²³ the latter being part of the same multi-protein complex with *BAP1*. Recently, a study showed that deregulation and mutations of *ASXL1* promote *BAP1* stabilization, its recruitment to chromatin and pro-cancer transcriptional signature driving leukemia.⁶²⁹ In UM, considering that *MBD4*-proficient and -deficient tumors share the same tumorigenic mechanisms, the *MBD4*-inactivated context and its associated increased mutational rate may favor and accelerate the apparition of rare driver mutations in UM metastases. Since Rodrigues' study

showed that UM metastatic clones diverge rapidly from the primary tumor after the first oncogenic events (only a minority of SNVs shared between the two), *MBD4* inactivation event in UM is thought to arise early in tumorigenesis, as a “set in stone” driver event supporting Harbour and colleague’s model of a punctuated evolution in UM with almost all driver events occurring in a small lapse of time.²⁶⁶ This hypothesis would also agree with the apparent similar age of disease onset in *MBD4*-proficient and deficient contexts, where *MBD4* inactivation results in genetic tumor heterogeneity soon after clonal expansion from the primary tumor.

DIFFERENT PIGMENTATION RISK LOCI FOR HIGH-RISK MONOSOMY 3 AND LOW-RISK DISOMY 3 UVEAL MELANOMAS

The main objective of this PhD was to identify and functionally characterize genetic predisposition factors in UM. Since UM is strongly associated with individuals of European ancestry and with individuals with light eye and skin color, we hypothesized that some UM risk alleles may be present in populations of European ancestry, or conversely, protective alleles in Asian and African populations. This came from the following observations: (i) the stable UM incidence over the past decades (unlike the rise in cutaneous melanoma) despite increased exposure to sunlight, the low tumor mutation burden in UM and the lack of ultra-violet (UV) mutational signature (except for iris melanoma) argue against a role for pigmentation protecting against UV radiation to explain UM epidemiology; (ii) familial forms of UM occur in 1% of cases, yet *BAP1* is the only known strongly predisposing UM gene and only explains a fraction of these familial occurrences, suggesting that other genetic factors are implicated in UM risk.

Aiming to identify additional genetic risk factors in UM, the team thus conducted a first genome-wide association study (GWAS) in UM in 2017, on 259 UM patients and 401 controls of European ancestry and identified a UM susceptibility region on chromosome 5 *TERT/CLPTM1L* locus.⁵²⁶ This complex locus has been shown to confer cancer risk in multiple tumor types; we aimed to functionally characterize this region and its biological implication in UM in a 3rd part of this work (Article 3). However, this locus could not explain the UM proneness in individuals of European ancestry as the frequency of its risk alleles were comparable to that in African and Asian populations. Aiming to identify additional UM susceptibility regions, we obtained germline DNA from additional UM patients treated at Curie Institute from 2013 to 2018, and European controls, and combined the previous GWAS dataset with these new datasets to perform a second phase of GWAS with increased statistical power on 1,142 UM patients and 882 European controls. This second article, now accepted for publication in the *Journal of the National Cancer Institute*, describes our findings from this second GWAS, in which the large sample size also allowed us to perform subgroup analysis of UMs at high (monosomy 3) or low (disomy 3) metastatic risk.

ARTICLE 2: DIFFERENT PIGMENTATION RISK LOCI FOR HIGH-RISK MONOSOMY 3 AND LOW-RISK DISOMY 3 UVEAL MELANOMAS

Abstract

Background: Uveal melanoma (UM), a rare malignant tumor of the eye, is predominantly observed in populations of European ancestry. UMs carrying a monosomy 3 (M3) frequently relapse mainly in the liver, whereas UMs with disomy 3 (D3) are associated with more favorable outcome. Here, we explored the UM genetic predisposition factors in a large genome-wide association study (GWAS) of 1,142 European UM patients and 882 healthy controls.

Methods: We combined two independent datasets (GSA array) with the dataset described in a previously published GWAS in UM (Omni5 array), which were imputed separately and subsequently merged. Patients were stratified according to their chromosome 3 status and identified UM risk loci were tested for differential association with M3 or D3 subgroups. All statistical tests were two-sided.

Results: We recapitulated the previously identified risk locus on chromosome 5 on CLPTM1L (rs421284: odds ratio [OR] = 1.58, 95% confidence interval [CI] = 1.35-1.86; $P = 1.98 \times 10^{-8}$) and identified two additional risk loci involved in eye pigmentation: IRF4 locus on chromosome 6 (rs12203592: OR = 1.76, 95% CI = 1.44-2.16; $P = 3.55 \times 10^{-8}$) and HERC locus on chromosome 15 (rs12913832: OR = 0.57, 95% CI = 0.48-0.67; $P = 1.88 \times 10^{-11}$). The IRF4 rs12203592 SNP was found to be exclusively associated with risk for the D3 UM subtype (ORD3 = 2.73, 95% CI = 1.87-3.97; $P = 1.78 \times 10^{-7}$), and the HERC2 rs12913832 SNP was exclusively associated with risk for the M3 UM subtype (ORM3 = 2.43, 95% CI = 1.79-3.29; $P = 1.13 \times 10^{-8}$). However, the CLPTM1L risk locus was equally statistically significant in both subgroups.

Conclusion: This work identified two additional UM risk loci known for their role in pigmentation. Importantly, we demonstrate that UM tumor biology and metastatic potential are influenced by patients' genetic backgrounds.

Keywords: GWAS; pigmentation; uveal melanoma.

© The Author(s) 2021. Published by Oxford University Press.

Different Pigmentation Risk Loci for High-Risk Monosomy 3 and Low-Risk Disomy 3 Uveal Melanomas

Lenha Mobuchon , PhD,^{1,†} Anne-Céline Derrien , MSc,^{1,†} Alexandre Houy , MSc,^{1,†} Thibault Verrier , BSc,¹ Gaëlle Pierron , PhD,² Nathalie Cassoux, MD, PhD,^{3,4} Maud Milder, MSc,⁵ Jean-François Deleuze , PhD,⁶ Anne Boland , PhD,⁶ Ghislaine Scelo, PhD,^{7,8} Géraldine Cancel-Tassin , PhD,^{9,10} Olivier Cussenot, MD, PhD,^{9,10} Manuel Rodrigues , MD, PhD,^{1,11} Josselin Noirel , PhD,¹² Mitchell J. Machiela , PhD,¹³ Marc-Henri Stern , MD, PhD^{1,*}

¹Inserm U830, DNA Repair and Uveal Melanoma (D.R.U.M), Equipe Labellisée par la Ligue Nationale Contre le Cancer, Institut Curie, PSL Research University, Paris, France; ²Somatic Genetic Unit, Department of Genetics, Institut Curie, PSL Research University, Paris, France; ³Department of Ocular Oncology, Institut Curie, Paris, France; ⁴Faculty of Medicine, University of Paris Descartes, Paris, France; ⁵Inserm CIC BT 1418, Institut Curie, PSL Research University, Paris, France; ⁶Université Paris-Saclay, CEA, Centre National de Recherche en Génomique Humaine, Evry, France; ⁷International Agency for Research on Cancer (IARC), Lyon, France; ⁸Cancer Epidemiology Unit, Department of Medical Sciences, University of Turin, Turin, Italy; ⁹CeRePP, Tenon Hospital, Paris, France; ¹⁰Sorbonne University, GRC n°5 Predictive Onco-Urology, AP-HP, Tenon Hospital, Paris, France; ¹¹Department of Medical Oncology, Institut Curie, PSL Research University, Paris, France; ¹²Laboratoire GBCM (EA7528), CNAM, HESAM Université, Paris, France and ¹³Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD 20892, USA

[†]These authors contributed equally to this work.

*Correspondence to: Marc-Henri Stern, MD, PhD, Institut Curie, Inserm U830, 26 rue d'Ulm, 75248 Paris cedex 05, France (e-mail: marc-henri.stern@curie.fr).

Abstract

Background: Uveal melanoma (UM), a rare malignant tumor of the eye, is predominantly observed in populations of European ancestry. UMs carrying a monosomy 3 (M3) frequently relapse mainly in the liver, whereas UMs with disomy 3 (D3) are associated with more favorable outcome. Here, we explored the UM genetic predisposition factors in a large genome-wide association study (GWAS) of 1142 European UM patients and 882 healthy controls. **Methods:** We combined 2 independent datasets (Global Screening Array) with the dataset described in a previously published GWAS in UM (Omni5 array), which were imputed separately and subsequently merged. Patients were stratified according to their chromosome 3 status, and identified UM risk loci were tested for differential association with M3 or D3 subgroups. All statistical tests were 2-sided. **Results:** We recapitulated the previously identified risk locus on chromosome 5 on *CLPTM1L* (rs421284: odds ratio [OR] = 1.58, 95% confidence interval [CI] = 1.35 to 1.86; $P = 1.98 \times 10^{-8}$) and identified 2 additional risk loci involved in eye pigmentation: *IRF4* locus on chromosome 6 (rs12203592: OR = 1.76, 95% CI = 1.44 to 2.16; $P = 3.55 \times 10^{-8}$) and *HERC2* locus on chromosome 15 (rs12913832: OR = 0.57, 95% CI = 0.48 to 0.67; $P = 1.88 \times 10^{-11}$). The *IRF4* rs12203592 single-nucleotide polymorphism was found to be exclusively associated with risk for the D3 UM subtype (OR_{D3} = 2.73, 95% CI = 1.87 to 3.97; $P = 1.78 \times 10^{-7}$), and the *HERC2* rs12913832 single-nucleotide polymorphism was exclusively associated with risk for the M3 UM subtype (OR_{M3} = 2.43, 95% CI = 1.79 to 3.29; $P = 1.13 \times 10^{-8}$). However, the *CLPTM1L* risk locus was equally statistically significant in both subgroups. **Conclusions:** This work identified 2 additional UM risk loci known for their role in pigmentation. Importantly, we demonstrate that UM tumor biology and metastatic potential are influenced by patients' genetic backgrounds.

Uveal melanoma (UM) arises from melanocytes in the uveal tract of the eye, including the choroid and, more rarely, ciliary body and iris. Prognosis is dismal when the disease spreads, frequently metastasizing to the liver (1). Loss of chromosome 3 and gain of chromosome 8 are associated with a higher risk of metastatic relapse (2,3). Monosomy 3 (M3) UMs are associated with BAP1 (3p21) mutations and a high risk of metastases (4). Conversely, disomy 3 (D3)

tumors carry SF3B1 or EIF1AX mutations (5-7) and are associated with late metastases and a better prognosis. These M3 and D3 subtypes are different not only in terms of mutational statuses but also at the cytogenetic, miRNome, methylome, and proteome levels, suggesting that they derive from 2 tumorigenic processes (8).

UM mainly affects populations of European ancestry, with a 10-fold lower incidence in individuals of African American or

Received: 4 May 2021; Revised: 6 July 2021; Accepted: 19 August 2021

© The Author(s) 2021. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Asian Pacific Islander ancestry (9,10). Fair skin and blue-gray eyes are also risk factors for UM (11). With the hypothesis that higher frequency of risk alleles exists in populations of European ancestry to explain UM epidemiology, we performed the first genome-wide association study (GWAS) in UM and identified rs421284 as the leading single-nucleotide polymorphism (SNP) on the *CLPTM1L/TERT* risk locus on chromosome 5p15.33. Moreover, a trend for association between variants in *OCA2* and UM was also observed (12). Recently, another UM GWAS identified 11 loci with a *P* value of association less than 10^{-5} , but none reached statistical significance (13).

The *CLPTM1L* risk allele identified by our first UM GWAS had a higher frequency in individuals of African American ancestry compared with Europeans and thus could not explain the peculiar prevalence of UM in individuals of European ancestry (12). To identify additional UM risk loci in the European population, we increased the power of our GWAS by performing genome-wide genetic imputation and by accruing 1142 UM patients and 882 controls, a threefold increase of our first study, allowing subgroup analysis depending on chromosome 3 status.

Methods

Study Populations

This study was approved by the ethical committee and internal review board at the Institut Curie. Blood samples were obtained from 946 UM patients who consented to participate in the study and from 496 control individuals of French origin from the KIDRISK consortium (US NCI U01CA155309; G. Scelo). Genotypes obtained on the Infinium Global Screening Array 24 v1.0 were called using default parameters in GenomeStudio (Illumina).

Genotyping, Imputation, and Merge

Genotypes from the previously published GWAS (dataset1) (12) and for the 2 new sets (dataset2 and dataset3) were filtered (Supplementary Methods, available online) and independently imputed on the Michigan Imputation Server using Eagle for the phasing and Haplotype Reference Consortium r1.1 as the reference dataset. Imputed datasets were merged together, and another quality control was performed (Supplementary Table 1, available online). Manual genotyping was also performed on selected SNPs and individuals (Supplementary Methods, available online). Patients and controls of European ancestry were stringently selected for further analyses (Supplementary Methods and Supplementary Figures 1 and 2, available online).

Statistical Analysis

For GWAS, firth logistic regression was performed using plink2 with covariates described in the Supplementary Methods (available online). An exact number of patients and controls used are indicated in the respective figures and tables for each analysis. Association of SNPs with UM risk was determined by odds ratios (ORs) with 95% confidence intervals (CIs), and SNPs with a *P* value less than 5.00×10^{-8} were considered to be statistically significant, and those with *P* value less than 1.00×10^{-5} only reached the tendency line. Eye color was predicted using IrisPlex tools (<https://hirisplex.erasmusmc.nl/>). Association of eye color with UM risk was calculated using a 2-sided Fisher test *P* value and odds ratio. Comparison of variant allele frequency (VAF) of SNPs in different populations were tested for statistical

significance using a 2-sided Fisher test *P* value. Expression quantitative trait loci (eQTL) were performed using linear regression. A *P* value of less than .05 was considered statistically significant for all tests other than GWAS firth logistic regression.

Results

Genome-Wide Association Study in UM

We combined 2 independent datasets (dataset2: 369 UM and 496 controls; dataset3: 577 UM, Global Screening Array) with that of our previous UM GWAS (dataset1 of 271 UM and 429 controls; Omni5 array) (12). The data were quality filtered (Figure 1; Supplementary Table 1, available online). The 3 datasets were imputed separately using the Haplotype Reference Consortium on the Michigan server and subsequently merged. Quality of the genotyping and imputation was further assessed by TaqMan genotyping on rs421284, rs12203592, and rs12913832 SNPs on 972 selected samples, with 95.2%, 99.1%, and 99.6% of good match, respectively (Supplementary Table 2, available online). Data from individuals of European ancestry were stringently selected from principal component analyses (PCA) using plink2 in which the first 2 principal components were used. Outliers were then excluded from those selected samples using SmartPCA with 10 iterative PCAs (Supplementary Figures 1-3, available online). The final dataset for the UM GWAS analysis consisted of 7 488 175 SNPs in 1142 patients and 882 controls (Figure 1).

The GWAS Manhattan plot showed 3 distinct loci reaching genome-wide significance (firth logistic regression $P < 5.00 \times 10^{-8}$) (chr5, *CLPTM1L/TERT* locus; chr6, *IRF4* locus; and chr15, *HERC2/OCA2* locus) (Figure 2; Supplementary Table 3, available online). Within the *HERC2/OCA2* locus, 8 SNPs in high linkage disequilibrium reached statistical significance. The most statistically significant SNPs at this locus were rs1129038 and rs12913832 (OR = 0.56, 95% CI = 0.48 to 0.66; $P = 5.97 \times 10^{-12}$; and OR = 0.57, 95% CI = 0.48 to 0.67; $P = 1.88 \times 10^{-11}$, respectively), located in *HERC2*. A single SNP located in *IRF4* was found to be well above the genome-wide significance: rs12203592 (OR = 1.76, 95% CI = 1.44 to 2.16; $P = 3.55 \times 10^{-8}$). Finally, the association study recapitulated the previously identified 5p15.33 risk locus (*TERT/CLPTM1L*) (12), with several SNPs in high linkage disequilibrium ($r^2 > 0.9$) reaching statistical significance (Supplementary Table 3, available online). The most statistically significant SNP was rs370348 (OR = 1.59, 95% CI = 1.35 to 1.86; $P = 1.48 \times 10^{-8}$). The leading risk SNP in our first GWAS, rs421284 (12), also showed high statistical significance (OR = 1.58, 95% CI = 1.35 to 1.86; $P = 1.98 \times 10^{-8}$) and was further analyzed in this study. A few other loci showed suggestive evidence for an association with UM but did not reach genome-wide significance ($P < 5.00 \times 10^{-8}$) (Supplementary Table 3, available online and Figure 2).

Conditional analyses enable the detection of secondary independent association signals within a genomic locus by conditioning on the primary associated SNP at the locus. At the *CLPTM1L*, *IRF4*, and *HERC2* loci, no other statistically significant SNP was found to be independently associated with UM when conditioning on rs421284, rs12203592, or rs12913832, respectively. Moreover, these 3 conditional analyses did not reveal any statistically significant regions other than *CLPTM1L*, *IRF4*, and *HERC2* (Supplementary Figure 4, available online).

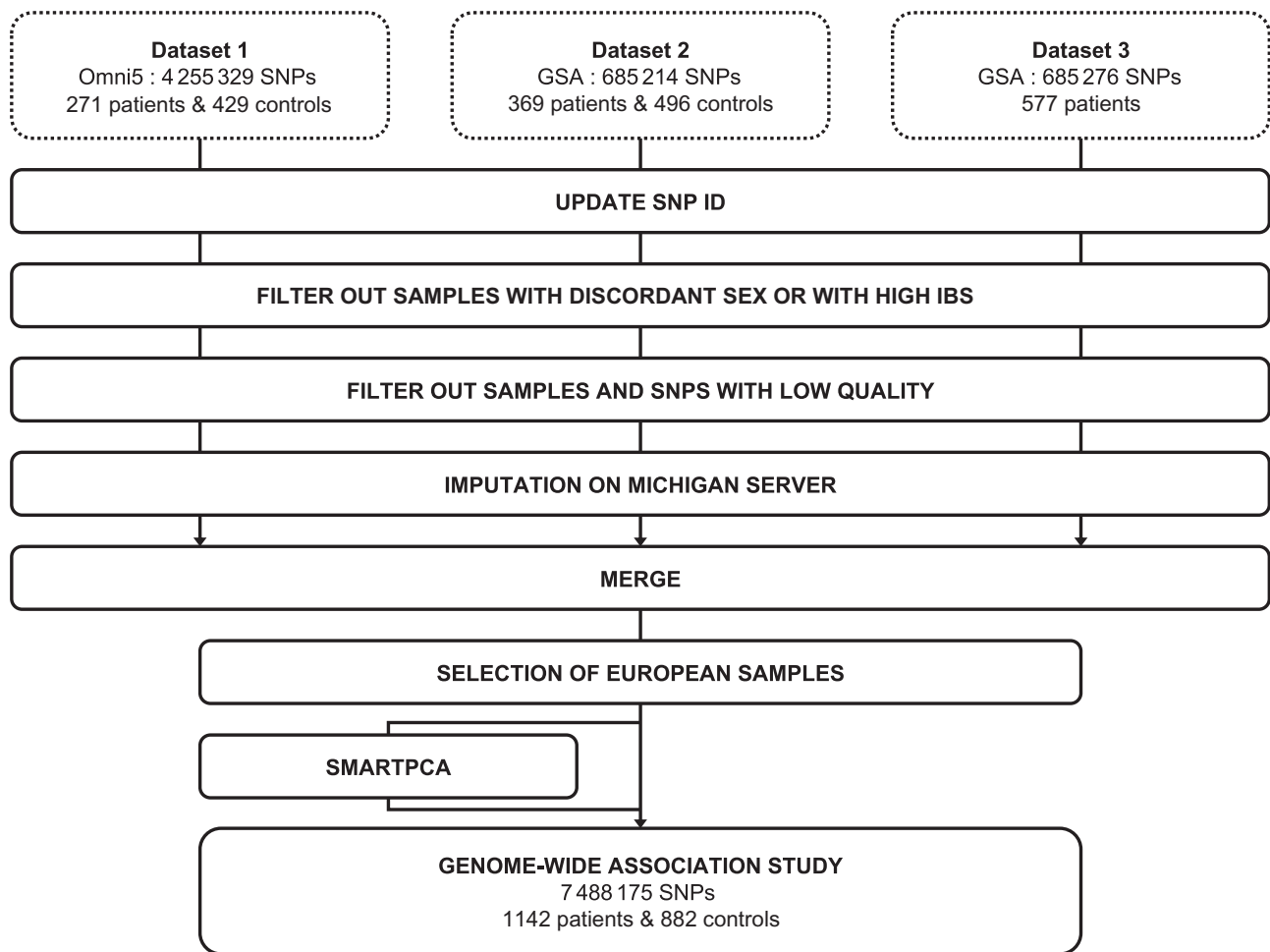


Figure 1. Files and pipeline used for the filtering and imputation of the Genome-Wide Association Study in uveal melanoma. GSA = Global Screening Array; ID = identification; SNP = single nucleotide polymorphism; IBS = Identify By State.

UM Risk Loci and Pigmentation

To evaluate the impact of risk SNPs on gene regulation, eQTL analyses were performed for the statistically significant loci using expression data from tumors of an in-house series of 73 UMs (14). We previously identified an association between *CLPTM1L* expression and rs421284 with higher expression of *CLPTM1L* in individuals carrying the risk allele (C) (12). Interestingly, the other 2 major risk loci identified in this association study, *IRF4* and *HERC2*, are known to be strongly implicated in the regulation of the pigmentation pathways determining eye and skin colors (15–17), prompting us to further investigate the expression of pigmentation genes in UM. *IRF4* expression was found to be strongly associated with rs12203592 alleles, with a decreased expression in tumors carrying the risk TT genotype (linear regression $P = 2.00 \times 10^{-6}$; Supplementary Figure 5, A, available online). Looking at eQTLs in the Genotype-Tissue Expression database, rs12203592 is linked to *IRF4* expression in most tissues, but the directionality of the association varies. As in UM, sun-exposed skin had a lower *IRF4* expression linked to the T allele, whereas a lower expression of *IRF4* is associated with the C allele in all other tissues, suggesting a tissue-specific regulation for this gene (Supplementary Figure 5, B, available online). At the *HERC2* locus, no correlation was found between rs12913832 alleles and expression of this gene in UM (Supplementary Figure 6, A, available online), in contrast to

whole blood, where there is a statistically significant decrease in *HERC2* expression associated with the G allele (Supplementary Figure 6, B, available online). However, expression of *OCA2*, a nearby gene known to be regulated by *HERC2* in melanocytes (17), was found with a highly statistically significant association with rs12913832 genotypes ($P = 9.08 \times 10^{-4}$) in UM, with decreased expression for tumors carrying the risk G allele (Supplementary Figure 6, C, available online).

Our finding of 2 major pigmentation loci is in accordance with the high prevalence of light eye color in UM patients of European ancestry (11). We investigated whether the risk of developing UM conferred by the risk alleles of *HERC2* and *IRF4* was fully linked to their determining role in eye pigmentation. We thus predicted the eye color of all UM and control individuals included in this study, using the algorithm developed in the IrisPlex System, based on the genotype combination of 6 SNPs (*HERC2* rs12913832, *OCA2* rs1800407, *SLC45A2* rs16891982, *TYR* rs1393350, *IRF4* rs12203592, and *LOC105370627*: intron variant) (18). We predicted the eye color of UM patients and controls to be brown (41.6% of patients vs 60.1% of controls, respectively), green (1.7% vs 1.1%), or blue (56.7% vs 38.9%), allowing us to confirm the statistically significant association of blue eye color (vs other eye colors) with UM risk (OR = 2.07, 95% CI = 1.72 to 2.49; 2-sided Fisher test $P = 1.21 \times 10^{-15}$) (Figure 3, A and B), confirming the recent study by Jager and colleagues (19). Strikingly, when we added eye color

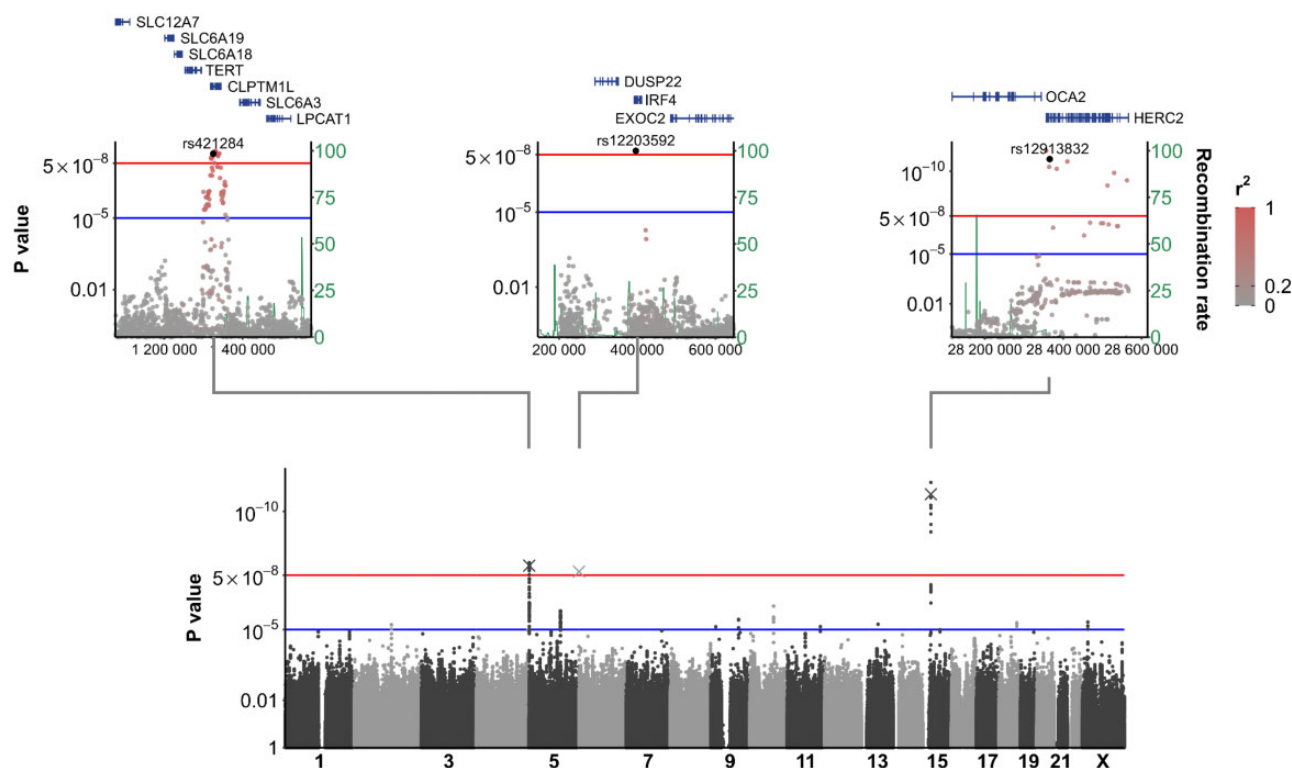


Figure 2. Manhattan plot and regional linkage disequilibrium plot for statistically significant loci. For the Manhattan plot, the association test P value (y-axis) is plotted against its physical chromosomal position (x-axis). Chromosomes are shown in alternating black and grey. SNPs above the **top horizontal line** represent those with a $P < 5.00 \times 10^{-8}$ and were considered to be statistically significantly associated with uveal melanoma. The **bottom horizontal line** represents the tendency line ($P < 1.00 \times 10^{-5}$). Statistical significance was measured using unconditional logistic regressions. For regional locus plots, genes are depicted with **rectangles** and SNPs are represented by **dots**. Shading of dots reflects the level of linkage disequilibrium (r^2) with the highlighted SNP of interest (**black circle** with rs number indicated). **Vertical bars** indicate recombination rates in human population. CLPTM1L = cleft lip and palate transmembrane protein 1-like; DUSP22 = dual specificity phosphatase 22; EXOC2 = exocyst complex component 2; HERC2 = HECT and RLD domain containing E3 ubiquitin protein ligase 2; IRF4 = interferon regulatory factor 4; LPCAT1 = lysophosphatidylcholine acyltransferase 1; SLC12A7 = solute carrier family 12 member 7; SCL6A18 = solute carrier family 6 member 18; SCL6A19 = solute carrier family 6 member 19; SCL6A3 = solute carrier family 6 member 3; OCA2 = oculocutaneous albinism II; TERT = telomerase reverse transcriptase.

(determined by the IrisPlex System) as a covariate in the association analysis, the resulting odds ratio remained unchanged for IRF4 (OR = 1.76, 95% CI = 1.44 to 2.16; firth logistic regression $P = 3.55 \times 10^{-8}$ without eye color covariate, vs OR = 1.76, 95% CI = 1.43 to 2.17; $P = 9.25 \times 10^{-8}$, with eye color covariate; **Figure 3, C**; **Supplementary Table 4**, available online). Conversely, the odds ratio of HERC2 risk SNP rs12913832 lost statistical significance with eye color covariate (OR = 0.57, 95% CI = 0.48 to 0.67; $P = 1.88 \times 10^{-11}$, without eye color covariate, vs OR = 0.76, 95% CI = 0.57 to 1.02; $P = 0.06$, with eye color covariate), in accordance with the major role of rs12913832 in the determination of eye pigmentation (17,18). As expected, the odds ratio of CLPTM1L, a gene with no known role in pigmentation, remained unchanged (rs421284: OR = 1.58, 95% CI = 1.35 to 1.86; $P = 1.98 \times 10^{-8}$, without eye color covariate vs OR = 1.58, 95% CI = 1.34 to 1.86; $P = 4.01 \times 10^{-8}$, with eye color covariate; **Figure 3, C**; **Supplementary Table 4**, available online). This indicates that the implication of the IRF4 locus in UM risk not only is explained by the prevalence of UM among individuals with light eye color but also points toward another role for this risk locus beyond pigmentation.

Pigmentation Risk Loci and UM Epidemiology

The higher prevalence of UM among individuals of European ancestry strongly supports the existence of inherited risk alleles

for the disease. The TERT/CLPTM1L risk locus does not account for this population bias, as the risk haplotype is more frequent in African American populations than those of European ancestry (rs421284: VAF = 0.597 vs 0.429, respectively) (**Supplementary Table 5**, available online; Genome Aggregation Database v2.1). However, the risk haplotypes of both IRF4 and HERC2 are found at statistically significantly higher frequencies in populations of non-Finnish European ancestry (NFE) than in those of African or African American and East Asian origins (populations defined by Genome Aggregation Database) (IRF4 rs12203592: VAF = 0.144, 0.034, and 0.000, respectively; HERC2 rs12913832: VAF = 0.803, 0.125, and 0.001, respectively; 2-sided Fisher test $P < 1.00 \times 10^{-20}$ for all statistical comparisons of NFE vs East Asian and NFE vs African and African or African American). Therefore, the higher frequency of the risk alleles of these 2 pigmentation loci may at least partly explain the higher prevalence of UM in European populations.

Association Study for the Two Major UM Subtypes

Loss of chromosome 3 is the strongest factor associated with poor metastatic outcome in UM and correlates with increased mortality (2,3). The genomic status was available for 384 UM patients, allowing us to test for differential association of UM risk loci according to chromosome 3 status. Association studies

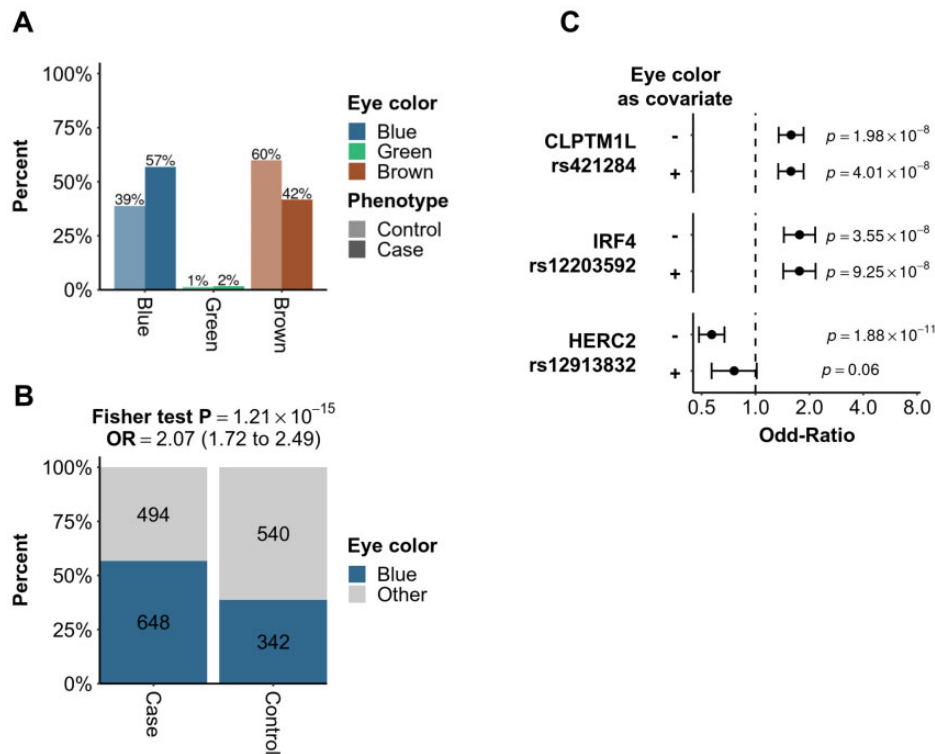


Figure 3. Eye pigmentation and uveal melanoma risk. **A)** Proportion of blue, green, and brown eye colors among uveal melanoma (UM) patients (dark shade) and controls (light shade), as predicted by the IrisPlex System (18). **B)** Proportion of blue eyes vs other eye colors in UM patients and controls. The number of individuals is indicated. The association of blue eye color with UM risk is indicated by the Fisher test P value and odds ratio (OR). The 95% confidence interval for the odds ratio is indicated within brackets. **C)** Effect of eye color as a GWAS covariate on the odds ratio for the 3 main SNPs of statistically significant UM risk loci (CLPTM1L, IRF4, and HERC2). The error bars indicate the 95% confidence intervals for the odds ratio. Statistical significance was assessed using a 2-sided Fisher test. The + and - indicate the inclusion or exclusion of eye color as a GWAS covariate, respectively. For each SNP and in both covariate conditions, association with UM risk is represented by the odds ratio (x-axis) and associated P value. The vertical dotted line is set at odds ratio = 1.00, indicating an absence of association with UM. All statistical tests were 2-sided. CLPTM1L = cleft lip and palate transmembrane protein 1-like; HERC2 = HECT and RLD domain containing E3 ubiquitin protein ligase 2; IRF4 = interferon regulatory factor 4.

were performed independently on UMs with D3 or M3 (246 M3 and 138 D3) vs controls (CTL), for the most statistically significant SNP of each risk locus identified by GWAS (Table 1). Interestingly, rs12203592 (IRF4 locus) showed a strong association with D3 UM, using a logistic regression model ($OR_{D3vsCTL} = 2.73$, 95% CI = 1.87 to 3.97; $P = 1.78 \times 10^{-7}$), whereas the association vanished completely in M3 UM ($OR_{M3vsCTL} = 1.01$, 95% CI = 0.7 to 1.47; $P = .95$). On the contrary, rs12913832 (HERC2 locus) showed a statistically significant high association with M3 UM but not with D3 UM ($OR_{M3vsCTL} = 2.43$, 95% CI = 1.79 to 3.29; $P = 1.13 \times 10^{-8}$; $OR_{D3vsCTL} = 1.10$, 95% CI = 0.80 to 1.52; $P = .56$). As for rs421284 (CLPTM1L locus), no preferential association was found in either UM subgroup ($OR_{D3vsCTL} = 2.26$, 95% CI = 1.61 to 3.17; $P = 2.64 \times 10^{-6}$; $OR_{M3vsCTL} = 1.55$, 95% CI = 1.18 to 2.03; $P = .001$) (Table 1). To further assess the statistical significance of the observed differential association of rs12203592 in M3 and D3, we compared both subgroups (OR_{M3vsD3}) for their association with UM risk SNPs (Supplementary Table 6, available online). As expected, the odds ratio of CLPTM1L rs421284 with M3 UMs or D3 UMs collapsed toward the value 1, indicating that this SNP was similarly associated with both subgroups ($OR_{M3vsD3} = 0.86$, 95% CI = 0.67 to 1.11; $P = .33$). Conversely, the low odds ratio M3 vs D3 and statistically significant P value obtained for IRF4 rs12203592 ($OR_{M3vsD3} = 0.38$, 95% CI = 0.27 to 0.52; $P = 8.46 \times 10^{-7}$) and the high odds ratio M3 vs D3 for HERC2 rs12913832 ($OR_{M3vsD3} = 1.81$, 95% CI = 1.38 to 2.38; $P = 3.87 \times$

10^{-4}) recapitulated the specific association of these risk regions for D3 UM and M3 UM, respectively.

These data strongly suggest that UM tumor biology is influenced by the genetic background predisposing to UM, with CLPTM1L SNPs predisposing to all UM types, IRF4 SNP predisposing specifically to risk in D3 UM, and HERC2 locus to risk in M3 UM.

Discussion

We extended our initial UM GWAS by including 1142 UM patients and performing genome-wide genotype imputation. This allowed us to recapitulate the previously described CLPTM1L risk locus and to further identify IRF4 and HERC2, 2 pigmentation loci, as UM genetic risk factors. Furthermore, we demonstrated that whereas CLPTM1L is a risk locus in all UM subgroups, IRF4 is specifically associated with D3 UM and HERC2 specifically with M3 UM.

The TERT/CLPTM1L region has frequently been associated in GWAS studies, with higher and lower tumor risk depending on cancer types (20). The function of CLPTM1L is not yet fully understood, but this protein is thought to contribute to RAS-dependent transformation and tumorigenesis, including in pancreatic tumorigenesis (21-23). On the other hand, TERT (on the same locus) plays a major role in telomere maintenance (24). In a previous study, we revealed a correlation between rs421284

Table 1. Main risk loci in uveal melanoma according to their chromosome 3 status

ID ^a	SNP ^b	Symbol	Alternative allele	Monosomy 3			Disomy 3		
				Total No. (patients/controls)	OR (95% CI)	P ^c	Total No. (patients/controls)	OR (95% CI)	P ^c
5:1325590: T > C	rs421284	CLPTM1L	C	1126 (244/882)	1.55 (1.18 to 2.03)	0.001	1018 (137/881)	2.26 (1.61 to 3.17)	2.64×10^{-6}
6:396321: C > T	rs12203592	IRF4	T	1126 (244/882)	1.01 (0.70 to 1.47)	0.95	1018 (137/881)	2.73 (1.87 to 3.97)	1.78×10^{-7}
15:28365618: A > G	rs12913832	HERC2	G	1126 (244/882)	2.43 (1.79 to 3.29)	1.13×10^{-8}	1018 (137/881)	1.10 (0.80 to 1.52)	0.56

^aID refers to chromosome number: chromosomal genomic position: reference allele > alternative allele, based on genome build GRCh37 (hg19). CI = confidence interval; OR = odds ratio.

^bSNP = single nucleotide polymorphism, according to the Single Nucleotide Polymorphism Database.

^cTwo-sided P values were calculated by general linear model.

genotype and CLPTM1L expression but not TERT, the latter being poorly expressed in UMs (12). Whether CLPTM1L or TERT is the target of this risk haplotype in UM tumorigenesis is still unclear.

We confirmed the association of the OCA2/HERC2 locus with UM risk, initially identified as candidate SNPs by Ferguson et al. (25). We confirmed the correlation between HERC2 rs12913832 and OCA2 expression in UM, with a decreased expression in individuals carrying the G allele (Supplementary Figure 6, C, available online). HERC2 is known to regulate the expression of OCA2, which codes for a protein involved in determining the melanin type and amount (26). These 2 genes are the main genetic determinants of iris color (18). In melanocytic cell lines, the transcription factor HLTF binds to the A but not the G allele of rs12913832, creating an activating loop for OCA2 transcription by the recruitment of MITF and LEF1 (17,27). The rs12913832 A allele is consequently associated with high expression of OCA2, production of melanin, brown eye color, and low UM risk, and conversely for the rs12913832 G allele.

The third UM risk locus identified in the present study is characterized by a single risk SNP on IRF4, rs12203592 (25). IRF4 regulates the expression of key pigmentation genes in association with MITF, including TYR involved in the production of melanin. The IRF4 locus is also associated with melanocytic naevus count, freckling, and tanning ability (28-30). TFAP2 α recognizes rs12203592 C allele in melanocytes, allowing the recruitment of MITF, YY1, and potentially LEF1 and increasing IRF4 expression (15,16). Conversely, rs12203592 T allele prevents TFAP2 α binding resulting in lower IRF4 expression. We showed that the rs12203592 UM risk allele T is associated with a dramatic decreased expression of IRF4 (Supplementary Figure 5, A, available online). Of note, only a minority of individuals (3 in our in-house series) carry the TT genotype. A similar eQTL pattern was reported in sun-exposed skin from Genotype-Tissue Expression, whereas an opposite direction was found in other tissues (Supplementary Figure 5, B, available online), strongly suggesting that IRF4 is regulated in a tissue-specific manner.

The present GWAS demonstrates the role of 2 pigmentation genes in the genetic risk of UM, in addition to the CLPTM1L/TERT risk locus. This is consistent with light iris color being a risk factor for UM (OR = 1.75) (11,19,31) similar to our finding (OR = 2.07). Iris pigmentation depends on the production and maturation of melanin as well as on the ratio of the 2 types of melanin: eumelanin (black-brown, densely packed) and pheomelanin (yellow-to-red, loosely packed). Melanin plays a major role in protecting against ultraviolet radiation (UVR) by absorbing free radicals and inhibiting UV-mediated damage (32). Pheomelanin, however, can also induce more oxidative damage on UVR than eumelanin (33), which was proposed to explain the contribution of light iris color in UM (34). However, the steady UM incidence despite increased UVR exposure, the low tumor mutation burden, and absence of UVR mutational signature in UM tumors ruled out this hypothesis (5,35). Interestingly, iris melanoma, a rare form of UM, is associated with high tumor mutation burden and a UVR signature (36), consistent with iris color being a risk factor for iris melanoma (37). However, our GWAS is restricted to choroid melanoma, a tissue that, unlike the iris, is not directly exposed to sunlight. In this respect, IRF4 and potentially HERC2/OCA2 SNPs may play a role outside from iris pigmentation to explain UM risk. However, a limitation of our study is that eye pigmentation is deduced from genotypes, which are also risk SNPs for UM, making it challenging to derive causal statements.

Status of chromosome 3 and BAP1 delineates 2 UM subtypes, M3 and BAP1-inactivated high-risk tumors and D3 and wild-

type BAP1 low-risk tumors (2-4,8). Strikingly, whereas CLPTM1L region confers similar susceptibility for M3 UM and D3 UM, we show that the risk for M3 UM is associated with the OCA2/HERC2 region and D3 UM with the IRF4 locus. How these processes influence the malignant transformation is unknown but most probably independent of the protective role of melanin against UVR. Furthermore, our data reinforce the idea that UM encompasses at least 2 diseases, with distinct clinicobiological characteristics (8,38-40) and distinct susceptibility loci.

Further studies should investigate the molecular mechanisms behind these UM genetic susceptibility loci to understand the role of pigmentation genes in UM risk. This study provides important insights in the genetics of UM and may lead to improvements in risk prediction and to a better understanding of the biological basis of UM.

Funding

A-CD was supported by the Horizon 2020 program and innovation program under the Marie Skłodowska-Curie grant agreement No 666003 and the Ligue Nationale Contre le Cancer. LM was supported by the Horizon 2020 program UM Cure (No 667787). This study was funded by the INCa Programme de recherche sur le cancer en Sciences Humaines et Sociales, Epidémiologie et Santé Publique (2017-1-PL SHS-02), the Horizon 2020 program UM Cure (No 667787), the Institut National de la Santé et de la Recherche Médicale (INSERM), the Institut Curie, the Ligue Nationale Contre le Cancer (Labellisation), and the Site de Recherche Intégrée sur le Cancer (SiRIC2) de l'Institut Curie. Genome-wide genotyping of the French controls including in the KIDRISK study was funded by the US National Institutes of Health (NIH), National Cancer Institute (U01CA155309) through the IARC-2 scan.

Notes

Role of the funders: The funders had no role in the design of the study; the collection, analysis, and interpretation of the data; the writing of the manuscript; and the decision to submit the manuscript for publication.

Disclosures: The authors have no conflict of interest to declare.

Author contributions: LM, A-CD, and AH contributed equally to this study. LM, A-CD, AH, and M-HS conceptualized the study and developed its methodology. G C-T, OC, and GS provided resources (GWAS control samples). A-CD, LM, AB, and J-FD, conducted research investigation (experiments). LM, AH, TV, and JN performed data curation and formal analysis. GP, NC, and MM provided resources. A-CD, AH, and LM conducted experiments, performed visualization/data presentation, and wrote and edited the manuscript. GC, MR, JN, and MR reviewed the manuscript. M-HS supervised the study. All authors reviewed and approved the final manuscript.

Acknowledgements: The authors are grateful to the patients for accepting to participate in this study. The authors would like to thank Antoine Chouteau, Manon Reverdy, Marion Gautier, and Khadija Ait Rais for their help in managing samples and are grateful for the Biological Resource Center of the Institut Curie and its members for preparing patients' DNAs.

Data Availability

Dataset2 and 3 genotyping data used in the analysis have been deposited and are available on the European Genome-Phenome Archive (EGA) (<https://ega-archive.org/>) under accession number EGAS00001005200. Previously published genotyping of dataset1 patients and controls are found on EGA under Accession number EGAS00001002334 and on the database for Genotypes and Phenotypes (dbGaP) under accession number phs001271.v1.p1., respectively. Previously published expression data (RNA-seq data) of 73 UM tumors are available at EGA under accession no. EGAS00001002932. PCAs were performed using HapMap3 (ftp://ftp.ncbi.nlm.nih.gov/hapmap/genotypes/hapmap3_r3). For expression quantitative trait loci (eQTL) analyses, data was obtained from the Genotype-Tissue Expression (GTEx) public database (<https://www.gtexportal.org/home/>). Allele frequency of SNPs of interest in different populations was obtained from the Genome Aggregation Database (GnomAD v2.1.1, <https://gnomad.broadinstitute.org>).

Code availability: The following web-based resources were used in the GWAS analysis: PLINK 1.9 and 2.0 (<https://www.cog-genomics.org/plink/1.9/>, <https://www.cog-genomics.org/plink/2.0/>), Michigan Imputation Server (<https://imputationserver.sph.umich.edu/index.html>), GitHub (<https://github.com/DReichLab/EIG>), and HIRISplex (<https://hirisplex.erasmusmc.nl/>). The code underlying this article will be shared on reasonable request to the corresponding author.

References

- Jager MJ, Shields CL, Cebulla CM, et al. Uveal melanoma. *Nat Rev Dis Primers*. 2020;6(1):24.
- Singh AD, Aronow ME, Sun Y, et al. Chromosome 3 status in uveal melanoma: a comparison of fluorescence in situ hybridization and single-nucleotide polymorphism array. *Invest Ophthalmol Vis Sci*. 2012;53(7):3331-3339.
- Cassoux N, Rodrigues MJ, Plancher C, et al. Genome-wide profiling is a clinically relevant and affordable prognostic test in posterior uveal melanoma. *Br J Ophthalmol*. 2014;98(6):769-774.
- Harbour JW, Onken MD, Roberson ED, et al. Frequent mutation of BAP1 in metastasizing uveal melanomas. *Science*. 2010;330(6009):1410-1413.
- Furney SJ, Pedersen M, Gentien D, et al. SF3B1 mutations are associated with alternative splicing in uveal melanoma. *Cancer Discov*. 2013;3(10):1122-1129.
- Harbour JW, Roberson ED, Anbunathan H, et al. Recurrent mutations at codon 625 of the splicing factor SF3B1 in uveal melanoma. *Nat Genet*. 2013;45(2):133-135.
- Martin M, Maßhöfer L, Temming P, et al. Exome sequencing identifies recurrent somatic mutations in EIF1AX and SF3B1 in uveal melanoma with disomy 3. *Nat Genet*. 2013;45(8):933-936.
- Robertson AG, Shih J, Yau C, et al.; for the TCGA Research Network. Integrative analysis identifies four molecular and clinical subsets in uveal melanoma. *Cancer Cell*. 2017;32(2):204-220.e15.
- Aronow ME, Topham AK, Singh AD. Uveal melanoma: 5-year update on incidence, treatment, and survival (SEER 1973-2013). *Ocul Oncol Pathol*. 2018;4(3):145-151.
- Bishop KD, Olszewski AJ. Epidemiology and survival outcomes of ocular and mucosal melanomas: a population-based analysis. *Int J Cancer*. 2014;134(12):2961-2971.
- Weis E, Shah CP, Lajoux M, et al. The association between host susceptibility factors and uveal melanoma: a meta-analysis. *Arch Ophthalmol*. 2006;124(1):54-60.
- Mobuchon L, Battistella A, Bardel C, et al. A GWAS in uveal melanoma identifies risk polymorphisms in the CLPTM1L locus. *NPJ Genom Med*. 2017;2(1):5.
- Thomsen H, Chattopadhyay S, Hoffmann P, et al. Genome-wide study on uveal melanoma patients finds association to DNA repair gene TDP1. *Melanoma Res*. 2020;30(2):166-172.
- Alsafadi S, Houy A, Battistella A, et al. Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. *Nat Commun*. 2016;7:10615.
- Visser M, Palstra RJ, Kayser M. Allele-specific transcriptional regulation of IRF4 in melanocytes is mediated by chromatin looping of the intronic rs12203592 enhancer to the IRF4 promoter. *Hum Mol Genet*. 2015;24(9):2649-2661.

16. Praetorius C, Grill C, Stacey SN, et al. A polymorphism in IRF4 affects human pigmentation through a tyrosinase-dependent MITF/TFAP2A pathway. *Cell*. 2013;155(5):1022–1033.
17. Visser M, Kayser M, Palstra RJ. HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter. *Genome Res*. 2012;22(3):446–455.
18. Walsh S, Liu F, Ballantyne KN, et al. IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. *Forensic Sci Int Genet*. 2011;5(3):170–180.
19. Houtzagers LE, Wierenga APA, Ruys AAM, et al. Iris colour and the risk of developing uveal melanoma. *Int J Mol Sci*. 2020;21(19):7172.
20. Wang Z, Zhu B, Zhang M, et al. Imputation and subset-based association analysis across different cancer types identifies multiple independent risk loci in the TERT-CLPTM1L region on chromosome 5p15.33. *Hum Mol Genet*. 2014;23(24):6616–6633.
21. James MA, Vikis HG, Tate E, et al. CRR9/CLPTM1L regulates cell survival signaling and is required for Ras transformation and lung tumorigenesis. *Cancer Res*. 2014;74(4):1116–1127.
22. Jia J, Bosley AD, Thompson A, et al. CLPTM1L promotes growth and enhances aneuploidy in pancreatic cancer cells. *Cancer Res*. 2014;74(10):2785–2795.
23. Clarke WR, Amundadottir L, James MA. CLPTM1L/CRR9 ectodomain interaction with GRP78 at the cell surface signals for survival and chemoresistance upon ER stress in pancreatic adenocarcinoma cells. *Int J Cancer*. 2019;144(6):1367–1378.
24. Shay JW, Wright WE. Telomeres and telomerase: three decades of progress. *Nat Rev Genet*. 2019;20(5):299–309.
25. Ferguson R, Vogelsang M, Ucisik-Akkaya E, et al. Genetic markers of pigmentation are novel risk loci for uveal melanoma. *Sci Rep*. 2016;6:31191.
26. Frudakis T, Terravainen T, Thomas M. Multilocus OCA2 genotypes specify human iris colors. *Hum Genet*. 2007;122(3–4):311–326.
27. Sturm RA, Larsson M. Genetics of human iris colour and patterns. *Pigment Cell Melanoma Res*. 2009;22(5):544–562.
28. Ainger SA, Jagirdar K, Lee KJ, et al. Skin pigmentation genetics for the clinic. *Dermatology*. 2017;233(1):1–15.
29. Duffy DL, Iles MM, Glass D, et al. GenoMEL. IRF4 variants have age-specific effects on nevus count and predispose to melanoma. *Am J Hum Genet*. 2010;87(1):6–16.
30. Han J, Qureshi AA, Nan H, et al. A germline variant in the interferon regulatory factor 4 gene as a novel skin cancer risk locus. *Cancer Res*. 2011;71(5):1533–1539.
31. Schmidt-Pokrzywniak A, Jockel KH, Bornfeld N, et al. Positive interaction between light iris color and ultraviolet radiation in relation to the risk of uveal melanoma: a case-control study. *Ophthalmology*. 2009;116(2):340–348.
32. Krol ES, Liebler DC. Photoprotective actions of natural and synthetic melanins. *Chem Res Toxicol*. 1998;11(12):1434–1440.
33. Ye T, Hong L, Garguilo J, et al. Photoionization thresholds of melanins obtained from free electron laser-photoelectron emission microscopy, femtosecond transient absorption spectroscopy and electron paramagnetic resonance measurements of oxygen photoconsumption. *Photochem Photobiol*. 2006;82(3):733–737.
34. d'Ischia M, Wakamatsu K, Cicoira F, et al. Melanins and melanogenesis: from pigment cells to human health and technological applications. *Pigment Cell Melanoma Res*. 2015;28(5):520–544.
35. Royer-Bertrand B, Torsello M, Rimoldi D, et al. Comprehensive genetic landscape of uveal melanoma by whole-genome sequencing. *Am J Hum Genet*. 2016;99(5):1190–1198.
36. Johansson PA, Brooks K, Newell F, et al. Whole genome landscapes of uveal melanoma show an ultraviolet radiation signature in iris tumours. *Nat Commun*. 2020;11(1):2408.
37. Rootman J, Gallagher RP. Color as a risk factor in iris melanoma. *Am J Ophthalmol*. 1984;98(5):558–561.
38. Durante MA, Rodriguez DA, Kurtenbach S, et al. Single-cell analysis reveals new evolutionary complexity in uveal melanoma. *Nat Commun*. 2020;11(1):496.
39. Field MG, Kuznetsov JN, Bussies PL, et al. BAP1 loss is associated with DNA methylomic repatterning in highly aggressive class 2 uveal melanomas. *Clin Cancer Res*. 2019;25(18):5663–5673.
40. Onken MD, Worley LA, Ehlers JP, et al. Gene expression profiling in uveal melanoma reveals two molecular classes and predicts metastatic death. *Cancer Res*. 2004;64(20):7205–7209.

SUPPLEMENTARY MATERIALS

Supplementary Methods

Study populations

This study was approved by the Ethical Committee and Internal Review Board at the Institut Curie. Blood (germline) samples were obtained from uveal melanoma (UM) patients who consented to participate to the study. Germline DNAs for 946 UMs qualified for our study and DNA was isolated from whole blood using DNeasy Blood & Tissue Kit (Qiagen) according to the manufacturer's instructions. DNAs from 496 control individuals of French origin were obtained from the KIDRISK consortium (US NCI U01CA155309; G. Scelo). Genotypes were obtained on the Infinium Global Screening Array 24 v1.0 (Illumina) at the Centre National de Génomique (Evry, France). Genotypes were called using default parameters in GenomeStudio (Illumina).

Filtering genotyping data

Genotypes for UMs and controls from the previously published GWAS (dataset1) (1) and genotypes for two new sets of UM cases (dataset2 and dataset3) were filtered using the following criteria in an automated pipeline. All SNPs were mapped, standardized with respect to strand and renamed according to the genomic location, the reference and the alternative alleles (CHR:POS:REF>ALT). Only biallelic SNPs with a unique location on GRCh37 were included in analyses. Gender of each sample was predicted using plink1.9 (v.20200219) and samples with difference between predicted and observed sexes were filtered out. Plink2 (v.20201028) was used to remove one sample of each related pair, with kinship coefficient >0.05. Samples with a genotype completion rate less than 95% were removed from the analysis. SNPs with less than 95% of calls or a minor allele frequency less than 1% in all samples and with a Hardy-Weinberg Equilibrium p-value lower than 10^{-3} in control samples were excluded from the analysis.

Imputation and merge

The three datasets were then independently imputed on the Michigan Imputation Server (<https://imputationserver.sph.umich.edu/>) using Eagle for the phasing and Haplotype Reference Consortium r1.1 as the reference dataset. SNPs were filtered out if imputation score was <0.7. All imputed datasets were merged together and another quality control was performed with the same threshold as before (Supplementary Table 1).

Preparation for principal component analyses (PCA)

Prior to each principal component analysis (PCA), data were processed as recommended in the plink2 documentation (<https://www.cog-genomics.org/plink/2.0/strat#pca>) for linkage disequilibrium (LD) pruning, with the following parameters: windows of 250kb, a sliding step of 1 variant and a r^2 threshold of 0.02. SNPs present in highly conserved regions or with high LD with SNPs of interest were removed. PCA was then performed using plink2 to select a population with European ancestry and using smartPCA from EIGENSOFT package v7.2.1 (<https://github.com/DReichLab/EIG>) to get the 10 first components used as covariates in the regression analysis.

Selection of individuals of European ancestry

To stringently select individuals of European ancestry, samples that passed the quality control were merged with the HapMap project data (Phase 3 with 1,397 individuals and 924,118 SNPs) and prepared as described above (Supplementary Figure 1). Center of the HapMap European cluster was determined by the mean of the two first components using plink2 of all samples coming from CEU (Utah residents with Northern and Western European ancestry from the CEPH collection) and TSI (Tuscany in Italia) populations. The maximal distance, in PC1-PC2 space, between this center and all European samples from HapMap increased by 10% was used as a threshold to define cases and controls of European ancestry (Supplementary Figure 2), and these selected samples were used for further analyses.

Principal component analysis

Data from samples of Europeans ancestry were processed as described in preparation for PCA. SmartPCA with default parameters (10 PCA iterations (numoutlierevec) ; 10 PCs to output (numoutevec) ; a maximum of 5 outliers to remove for each iteration (numoutlieriter)) was then used to extract the ten first components.

Statistical analyses

For whole genome analysis comparing all cases against all controls, Firth logistic regression was performed using plink2 with age, sex, dataset (dataset1, dataset2 and dataset3) and the first ten principal components from smartPCA as covariates. Exact number of cases and controls used for GWAS analyses are indicated. Association of SNPs with UM risk was determined by Odds Ratio (OR) [CI 95%] (Table 1, Supplementary Table 3), and SNPs with a P value $<5 \times 10^{-8}$ for this analysis were considered to be significant while those with P value $<1 \times 10^{-5}$ only reached the tendency line (Figure 2). Association of blue eye color with UM risk was calculated using a two-sided Fisher test p -value and OR (Figure 3). Comparison of

Variant Allele Frequency of SNPs of interest in different populations were also tested for significance using a two-sided Fisher test p-value (Supplementary Table 5). Expression Quantitative Trait Loci (eQTL) were performed using linear regression. For analyses stratified by subtype comparing cases with or without loss of chromosome 3 (monosomy 3 [M3] or disomy 3 [D3], respectively) against controls, Firth logistic regression was also performed using plink2 with the same covariates as mentioned above. For risk loci analysis comparing M3 cases against D3 cases, logistic regression was performed using R with age, sex and dataset as covariates. For M3 vs. D3 stratification analysis, logistic regression with sex, age and dataset as covariates was used for analysis and results appear as OR [CI 95%] (Table 1, Supplementary Table 6). A *P* value of <0.05 was considered statistically significant for all tests except for GWAS Firth logistic regression mentioned above.

Conditional analysis

For conditional analyses, rs421284, rs12203592 or rs12913832 were added to the list of covariates (age, sex, dataset and the first ten principal components) in three independent, genome-wide analyses through logistic regressions.

Targeted study

To validate genotyping and imputation quality, manual genotyping was performed using TaqMan® allele specific primers (rs421284, C_2396811_10; rs12203592, C_31918199_10; rs12913832, C_30724404_10) and PCR primers (#cat: 4351379, Thermofisher). PCRs were carried out according to the manufacturer's instructions on 10 ng of genomic DNA using the following cycling conditions: initial denaturation at 95°C, 10 min; 40 cycles of 95°C, 15 s; 60°C, 1 min. Endpoint analyses were carried out using the Applied Biosystems 7500HT Fast Real-Time PCR System and collected using Applied Biosystems TaqMan® Genotyper Software Version 1.3.

Eye color prediction

Eye color was predicted using IrisPlex tools (<https://hirisplex.erasmusmc.nl/>) based on the genotypes of the six required SNPs (*HERC2* rs12913832; *OCA2* rs1800407; *LOC105370627* rs12896399; *SLC45A2* rs16891982; *TYR* rs1393350; and *IRF4* rs12203592) for European patient and control samples as input.

Expression analyses

Expression analyses were performed on RNA-seq data from 73 in-house UM tumor samples (2). Tumors with a copy number alteration at the studied locus were removed from eQTL analyses to avoid confounding by copy number. Correlation between genotype and

expression was analyzed with linear regression using the MatrixEQTL package in R (3), and statistical significance was set at $p < 0.05$.

Data availability

Dataset2 and 3 genotyping data used in the analysis are have been deposited and are available on the European Genome-Phenome Archive (EGA) (<https://ega-archive.org/>) under accession number EGAS00001005200. Previously published genotyping of dataset1 cases and controls are found on EGA under Accession number EGAS00001002334 and on the database for Genotypes

and Phenotypes (dbGaP) under accession number phs001271.v1.p1., respectively.

Previously published expression data (RNA-seq data) of 73 UM tumors are available at EGA under accession no. EGAS00001002932). PCAs were performed using HapMap3

(ftp://ftp.ncbi.nlm.nih.gov/hapmap/genotypes/hapmap3_r3). For expression quantitative trait loci (eQTL) analyses, data was obtained from the Genotype-Tissue Expression (GTEx) public database (<https://www.gtexportal.org/home/>). Allele frequency of SNPs of interest in different populations was obtained from the Genome Aggregation Database (GnomAD v2.1.1, <https://gnomad.broadinstitute.org>).

Code availability The following web-based resources were used in the GWAS analysis:

PLINK 1.9 and 2.0 (<https://www.cog-genomics.org/plink/1.9/>, <https://www.cog-genomics.org/plink/2.0/>), Michigan Imputation Server

(<https://imputationserver.sph.umich.edu/index.html>), GitHub

(<https://github.com/DReichLab/EIG>), and HIRISplex (<https://hirisplex.erasmusmc.nl/>).

The code generated during this study is available at https://gitlab.curie.fr/ahouy/gwas/-/blob/master/GWAS_analysis.html.

Supplementary references

1. Mobuchon L, Battistella A, Bardel C, *et al*. A GWAS in uveal melanoma identifies risk polymorphisms in the CLPTM1L locus. NPJ Genom Med 2017;2(1):5.
2. Alsafadi S, Houy A, Battistella A, *et al*. Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. Nat Commun 2016;7:10615.
3. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics 2012;28(10):1353-8.

Supplementary Table 1 : Number of samples and SNPs before and after the merging step of the 3 datasets included in the GWAS analysis.

BEFORE MERGING	Dataset 1			Dataset 2			Dataset 3		
	Cases	Controls	SNPs	Cases	Controls	SNPs	Cases	Controls	SNPs
Raw data	271	429	4,255,329	369	496	685,214	577	0	685,276
-check-sex	261	428	4,255,329	361	496	685,214	576	0	685,276
-king-table-filter 0.05	260	421	4,255,329	361	494	685,214	575	0	685,276
-mind 0.05	260	421	4,255,329	361	494	685,214	574	0	685,276
-geno 0.05	260	421	4,147,006	361	494	674,658	574	0	676,370
-maf 0.01	260	421	2,475,290	361	494	503,432	574	0	505,188
-hwe 10^{-3} (only on controls)	260	421	2,469,061	361	494	502,088	skip (no controls)		
Imputed data	260	421	18,839,472	361	494	18,181,178	574	0	16,748,402
rs > 0.7	260	421	16,273,929	361	494	13,884,737	574	0	13,087,057
AFTER MERGING	Dataset								
	Cases	Controls	SNPs						
Merged	1,195	915	10,061,295						
IBD	1,194	914	10,061,295						
-mind 0.05	1,194	914	10,061,295						
-geno 0.05	1,194	914	10,061,295						
-maf 0.01	1,194	914	7,527,748						
-hwe 10^{-3} (only on controls)	1,194	914	7,520,959						
European selection (HapMap)	1,143	883	7,520,959						
Outlier removal (smartpca)	1,142	882	7,520,959						
-mind 0.05	1,142	882	7,520,959						
-geno 0.05	1,142	882	7,520,959						
-maf 0.01	1,142	882	7,489,498						
-hwe 10^{-3} (only on controls)	1,142	882	7,488,175						

Supplementary Table 2: Validation by TaqMan genotyping of imputed genotypes used in the GWAS analysis: genotyping of 972 selected samples on rs421284, rs12203592 and rs12913832.

Available for separate download

Supplementary Table 3 : Loci reaching genome-wide significance, 5×10^{-8} , or tendency, 10^{-5} , for association with uveal melanoma.

ID ^a	SNP	Symbol	A1 ^b	Total (Case + Control)	P	OR [CI 95%] ^c	Threshold
2:134821140:A>G	rs6733440	NA	G	2024 (1142 + 882)	6.30E-06	1.51 [1.26;1.81]	Tendency
2:134821253:C>G	rs6718849	NA	G	2024 (1142 + 882)	6.30E-06	1.51 [1.26;1.81]	Tendency
2:134826007:C>T	rs4954103	NA	T	2024 (1142 + 882)	9.44E-06	1.5 [1.25;1.8]	Tendency
5:1300429:T>G	rs2735946	NA	T	2024 (1142 + 882)	7.99E-07	1.51 [1.28;1.78]	Tendency
5:1302144:T>C	rs2736102	NA	T	2024 (1142 + 882)	4.42E-06	1.45 [1.24;1.7]	Tendency
5:1302914:G>A	rs2853666	NA	G	2024 (1142 + 882)	4.42E-06	1.45 [1.24;1.7]	Tendency
5:1303901:T>C	rs2735945	NA	T	2024 (1142 + 882)	4.42E-06	1.45 [1.24;1.7]	Tendency
5:1306165:T>C	rs4404721	NA	T	2024 (1142 + 882)	3.21E-06	1.46 [1.24;1.71]	Tendency
5:1306331:T>C	rs4530805	NA	T	2024 (1142 + 882)	4.04E-06	1.45 [1.24;1.7]	Tendency
5:1306765:C>G	rs11133727	NA	C	2024 (1142 + 882)	1.34E-06	1.48 [1.26;1.73]	Tendency
5:1309168:T>C	rs61574973	NA	T	2024 (1142 + 882)	2.68E-06	1.46 [1.25;1.71]	Tendency
5:1309904:G>A	rs60622800	NA	G	2024 (1142 + 882)	1.34E-06	1.48 [1.26;1.73]	Tendency
5:1310152:G>A	rs6554758	NA	G	2024 (1142 + 882)	1.34E-06	1.48 [1.26;1.73]	Tendency
5:1311693:T>C	rs6866294	NA	T	2024 (1142 + 882)	1.12E-06	1.48 [1.26;1.74]	Tendency
5:1312020:T>C	rs6866783	NA	T	2024 (1142 + 882)	2.68E-06	1.46 [1.25;1.72]	Tendency
5:1312457:G>A	rs13356727	NA	G	2024 (1142 + 882)	7.93E-07	1.49 [1.27;1.75]	Tendency
5:1312935:T>C	rs13355267	NA	T	2024 (1142 + 882)	9.40E-07	1.49 [1.27;1.74]	Tendency
5:1313701:A>G	rs28379291	NA	A	2024 (1142 + 882)	2.69E-06	1.46 [1.25;1.72]	Tendency
5:1314009:C>T	rs10078017	NA	C	2024 (1142 + 882)	3.11E-06	1.46 [1.25;1.71]	Tendency
5:1315343:G>A	rs4975615	NA	G	2024 (1142 + 882)	1.64E-06	1.48 [1.26;1.73]	Tendency
5:1315660:G>A	rs4975616	NA	G	2024 (1142 + 882)	7.42E-07	1.49 [1.27;1.75]	Tendency
5:1317820:A>G	rs3816659	NA	A	2024 (1142 + 882)	1.59E-07	1.53 [1.31;1.8]	Tendency
5:1320136:G>A	rs421629	CLPTM1L	A	2024 (1142 + 882)	3.25E-08	1.57 [1.34;1.85]	Significancy
5:1320247:G>A	rs380286	CLPTM1L	A	2024 (1142 + 882)	2.83E-08	1.58 [1.34;1.85]	Significancy
5:1322087:C>T	rs401681	CLPTM1L	T	2024 (1142 + 882)	7.61E-08	1.55 [1.32;1.82]	Tendency
5:1322468:G>A	rs381949	CLPTM1L	A	2024 (1142 + 882)	1.65E-07	1.53 [1.31;1.8]	Tendency
5:1323212:C>T	rs13178866	CLPTM1L	T	2024 (1142 + 882)	2.85E-08	1.57 [1.34;1.85]	Significancy
5:1324121:G>A	rs414965	CLPTM1L	A	2024 (1142 + 882)	1.02E-07	1.54 [1.32;1.81]	Tendency
5:1325590:T>C	rs421284 ^d	CLPTM1L	C	2024 (1142 + 882)	1.98E-08	1.58 [1.35;1.86]	Significancy
5:1325767:A>G	rs466502	CLPTM1L	G	2024 (1142 + 882)	2.26E-08	1.58 [1.35;1.85]	Significancy
5:1325803:A>G	rs465498	CLPTM1L	G	2024 (1142 + 882)	1.67E-08	1.59 [1.35;1.86]	Significancy
5:1327851:C>T	rs383009	CLPTM1L	T	2024 (1142 + 882)	4.85E-08	1.56 [1.33;1.83]	Significancy
5:1330253:T>C	rs452932	CLPTM1L	C	2024 (1142 + 882)	1.67E-08	1.59 [1.35;1.86]	Significancy
5:1330840:T>C	rs452384	CLPTM1L	C	2024 (1142 + 882)	1.67E-08	1.59 [1.35;1.86]	Significancy
5:1331219:A>G	rs370348 ^d	CLPTM1L	G	2024 (1142 + 882)	1.48E-08	1.59 [1.35;1.86]	Significancy
5:1333077:A>G	rs2447853	CLPTM1L	G	2024 (1142 + 882)	1.76E-08	1.58 [1.35;1.86]	Significancy
5:1336178:A>T	rs457130	CLPTM1L	T	2024 (1142 + 882)	7.02E-08	1.55 [1.32;1.82]	Tendency
5:1336221:T>C	rs467095	CLPTM1L	C	2024 (1142 + 882)	2.21E-08	1.58 [1.35;1.85]	Significancy
5:1336243:A>G	rs455433	CLPTM1L	G	2024 (1142 + 882)	1.92E-08	1.58 [1.35;1.86]	Significancy
5:1336459:T>C	rs460073	CLPTM1L	C	2024 (1142 + 882)	1.92E-08	1.58 [1.35;1.86]	Significancy
5:1336626:T>A	rs462608	CLPTM1L	A	2024 (1142 + 882)	7.02E-08	1.55 [1.32;1.82]	Tendency
5:1337070:T>C	rs456366	CLPTM1L	C	2024 (1142 + 882)	2.21E-08	1.58 [1.35;1.85]	Significancy
5:1337106:T>A	rs459961	CLPTM1L	A	2024 (1142 + 882)	1.92E-08	1.58 [1.35;1.86]	Significancy
5:1337906:A>T	rs31484	CLPTM1L	T	2024 (1142 + 882)	2.51E-08	1.58 [1.34;1.85]	Significancy
5:1341101:G>C	rs31487	CLPTM1L	C	2024 (1142 + 882)	1.95E-08	1.58 [1.35;1.85]	Significancy
5:1342714:C>A	rs31489	CLPTM1L	A	2024 (1142 + 882)	2.09E-07	1.53 [1.31;1.79]	Tendency
5:1344458:G>A	rs31490	CLPTM1L	A	2024 (1142 + 882)	7.64E-08	1.55 [1.32;1.82]	Tendency
5:1345474:A>G	rs27996	NA	G	2024 (1142 + 882)	1.05E-06	1.48 [1.27;1.74]	Tendency
5:1346303:G>C	rs27070	NA	C	2024 (1142 + 882)	8.29E-07	1.49 [1.27;1.75]	Tendency
5:1347128:C>T	rs27069	NA	T	2024 (1142 + 882)	1.74E-06	1.47 [1.26;1.72]	Tendency
5:1348798:A>T	rs37011	NA	T	2024 (1142 + 882)	8.86E-07	1.49 [1.27;1.74]	Tendency
5:1349535:A>G	rs37010	NA	G	2024 (1142 + 882)	1.25E-06	1.48 [1.26;1.73]	Tendency
5:1350339:C>T	rs37009	NA	T	2024 (1142 + 882)	2.39E-06	1.46 [1.25;1.71]	Tendency
5:1350397:G>A	rs40182	NA	A	2024 (1142 + 882)	2.19E-06	1.47 [1.25;1.72]	Tendency
5:1351538:G>A	rs37008	NA	A	2024 (1142 + 882)	9.58E-07	1.49 [1.27;1.74]	Tendency
5:1352372:G>C	rs37007	NA	C	2024 (1142 + 882)	7.30E-07	1.49 [1.27;1.75]	Tendency
5:1354462:G>T	rs40181	NA	T	2024 (1142 + 882)	4.01E-07	1.51 [1.28;1.76]	Tendency
5:1355058:C>T	rs37006	NA	T	2024 (1142 + 882)	3.02E-07	1.51 [1.29;1.77]	Tendency
5:1356450:C>T	rs37005	NA	T	2024 (1142 + 882)	4.32E-07	1.5 [1.28;1.76]	Tendency
5:1356684:C>T	rs37004	NA	T	2024 (1142 + 882)	7.40E-06	1.51 [1.26;1.81]	Tendency
5:1356771:A>C	rs37003	NA	C	2024 (1142 + 882)	4.32E-07	1.5 [1.28;1.76]	Tendency
5:1361669:T>G	rs31494	NA	G	2024 (1142 + 882)	9.87E-06	0.68 [0.57;0.81]	Tendency
5:113999670:C>T	rs285877	NA	C	2024 (1142 + 882)	5.66E-06	0.67 [0.56;0.79]	Tendency
5:114009685:A>C	rs12719183	NA	A	2024 (1142 + 882)	5.34E-06	0.67 [0.56;0.79]	Tendency
5:114013830:T>G	rs608861	NA	T	2024 (1142 + 882)	3.44E-06	0.66 [0.55;0.79]	Tendency
5:114014162:A>G	rs493715	NA	A	2024 (1142 + 882)	3.44E-06	0.66 [0.55;0.79]	Tendency
5:114017921:C>G	rs287668	NA	C	2024 (1142 + 882)	3.30E-06	0.66 [0.55;0.79]	Tendency
5:114023037:T>C	rs285896	NA	T	2024 (1142 + 882)	2.87E-06	0.66 [0.55;0.78]	Tendency
5:114023462:G>C	rs285895	NA	G	2024 (1142 + 882)	2.65E-06	0.66 [0.55;0.78]	Tendency
5:114027189:C>T	rs285913	NA	C	2024 (1142 + 882)	2.92E-06	0.66 [0.55;0.78]	Tendency
5:114027841:G>C	rs285912	NA	G	2024 (1142 + 882)	2.92E-06	0.66 [0.55;0.78]	Tendency
5:114031231:T>C	rs285906	NA	T	2024 (1142 + 882)	1.72E-06	0.65 [0.54;0.78]	Tendency
5:114032425:A>G	rs285902	NA	A	2024 (1142 + 882)	1.62E-06	0.65 [0.55;0.78]	Tendency
5:114034445:G>A	rs2030374	NA	G	2024 (1142 + 882)	2.74E-06	0.66 [0.55;0.78]	Tendency
5:114035271:G>T	rs65174	NA	G	2024 (1142 + 882)	2.74E-06	0.66 [0.55;0.78]	Tendency
5:114040362:T>C	rs2416376	NA	T	2024 (1142 + 882)	2.31E-06	0.66 [0.55;0.78]	Tendency
5:114045412:C>T	rs285884	NA	C	2024 (1142 + 882)	5.00E-06	0.67 [0.56;0.79]	Tendency
5:114045730:G>A	rs285882	NA	G	2024 (1142 + 882)	5.75E-06	0.67 [0.56;0.79]	Tendency
5:114045774:T>C	rs285881	NA	T	2024 (1142 + 882)	5.75E-06	0.67 [0.56;0.79]	Tendency
5:114046746:A>C	rs285880	NA	A	2024 (1142 + 882)	1.86E-06	0.67 [0.56;0.79]	Tendency
5:114047808:T>C	rs157980	NA	T	2024 (1142 + 882)	5.02E-06	0.66 [0.56;0.79]	Tendency
5:114050162:C>T	rs1520256	NA	C	2024 (1142 + 882)	7.07E-06	0.67 [0.56;0.8]	Tendency
5:114057235:C>T	rs287685	NA	C	2024 (1142 + 882)	5.34E-06	0.67 [0.56;0.79]	Tendency
5:114069653:C>T	rs34310	NA	C	2024 (1142 + 882)	8.12E-06	0.67 [0.57;0.8]	Tendency
5:114070883:G>C	rs40057	NA	G	2024 (1142 + 882)	4.52E-06	0.66 [0.56;0.79]	Tendency
5:114075021:T>C	rs34311	NA	T	2024 (1142 + 882)	4.79E-06	0.66 [0.55;0.79]	Tendency
6:396321:C>T	rs12203592 ^d	IRF4	T	2024 (1142 + 882)	3.55E-08	1.76 [1.44;2.16]	Significancy
9:18258312:G>T	rs7846812	ADAMTSL1	T	2024 (1142 + 882)	7.68E-06	0.68 [0.57;0.81]	Tendency
9:100900782:A>G	rs4743170	CORO2A	G	2024 (1142 + 882)	3.74E-06	1.47 [1.25;1.73]	Tendency
9:100911671:T>C	rs814027	CORO2A	C	2024 (1142 + 882)	8.39E-06	0.7 [0.6;0.82]	Tendency

9:100911730:A>G	rs774122	CORO2A	A	2024 (1142 + 882)	3.99E-06	1.51 [1.27;1.8]	Tendency
10:86357721:T>G	rs11201120	NA	G	2024 (1142 + 882)	4.87E-06	1.51 [1.27;1.8]	Tendency
10:86358506:G>C	rs2488636	NA	G	2024 (1142 + 882)	1.02E-06	1.52 [1.29;1.8]	Tendency
10:86359514:T>C	rs2488638	NA	T	2024 (1142 + 882)	2.99E-06	1.5 [1.27;1.78]	Tendency
10:86361389:A>G	rs11813775	NA	A	2024 (1142 + 882)	3.70E-06	1.5 [1.26;1.77]	Tendency
11:119923914:A>G	rs142100881	NA	G	2024 (1142 + 882)	7.61E-06	0.16 [0.07;0.36]	Tendency
13:59970348:C>T	rs113515519	NA	T	2024 (1142 + 882)	6.00E-06	3.79 [2.13;6.74]	Tendency
15:28356859:C>T	rs1129038 ^d	HERC2	C	2024 (1142 + 882)	5.97E-12	0.56 [0.48;0.66]	Significance
15:28364059:A>G	rs7494942	HERC2	A	2024 (1142 + 882)	5.56E-11	0.53 [0.44;0.64]	Significance
15:28365618:A>G	rs12913832 ^d	HERC2	A	2024 (1142 + 882)	1.88E-11	0.57 [0.48;0.67]	Significance
15:28374012:T>C	rs3935591	HERC2	T	2024 (1142 + 882)	2.58E-07	0.59 [0.48;0.72]	Tendency
15:28383565:T>C	rs7403279	HERC2	T	2024 (1142 + 882)	7.27E-11	0.53 [0.44;0.64]	Significance
15:28410491:C>T	rs12916300	HERC2	C	2024 (1142 + 882)	2.62E-11	0.57 [0.48;0.67]	Significance
15:28453215:A>G	rs2238289	HERC2	G	2024 (1142 + 882)	7.54E-07	0.59 [0.48;0.73]	Tendency
15:28468723:T>G	rs3940272	HERC2	T	2024 (1142 + 882)	1.28E-07	0.58 [0.47;0.71]	Tendency
15:28496868:G>C	rs6497293	HERC2	G	2024 (1142 + 882)	1.38E-07	0.58 [0.47;0.71]	Tendency
15:28502279:A>G	rs11631797	HERC2	A	2024 (1142 + 882)	1.38E-07	0.58 [0.47;0.71]	Tendency
15:28513364:T>C	rs916977	HERC2	T	2024 (1142 + 882)	7.41E-10	0.55 [0.45;0.66]	Significance
15:28516084:C>T	rs8039195	HERC2	C	2024 (1142 + 882)	1.63E-07	0.58 [0.47;0.71]	Tendency
15:28530182:C>T	rs1667394	HERC2	C	2024 (1142 + 882)	1.29E-10	0.53 [0.44;0.65]	Significance
15:28537701:C>T	rs1667390	HERC2	C	2024 (1142 + 882)	2.06E-07	0.58 [0.48;0.72]	Tendency
15:28539834:C>T	rs1635166	HERC2	C	2024 (1142 + 882)	2.06E-07	0.58 [0.48;0.72]	Tendency
15:28562998:T>C	rs1614575	HERC2	T	2024 (1142 + 882)	3.57E-10	0.54 [0.45;0.66]	Significance
18:65494452:T>G	rs17184817	NA	G	2024 (1142 + 882)	5.23E-06	2.52 [1.69;3.76]	Tendency
18:65497221:G>A	rs111785612	NA	A	2024 (1142 + 882)	5.23E-06	2.52 [1.69;3.76]	Tendency
18:65499037:C>T	rs9962198	NA	T	2024 (1142 + 882)	7.19E-06	2.48 [1.67;3.68]	Tendency
18:65501129:G>T	rs9951755	NA	T	2024 (1142 + 882)	7.19E-06	2.48 [1.67;3.68]	Tendency
18:65503256:A>G	rs9960419	NA	G	2024 (1142 + 882)	7.19E-06	2.48 [1.67;3.68]	Tendency
18:65503412:T>C	rs9949829	NA	C	2024 (1142 + 882)	7.19E-06	2.48 [1.67;3.68]	Tendency
18:65504506:G>C	rs7239668	NA	C	2024 (1142 + 882)	5.45E-06	2.52 [1.69;3.75]	Tendency
18:65504758:G>A	rs9964442	NA	A	2024 (1142 + 882)	5.45E-06	2.52 [1.69;3.75]	Tendency
X:22567684:T>C	rs147479986	NA	C	2024 (1142 + 882)	9.74E-06	1.7 [1.34;2.16]	Tendency
X:22574447:G>C	rs147552497	NA	C	2024 (1142 + 882)	7.08E-06	1.71 [1.35;2.15]	Tendency
X:22574497:G>C	rs146125348	NA	C	2024 (1142 + 882)	4.85E-06	1.73 [1.37;2.19]	Tendency
X:22574644:T>C	rs16982108	NA	C	2024 (1142 + 882)	7.08E-06	1.71 [1.35;2.15]	Tendency

^aID refers to chromosome number : chromosomal genomic position : reference allele > alternative allele, based on genome build hg19.

^bA1 refers to the minor allele.

^cOR: odds ratio. To calculate ORs, results were corrected for sex, age and for the first ten components of the smartpca.

^dSNPs of interest in this study.

Supplementary Table 4: Effect of adding eye color as a covariate on the Odds Ratio for the main SNPs of the 3 significant UM risk loci.

ID ^a	SNP	Symbol	A1 ^b	Total (Cases + Controls)	Without eye covariate		With eye covariate	
					P	OR ^c [CI 95%]	P	OR ^c [CI 95%]
5:1325590:T>C	rs421284	CLPTM1L	C	2024 (1142 + 882)	1.98E-08	1.58 ^d [1.35;1.86]	4.01E-08	1.58 ^d [1.34;1.86]
6:396321:C>T	rs12203592	IRF4	T	2024 (1142 + 882)	3.55E-08	1.76 ^d [1.44;2.16]	9.25E-08	1.76 ^d [1.43;2.17]
15:28365618:A>G	rs12913832	HERC2	A	2024 (1142 + 882)	1.88E-11	0.57 ^d [0.48;0.67]	6.44E-02	0.76 [0.57;1.02]

^aID refers to chromosome number : chromosomal genomic position : reference allele > alternative allele, based on genome build hg19.

^bMinor allele

^cOR: odds ratio (with 95% confidence interval in brackets) resulting from the generalized linear model (GLM) from GWAS imputed data with correction for sex, age, dataset and the ten first components from smartpca, and with or without eye color as predicted by the IrisPlex tool as covariate.

^dOdds ratios that show an increase of the risk linked to the risk allele.

Supplementary Table 5: Allele frequencies for the 3 main SNPs of UM risk loci in different populations defined by the Genome Aggregation Database.

Gene	SNP	ID ^a	VAF ^b			Two-sided fisher test p-value	
			European (NFE) ^c	African/African-American	East Asian	NFE vs. African	NFE vs. East Asian
<i>CLPTM1L</i>	rs421284	5:1325590:T>C	0.429	0.597	0.200	3.71×10^{-138}	6.9×10^{-75}
<i>IRF4</i>	rs12203592	6:396321:C>T	0.144	0.034	0.000	4.48×10^{-185}	3.65×10^{-100}
<i>HERC2</i>	rs12913832	15:28365618:A>G	0.803	0.125	0.001	0	0

^aID refers to chromosome number : chromosomal genomic position : reference allele > alternative allele, based on genome build hg19.

^bVAF: Variant allele frequency

^cNon-Finnish Europeans

Supplementary Table 6 : Results of the Generalized Linear Model, GLM, for the main UM risk loci according to chromosome 3 status.

ID ^a	SNP	Symbol	A1 ^b	VAF ^c	Total (M3 + D3 cases ^d)	P	OR ^e [CI 95%]
5:1325590:T>C	rs421284	<i>CLPTM1L</i>	C	53%	381 (244 + 137)	3.28E-01	0.86 [0.67;1.11]
6:396321:C>T	rs12203592	<i>IRF4</i>	T	22%	381 (244 + 137)	8.46E-07	0.38 [0.27;0.52] ^f
15:28365618:A>G	rs12913832	<i>HERC2</i>	G	70%	381 (244 + 137)	3.87E-04	1.81 [1.38;2.38] ^g

^aID refers to chromosome number : chromosomal genomic position : reference allele > alternative allele, based on genome build hg19.

^bA1: Minor allele

^cVAF: variant allele frequency

^dTotal number of cases in the monosomy 3 (M3) and disomy 3 (D3) UM subgroups used in the analysis.

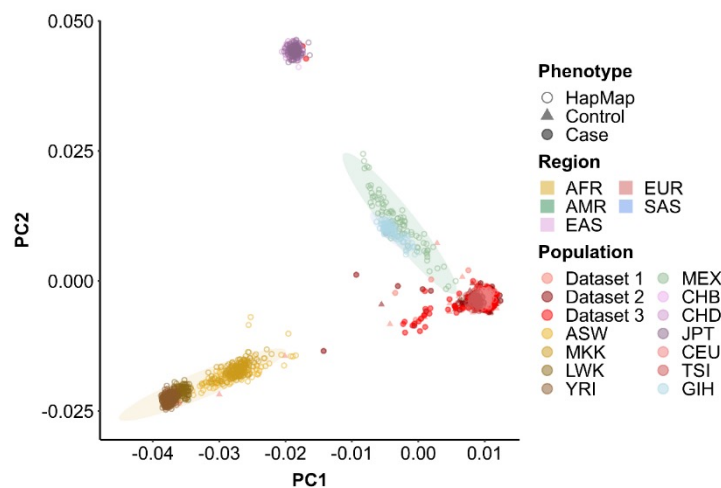
^eOR: odds ratio resulting from the GLM comparison of UM cases from the 3 merged datasets with M3 against those with D3, from imputed data with correction for sex, age and dataset.

^fOdds Ratio shows a statistically significant increase of UM risk with the risk allele in D3 cases.

^gOdds Ratio shows a statistically significant increase of UM risk with the risk allele in M3 cases.

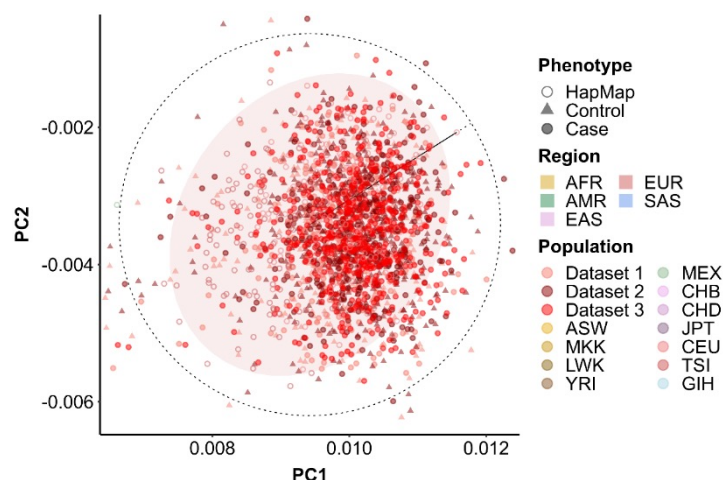
Supplementary Figures 1 – 6

Supplementary Figure 1



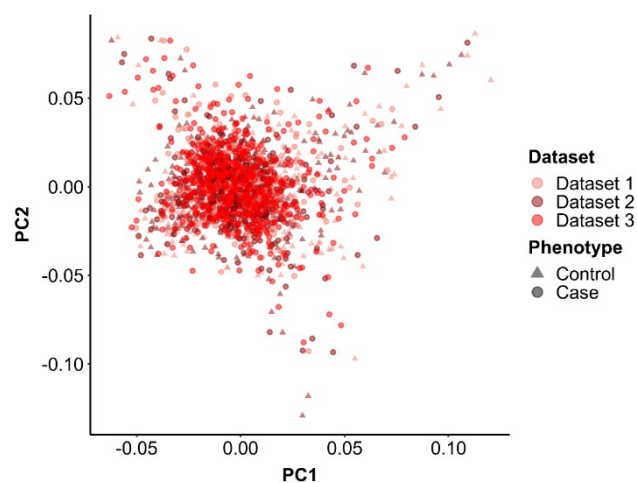
Supplementary Figure 1: HapMap Principal Component Analysis (PCA) for the merged imputed GWAS dataset. Each dot represents the projection of the first two components (PC1 and PC2) of the PCA combining the imputed GWAS with the HapMap dataset. Colored ellipses correspond to the 5 super-populations (regions) where AFR = African, AMR = Ad-Mixed American, EAS = East Asian, EUR = European and SAS = South Asian. Among populations other than merged datasets 1, 2 and 3, appear the following: AFR populations : ASW = African ancestry in Southwest USA; MKK = Maasai in Kinyawa, Kenya ; LWK = Luhya in Webuye, Kenya ; YRI = Yoruba in Ibadan, Nigeria. AMR population : MEX = Mexican ancestry in Los Angeles, California. EAS populations : CHB = Han Chinese in Beijing, China ; CHD = Chinese in Metropolitan Denver, Colorado; JPT = Japanese in Tokyo, Japan. EUR populations : CEU = Utah residents with Northern and Western European ancestry from the CEPH collection ; TSI = Toscani in Italia. SAS population : GIH = Gujarati Indians in Houston, Texas.

Supplementary Figure 2



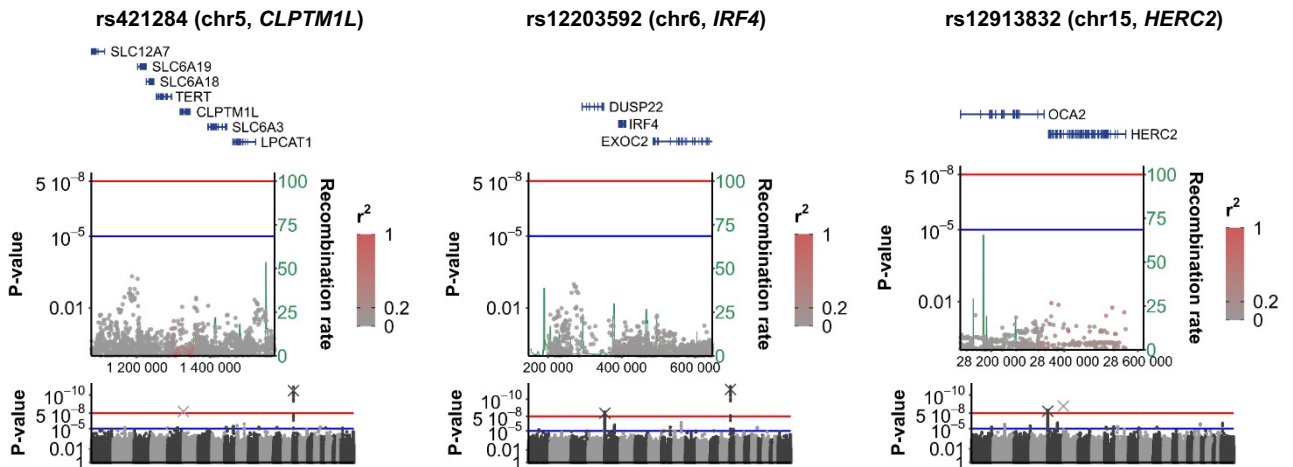
Supplementary Figure 2: Zoomed-in HapMap Principal Component Analysis (PCA) for imputed dataset on the European population cluster. Each dot is the projection of the two first components of the PCA combining the imputed GWAS with the HapMap dataset. Red ellipses correspond to the European super population. The dotted circle represents the threshold used to select European samples: it is centered on the median of PC1 and PC2 of all European samples from HapMap, and its radius is the distance between the furthest European samples from the center (shown with black line) multiplied by 1.1. All samples from GWAS datasets inside this circle are considered as European samples and will be kept for further analyses. Colored ellipses correspond to the 5 super-populations (regions) where AFR = African, AMR = Ad-Mixed American, EAS = East Asian, EUR = European and SAS = South Asian. Among populations other than merged datasets 1, 2 and 3, appear the following: AFR populations : ASW = African ancestry in Southwest USA; MKK = Maasai in Kinyawa, Kenya ; LWK = Luhya in Webuye, Kenya ; YRI = Yoruba in Ibadan, Nigeria. AMR population : MEX = Mexican ancestry in Los Angeles, California. EAS populations : CHB = Han Chinese in Beijing, China ; CHD = Chinese in Metropolitan Denver, Colorado; JPT = Japanese in Tokyo, Japan. EUR populations : CEU = Utah residents with Northern and Western European ancestry from the CEPH collection ; TSI = Toscani in Italia. SAS population : GIH = Gujarati Indians in Houston, Texas.

Supplementary Figure 3



Supplementary Figure 3: First two components of the PCA on merged GWAS datasets 1, 2 and 3, obtained with smartpca.

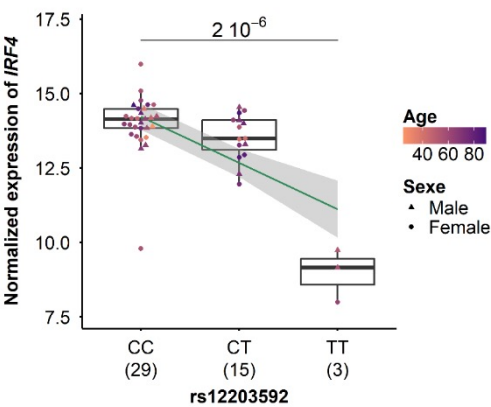
Supplementary Figure 4



Supplementary Figure 4: Manhattan plot and regional linkage disequilibrium (LD) plot for significant UM risk loci after conditional analyses on the three main loci. For the Manhattan plot (bottom), the P-value of the association test (y-axis) is plotted against its physical chromosomal position (x-axis). Chromosomes are shown in alternating black and grey colors. SNPs above the red lines represent those with a P-value $< 5 \times 10^{-8}$ and were considered as significantly associated with uveal melanoma. The blue lines represent the suggestive lines (P-value $< 1 \times 10^{-5}$). SNPs of interest are indicated with a cross on the Manhattan plot. Significance was measured using unconditional logistic regressions. For the regional locus plots (top), genes are depicted with blue rectangles and SNPs appear in colored dots. The color intensity of dots reflects the level of linkage disequilibrium with the highlighted SNP of interest in each UM risk locus (indicated on top). The green bars indicate recombination rates. Chr: chromosome.

Supplementary Figure 5

A. Uveal Melanoma



B. The Genotype-Tissue Expression (GTEx)

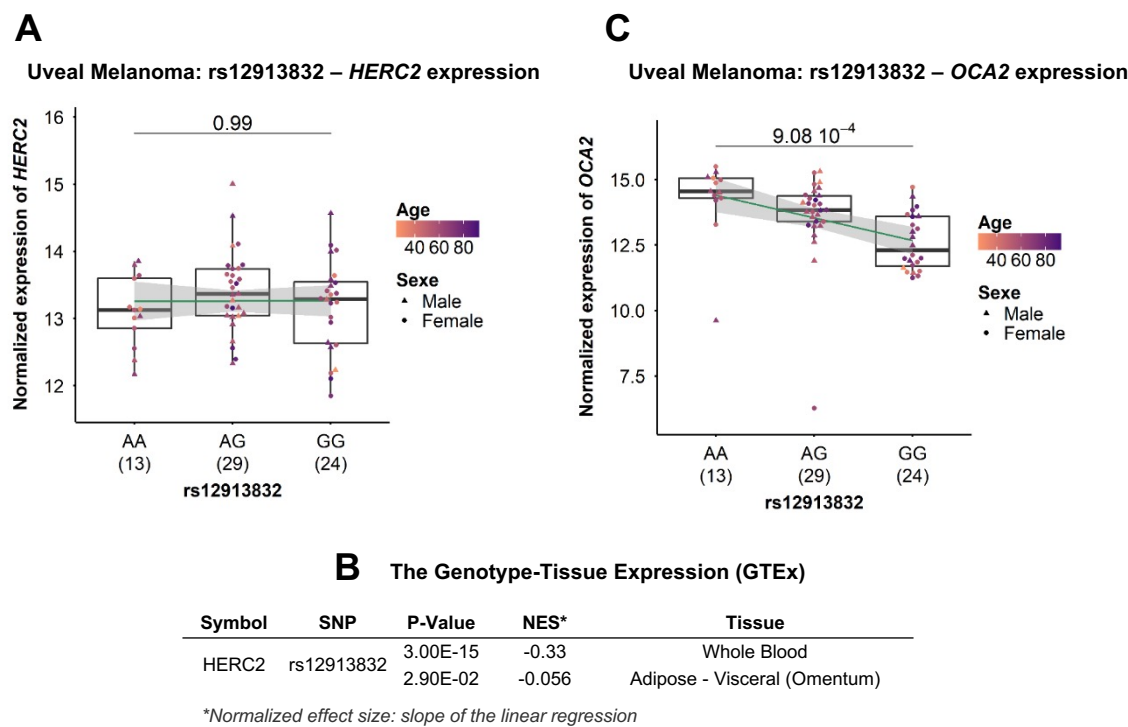
Symbol	SNP	P-Value	NES*	Tissue
IRF4	rs12203592	8.10E-14	0.34	Lung
		2.10E-12	0.22	Whole Blood
		1.50E-10	0.33	Spleen
		3.10E-08	0.36	Testis
		1.10E-07	-0.32	Skin - Sun Exposed (Lower leg)
		1.30E-06	0.27	Small Intestine - Terminal Ileum
		1.60E-06	0.31	Nerve - Tibial
		2.70E-06	0.2	Esophagus - Mucosa
		3.20E-06	0.31	Adipose - Visceral (Omentum)
		6.40E-06	0.54	Cells - EBV-transformed lymphocytes
		9.10E-06	0.18	Colon - Transverse
		2.00E-05	0.17	Stomach
		3.10E-05	0.19	Thyroid
		3.60E-03	0.16	Artery - Aorta

*Normalized effect size: slope of the linear regression

Supplementary Figure 5: Expression quantitative trait loci (eQTL) analysis for rs12203592 on *IRF4*.(A)

Analyses have been performed on normalized gene expression from RNA-seq data (EGAS00001002932) from 47 uveal melanoma tumors. (B) Data are from the Genotype-Tissue Expression (GTEx) public database (<https://gtexportal.org/home/>).

Supplementary Figure 6



Supplementary Figure 6: Expression quantitative trait loci (eQTL) analysis for rs12913832 on *OCA2*/*HERC2*. (A,C) Analyses have been performed on normalized gene expression from RNA-seq data (EGAS00001002932) from 66 uveal melanoma tumors, on *HERC2* (A) and *OCA2* (C). (B) Data are from the Genotype-Tissue Expression (GTEx) public database (<https://gtexportal.org/home/>).

DIFFERENT PIGMENTATION RISK LOCI FOR HIGH-RISK MONOSOMY 3 AND LOW-RISK DISOMY 3 UVEAL MELANOMAS

Our most novel finding in this work, and probably also the most intriguing, is the identification of two pigmentation risk loci in UM that are differentially associated with the monosomy 3 (M3) and disomy 3 (D3) UM subgroups, which respectively cluster with high and low metastatic risk, suggesting two distinct etiologies of UM genetic risk. These distinct genetic risk loci represent one additional factor that delineates the two UM subgroups, in addition to the established distinction between high-risk and low-risk UMs in terms of driver genetic alterations, DNA methylation, transcriptional features and multiple others.²⁶⁰ Our results may thus have clinical significance, as these differential genotypes could help determine clinical outcome and metastatic risk in UM. Additional studies will be required to investigate the molecular mechanisms by which *IRF4* SNPs essentially predispose to D3 “low-risk” while *HERC2* SNPs are mainly associated with M3 “high-risk” UM. We suggest that these mechanisms may act, at least partly, independently from their known role in pigmentation, especially *IRF4*.

The two pigmentation loci are characterized by SNPs on the *OCA2/HERC2* and *IRF4* loci, confirming Ferguson’s findings⁵⁵¹ where he demonstrated that UM and cutaneous melanoma (CM) share some common genetic risk regions, even if major driver events and clinical features in these two tumor types are very distinct.²²⁴ The *OCA2/HERC2* risk variants are associated with the “blue eye” phenotype previously discussed, mainly through a decreased expression of *OCA2*²⁴⁰ resulting in low eumelanin synthesis and consequently a higher pheomelanin : eumelanin ratio leading to lighter eye color. As for *IRF4*, although its function does not involve pigmentation (*IRF4* mainly acts as a transcription factor that regulates interferon-inducible genes and is involved in lymphocyte differentiation),⁶³⁰ *IRF4* rs12203592-T polymorphism (UM risk allele) has been linked to pigmentation through disruption of a binding site of TFAP2 α at the SNP position, resulting in a reduction of the tyrosinase (TYR) enzyme key to eumelanin synthesis,⁶³¹ thereby favoring the blue eye phenotype as well.

These common pigmentation markers are clearly consistent with the epidemiology of UM and CM, which are prevalent in individuals of fair skin and light eye color. However, unlike in CM for which a role for melanin synthesis (mostly dark eumelanin pigment) against ultra-violet radiation is a known protective mechanism, the role of *IRF4* and *OCA2/HERC2* polymorphisms in UM is less obvious. It is intriguing that they confer UM risk since light and ultra-violet radiation (UVR) hardly penetrate uveal melanocytes other than within the iris, and there is an absence of UVR-associated signature in UM.²⁵⁹ Our results taking into account eye color of UM patients (as determined by the IrisPlex system)²⁴² further support a role for these risk polymorphisms outside from pigmentation, especially *IRF4* that remains a significant UM risk factor when adding eye color as a covariate. Instead, the risk conferred by differences between lighter and darker eye pigmentation may rather implicate the pro-oxidant nature of pheomelanin compared to eumelanin anti-oxidant properties as a scavenger of ROS, in a UVR-independent mechanism.^{217,223,246} In the blue eye phenotype, lower levels of OCA2 protein, core melanogenic enzyme tyrosinase (TYR) is retained in the ER leading to an accumulation of the eumelanin precursor DHICA that further generates ROS and DNA damage,⁶³² which in turn may participate in the faster accumulation of somatic mutations in individuals with lower eumelanin : pheomelanin ratios and light eye color (Figure 20). Pheomelanin itself has been reported to generate ROS and other oxidants and suggested to have some UV-independent carcinogenic properties.^{217,633}

An important limitation to our study is the fact that eye pigmentation of UM patients was deduced from genotypes, some of which are also UM risk SNPs, making it challenging to assess causality between eye color and UM risk. We are currently trying to gain access to cohorts from the large-scale UK Biobank database and the Rotterdam study, for which eye color phenotype of some individuals is accessible. It would be interesting to validate our results in an independent cohort in which iris color is physically determined rather than by assessment of genotypes of pigmentation markers.

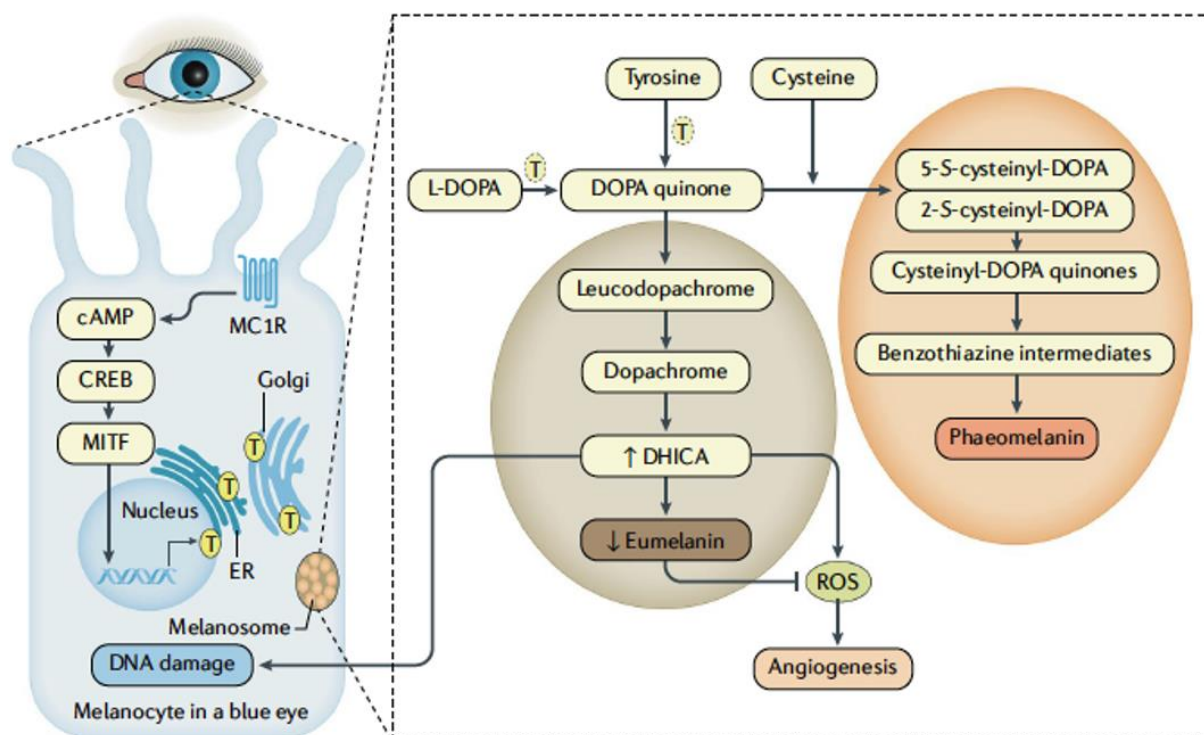


Figure 20. Low eumelanin to pheomelanin ratios and defective eumelanin synthesis may lead to the generation of reactive oxygen species and DNA damage in the eye

Yellow-circled T represents the tyrosinase enzyme retained in the ER and not properly transferred to melanosomes following reduction in OCA2 levels, leading to DHICA intermediate accumulation and generation of reactive oxygen species (ROS). Figure from Jager *et al.*²²³

Lastly, we confirmed in this work the genetic risk conferred by the *TERT/CLPTM1L* locus on chromosome 5 in UM, which was the only significant association signal identified in the first GWAS carried out by our team.⁵²⁶ Unlike pigmentation loci, the 5p15.33 susceptibility region was equally associated with D3 and M3 UM, suggesting shared oncogenic mechanisms in UM, although this risk region is not expected to explain UM epidemiology given the comparable risk allele frequencies in different populations. *TERT/CLPTM1L* is a multi-cancer risk locus, composed of multiple independent association signals and in which directionality of risk intriguingly depends on the cancer type.⁵²⁷⁻⁵²⁹ In the last part of this PhD work, we functionally characterize the *TERT/CLPTM1L* locus, in an attempt to understand its biological consequences in UM tumorigenesis. Similar studies on *IRF4* and *HERC2/OCA2* risk loci may also allow to further elucidate the mechanisms underlying genetic predisposition to UM.

INTRODUCTION TO ARTICLE 3

FUNCTIONAL CHARACTERIZATION OF CLPTM1L/TERT 5p15.33 RISK LOCUS IN UVEAL MELANOMA IDENTIFIES RS452384 AS A FUNCTIONAL VARIANT REGULATING CLPTM1L AND TERT EXPRESSION THROUGH NKX2.4 ALLELE- SPECIFIC BINDING

The initial genome-wide association study (GWAS) carried out by the team in 2017 identified one UM susceptibility region on chromosome 5p15.33 at the *TERT/CLPTM1L* locus, which was further validated in an independent cohort (combined odds ratio and p-value for lead SNPs rs421284 and rs452932 in the discovery and validation sets: OR=1.71, $p=5 \times 10^{-9}$; and OR=1.72, $p=2 \times 10^{-9}$, respectively).⁵²⁶ Risk alleles were located within intronic regions of the *CLPTM1L* gene, and displayed patterns of high linkage disequilibrium (LD) indicating that they belong to a same haploblock. In this work, our team showed by expression quantitative trait loci (eQTL) analysis that *CLPTM1L* expression was higher in individuals carrying the risk (C) allele of rs421284. No correlation with *TERT* was observed, at least partly explained by its marginal expression in UM. Our recent second GWAS (Article 2) confirmed the association of the 5p region with UM and further showed by conditional analysis that the risk locus was characterized by a unique association signal (led by rs421284, rs370348 and including other SNPs in high LD) as no other SNP was independently associated with UM when conditioning on rs421284.

The *TERT/CLPTM1L* is, as previously stated, a complex multi-cancer risk locus.⁵²⁷⁻⁵²⁹ Only few studies have functionally characterized the 5p15.33 risk signals, in cancers other than UM.^{183,543,634-636} Besides, this locus is thought to be regulated by tissue-specific processes, given the different directionalities of risk (risk and protective alleles of a same SNP are opposite in some cancers) and given the implication of various independent signals depending on the tumor type.⁵²⁷⁻⁵²⁹ We therefore aimed to decipher some of the biological mechanisms underlying the *TERT/CLPTM1L* risk region to understand their role in UM tumorigenesis. Findings from this project, which was carried out from the beginning of this PhD project until now, are described in the following prepared manuscript, which will soon be submitted for consideration in a scientific journal.

ARTICLE 3: FUNCTIONAL CHARACTERIZATION OF
CLPTM1L/TERT 5P15.33 RISK LOCUS IN UVEAL MELANOMA
IDENTIFIES RS452384 AS A FUNCTIONAL VARIANT REGULATING
CLPTM1L AND TERT EXPRESSION THROUGH ALLELE-SPECIFIC
BINDING OF NKX2.4

Abstract

Uveal melanoma (UM) is the most common intraocular malignancy in adults and major risk factors include fair skin, light eye color and European descent, suggesting for a role of genetic predisposition factors in UM. A genome-wide association study (GWAS) in UM identified a risk locus on chromosome (chr) 5p15.33 at the *TERT/CLPTM1L* locus, including multiple highly correlated alleles. Through characterization of a region within this locus enriched in active chromatin marks and enhancer elements, we functionally assessed three UM risk variants using luciferase reporter assays and identified rs452384 as a functional variant at the 5p risk locus, risk allele C was associated with higher gene expression. Electrophoretic mobility shifts assays and subsequent quantitative mass spectrometry identified NKX2.4 transcription factor as a rs452384-T specific binding protein while GATA4 interacted with both alleles and preferentially with rs452384-C. Knockdown of *NKX2.4* but not *GATA4* resulted in increased expression of *TERT* and *CLPTM1L*. Finally, relative telomere length measurement in germline DNA from peripheral blood lymphocytes (PBL) and tumor DNA of UM patients revealed a dramatic decrease in telomere length in tumor compared to PBL, and longer telomeres associated with rs452384-CC risk genotype compared to TT in PBL from UM patients and controls. Taken together, our results indicate that rs452384 is a functional variant at chr5p15.33 explaining some of the risk conferred by this locus in UM, through allele-specific protein binding, differential gene regulation and decreased *CLPTM1L* and *TERT* expression in protective T allele carriers, consistent with rs452384 CC genotype being associated with longer telomeres than rs452384 TT in DNA from PBLs. This study represents the first steps in unraveling the biological consequences of 5p15.33 genetic risk locus in UM genetic predisposition.

Functional characterization of *CLPTM1L/TERT* 5p15.33 risk locus in uveal melanoma identifies rs452384 as a functional variant regulating *CLPTM1L* and *TERT* expression through allele-specific binding

Anne-Céline Derrien¹, Alexandre Houy¹, Olivier Ganier¹, Florent Dingli², Massih Ningarhari³, Valentin Sabatet², Lenha Mobuchon¹, María Isabel Espejo Díaz¹, Damarys Loew², Jessica Zucman-Rossi^{3,4}, Nathalie Cassoux^{5,6}, Olivier Cussenot^{7,8}, Géraldine Cancel-Tassin^{7,8}, Manuel Rodrigues^{1,9}, Marc-Henri Stern^{1,*}.

¹Inserm U830, DNA Repair and Uveal Melanoma (D.R.U.M.), Equipe labellisée par la Ligue Nationale Contre le Cancer, Institut Curie, PSL Research University, Paris, 75005, France

²Institut Curie, PSL Research University, Centre de Recherche, Laboratoire de Spectrométrie de Masse Protéomique, 26 rue d'Ulm, Paris 75248 Cedex 05, France.

³Centre de Recherche des Cordeliers, Sorbonne Université, Inserm, Université de Paris, F-75006, Paris, France; Functional Genomics of Solid Tumors laboratory, Équipe labellisée Ligue Nationale contre le Cancer, Labex OncoImmunology, F-75006, Paris, France.

⁴Hôpital Européen Georges Pompidou, APHP, F-75015, Paris, France.

⁵Department of Ocular Oncology, Institut Curie, Paris, 75005, France.

⁶Faculty of Medicine, University of Paris Descartes, Paris, 75005, France.

⁷CeRePP, Tenon Hospital, Paris, 75020, France

⁸Sorbonne University, GRC n°5 Predictive Onco-Urology, AP-HP, Tenon Hospital, Paris 75020, France.

⁹Institut Curie, PSL Research University, Department of Medical Oncology, Paris, 75005, France.

* Corresponding author: marc-henri.stern@curie.fr; +33 1 56 24 66 46; 26 rue d'Ulm, 75248 Paris Cedex 05, France.

Abstract

Uveal melanoma (UM) is the most common intraocular malignancy in adults and major risk factors include fair skin, light eye color and European descent, suggesting for a role of genetic predisposition factors in UM. A genome-wide association study (GWAS) in UM identified a risk locus on chromosome (chr) 5p15.33 at the *TERT/CLPTM1L* locus, including multiple highly correlated alleles. Through characterization of a region within this locus enriched in active chromatin marks and enhancer elements, we functionally assessed three UM risk variants using luciferase reporter assays and identified rs452384 as a functional variant at the 5p risk locus, risk allele C was associated with higher gene expression. Electrophoretic mobility shifts assays and subsequent quantitative mass spectrometry identified NKX2.4 transcription factor as a rs452384-T specific binding protein while GATA4 interacted with both alleles and preferentially with rs452384-C. Knockdown of *NKX2.4* but not *GATA4* resulted in increased expression of *TERT* and *CLPTM1L*. Finally, relative telomere length measurement in germline DNA from peripheral blood lymphocytes (PBL) and tumor DNA of UM patients revealed a dramatic decrease in telomere length in tumor compared to PBL, and longer telomeres associated with rs452384-CC risk genotype compared to TT in PBL from UM patients and controls. Taken together, our results indicate that rs452384 is a functional variant at chr5p15.33 explaining some of the risk conferred by this locus in UM, through allele-specific protein binding, differential gene regulation and decreased *CLPTM1L* and *TERT* expression in protective T allele carriers, consistent with rs452384 CC genotype being associated with longer telomeres than rs452384 TT in DNA from PBLs. This study represents the first steps in unraveling the biological consequences of 5p15.33 genetic risk locus in UM genetic predisposition.

Introduction

Uveal Melanoma (UM) is a rare cancer but the most frequent type of intraocular malignancy in adults, arising from the malignant transformation of melanocytes from the uveal tract composed of the choroid, ciliary body and iris, with a large predominance for choroidal UM. In 30% to 50% of cases, the disease metastasizes, almost invariably to the liver,^{1,2} associated with dismal prognosis characterized by a median survival of 9–12 months and, until very recently, a lack of effective therapeutic treatment. UM is a genetically simple tumor, with two main oncogenic driver events characterized by the constitutive activation of the $G\alpha_q$ pathway through mutually exclusive mutations of *GNA11*, *GNAQ* or more rarely *CYSLTR2* or *PLCB4* genes,^{3,4} and a second event where almost mutually exclusive mutations of *BAP1*, *SF3B1* or *EIF1AX* define UM subgroups with high, middle and low metastatic risk, respectively.⁵⁻⁷

UM represents 4 to 5% of all melanoma types and is very distinct from cutaneous melanoma on multiple aspects, including oncogenic driver events and clinical manifestations.⁸ UM has an incidence rate of 5.6 cases per million persons per year and, strikingly, is more frequent in populations of European ancestry compared to that of Asian or African-American origins, with a 10 to 20-fold lower incidence in these two populations compared to those of European ancestry.^{9,10} In addition, meta-analyses have demonstrated a prevalent association of UM with light eye color and fair skin,¹¹ which initially pointed towards a potential role for exposure to sunlight and associated ultraviolet (UV) radiation in UM, corroborated by the existence of pigmentation genetic markers as two UM risk loci.^{12,13} However, the stable incidence in UM observed in the last decade¹⁴ despite the rise in skin melanoma incidence, the very low tumor mutation burden in UM and lack of associated UV mutational signature⁶ (except for iris melanoma) rule out a role for pigmentation traits protecting against UV radiation to explain this epidemiology, and suggest the existence of genetic predisposing factors in UM. Besides, rare germline pathogenic variants of *BAP1* are seen in UM⁵ but only explain a fraction of familial cases of UM, suggesting the presence of additional genetic risk factors in UM.

We conducted the first genome-wide association (GWAS) in UM on 259 UM patients and 401 controls of European ancestry and identified a UM susceptibility region on chromosome (chr) 5p15.33 on the *TERT/CLPTM1L* locus,¹⁵ led by rs452932 and rs421284 (odds ratio OR=1.71 and 1.72 respectively, confidence interval CI [1.43–2.05] and [1.44–2.06] respectively) in high linkage disequilibrium (LD) with one another ($r^2 > 0.9$) and with multiple SNPs in the same region spanning *CLPTM1L* introns. This prompted us to investigate the biological consequences of the *CLPTM1L/TERT* risk locus in UM. We recently published a second phase of GWAS in UM with genetic imputation, combining the first GWAS with two

additional datasets, extending the initial series to a total of 1,142 UM patients and 882 healthy controls.¹³ This study confirmed the association of 5p15.33 *TERT/CLPTM1L* locus with UM, led by rs370348 (OR= 1.59 [1.35; 1.86]), part of the same haploblock as rs421284, and identified two additional risk regions involved in pigmentation traits, on *IRF4* (chr6) and *HERC2/OCA2* (chr15) loci, confirming Ferguson's candidate SNP analysis in UM that demonstrated shared etiologic factors between UM and cutaneous melanoma (CM).¹² Another GWAS was recently performed by another team on 590 UMs, but no locus reached statistical significance.¹⁶

The 5p15.33 *TERT/CLPTM1L* locus has been extensively characterized as a multi-cancer risk locus, both in cancer-specific association studies and in meta-analyses.¹⁷⁻²⁰ Well over 10 tumor types in addition to UM have been reported for cancer association at the 5p15.33 locus, including breast (both specific subtypes and overall),²¹ colorectal,²² lung,²³ pancreatic,²⁴ prostate,²⁵ renal,²⁶ ovarian,²⁷ head and neck,²⁸ esophageal,²⁹ endometrial,³⁰ testicular germ cell³¹ cancers as well as cutaneous melanoma^{32,33} and glioma.³⁴ Importantly, up to ten independent risk loci have now been identified within this genomic region of almost 1Mb encompassing both *CLPTM1L* and *TERT* genes,¹⁷ extending the list of seven distinct association loci previously reported by Wang and colleagues;¹⁹ in addition, some alleles have been associated with multiple cancers with different directionality of effect (risk or protective alleles) depending on the tumor type,^{17,19,35} suggesting tissue-specific regulation of this multi-cancer risk locus. Conditional analyses in our second GWAS study revealed that no secondary independent locus was associated with UM risk other than the identified locus marked by rs421284. Notably, rs421284 is in high LD ($r^2 = 0.93$) with rs401681, another UM risk variant that also marks a shared association locus (termed 'Region 2' in Wang *et al*'s study)¹⁹ with lung, pancreatic, bladder cancers and CM, the latter for which the C allele is protective, as in UM and in pancreatic cancer, opposite to lung and bladder carcinomas for which C is the risk allele.^{20,36,37} This cross-cancer association signal was recently confirmed by Cheng and colleague¹⁷ where index SNP rs465498 marks one of ten shared independent signals at 5p15.33 region, and this 10th region is associated with risk of lung cancer and inverse risk for melanoma and pancreatic cancer (as in UM).

Although the *CLPTM1L/TERT* locus is a genetic predisposing region to many cancer types, little is known about the biological contribution of variants underlying these risk loci leading to cancer, especially in UM. Functional genomics studies, or "post-GWAS" analyses, aim to decipher the mechanisms underlying a risk locus, and identify among the numerous genetically linked polymorphisms, causal variants that exert active roles in tumorigenic processes, shedding light on cancer etiology. The two plausible gene candidates at the

5p15.33 UM risk locus are *CLPTM1L*, harboring UM risk SNPs in its intronic sequence, and *TERT*, sharing the same locus and frequently reported to play a role in cancer, and particularly melanoma.^{20,38} *CLPTM1L*, which encodes the cleft lip and palate associated transmembrane 1-like protein, is overexpressed in pancreatic and lung cancer, where it promotes growth and survival.^{39,40} The protein has also been shown to confer resistance to chemotherapy by regulation of Bcl-xL apoptotic factor in lung cancer.⁴¹ In UM, risk haplotype of the *CLPTM1L* locus is associated with an increase in *CLPTM1L* expression.¹⁵ *TERT*, on the other hand, encodes the catalytic subunit of the telomerase enzyme, which elongates telomeres that protect the end of chromosomes to maintain genome integrity.⁴² Telomere maintenance has been extensively linked to cancer; in some tumor types, cancer cells continue to divide despite having reached critically short telomere length (which normally results in cell senescence to prevent genomic instability), either by upregulating *TERT* or by activating the alternative lengthening of telomeres mechanisms (ALT), thereby creating genomic instability and abnormal cell survival, promoting cancer development. Highly recurrent mutations in the promoter sequence of *TERT* are seen in cutaneous melanoma.³⁸ The *TERT* locus has frequently been identified for conferring tumor risk in many cancers;²⁰ notably, Fang and colleagues reported a role for a SNP intronic to *CLPTM1L*, rs36115365, in the regulation of *TERT* but not *CLPTM1L* in pancreatic, testicular, melanoma and lung cancers.¹⁸

We sought to functionally characterize the UM risk region underlying the *CLPTM1L/TERT* locus on chr5p15.33, as the oncogenic mechanisms behind this predisposing locus remain unknown in UM. We investigated the presence of active chromatin marks within the genomic region harbouring UM risk SNPs, functionally tested candidate risk variants for allele-specific activity and identified rs452384 as a functional variant in UM, characterized by allele-specific binding of nuclear factors and transcriptional regulation. We present here the functional investigation of a sub-region of the 5p15.33 risk locus in UM, and the molecular mechanisms underlying rs452384 allele-specific regulation of *TERT* and *CLPTM1L*.

Materials and Methods

Subjects

Subjects that were part of the initial GWAS study that first identified the chr5 *TERT/CLPTM1L* uveal melanoma (UM) risk locus include 259 UM patients and 401 controls of European ancestry (“dataset 1”).¹⁵ Genotyping data of cases are deposited on the European Genome-Phenome Archive (EGA) (accession number EGAS00001002334) while those of controls are accessible under dbGaP Study Accession number phs001271.v1.p1). A second association study, extending our initial GWAS study to a total of 1,142 European UM patients treated at Institut Curie from 2013 to 2018 and 882 healthy controls (datasets 1 initially described, combined with datasets 2 and 3) with imputation, further recapitulated *CLPTM1L/TERT* risk locus.¹³ Genotyping data of cases from datasets 2 and 3 are available on EGA under EGAS0000100233. The study was approved by the Ethical Committee and Internal Review Board at the Institut Curie, and all patients and controls consented to participate to the study. Controls are of French origin and were selected from the KIDRISK consortium (US NCI U01CA155309; G. Scelo) as part of a parallel GWAS on renal cell carcinoma. For relative telomere length measurement, a subpopulation from our GWAS metanalysis was randomly chosen, including germline DNA of 480 UM patients (from datasets 1, 2 and 3) and 120 controls from the KIDRISK study. Tumor DNA samples from an in-house additional series of 72 UM patients diagnosed at Institut Curie and treated by primary enucleation were also included in the analysis (previously published cohort described in Alsafadi *et al*⁴³ and expression data deposited on EGA under no. EGAS00001002334). Characteristics of patients included in this telomere length analysis (age, sex, rs452384 genotype) are provided in Supplementary Table 1.

Cell lines and culture

All cell lines described in this study are derived from UM. MP41 and MM66 cell lines were derived at Institut Curie and described in Amirouchene-Angelozzi *et al*⁴⁴. Mel202 cell line was purchased from the European Searchable Tumour Line Database (Tubingen University, Germany). OMM1 and OMM2.5 (also called OMM1.5) cell lines were kindly provided by P.A. van der Velden (Leiden University Medical Center, The Netherlands). Cell lines were cultured in RPMI 1640 (Gibco) supplemented with 10% Fetal Bovine Serum (FBS) (MP41, Mel202, OMM1, OMM2.5) or 20% FBS (MM66). Authentication of cell lines was verified by RNA-Seq / verified by short tandem repeat (STR) test. Cells were regularly tested for Mycoplasma and found to be negative each time.

RNA and DNA extraction

RNA and genomic DNA extraction from UM cell lines were performed using Qiagen's RNeasy Plus Mini Kit and DNeasy Blood & Tissue kit respectively, following manufacturer's instructions. RNA and DNA quality and yield were obtained using a Nanodrop 2000 spectrophotometer. Tumor DNA and RNA samples (priorly extracted) from UM patients were further purified from melanin and other contaminants using the OneStep PCR Inhibitor Removal Kit (Zymo Research).

Cloning and luciferase assay

The genomic region surrounding rs452384 and other UM risk SNPs with high Odds Ratio on chr5p15.33 (*CLPTM1L* introns 7 and 8) (Figure 1) was PCR amplified from UM MP41 cell line, which is heterozygous (CT) for rs452384 and other UM risk SNPs in high linkage disequilibrium (LD).¹⁵ In-Fusion online tool and HD Cloning Plus kit were used to design primers used for PCR amplification of the 3.04 kb fragment ("full vector" fragment on Figure 1) and to subsequently subclone the insert into a pGL3-Promoter (pGL3-P) Firefly luciferase reporter vector (Promega) via its *BglIII* site (vector designed to clone potential enhancers upstream of an SV40 promoter driving luciferase expression). The same technique was used in the deletion analysis to further define the enhancer region (fragments A to G of the full length insert, Figure 1A&D). All primers are listed in Supplementary Table 2. Plasmids were amplified into Sure2 supercompetent cells (Agilent) and verified at the insert position by Sanger sequencing, further allowing to select clones of both genotypes (high-risk C and low-risk T for rs452384). Plasmids were then co-transfected with pRL-CMV Renilla luciferase control vector (Promega) in UM cell lines (MP41 and OMM2.5) using Lipofectamine 2000 reagent (ThermoFisher Scientific) following manufacturer's instructions. Cells were collected 48h post-transfection and luciferase activity was measured using the Dual Luciferase Reporter Assay System (Promega). Luminescence signal from Firefly luciferase was normalized to that of Renilla luciferase, and resulting normalized luminescence was compared to that of transfection with an empty pGL3-P vector. Experiments were performed in triplicates and repeated at least 5 times. Two-tailed t-tests were used to assess statistical significance of differences in luciferase activity.

Site-directed mutagenesis

Site-directed mutagenesis was performed on the full-length pGL3-P-*CLPTM1L* vector (*CLPTM1L* intronic enhancer sequence described above) to evaluate the individual effects of the 3 SNPs identified in the UM GWAS within the insert sequence: rs452932T/C, rs452384T/C and rs370348A/G. The QuickChange II Site-directed mutagenesis kit (Agilent) was used to

specifically mutate the full-length pGL3-P-*CLPTM1L* vector at the 3 individual positions, using 35 to 45bp mutagenic primers (Supplementary Table 2). After mutant strand synthesis using *PfuUltra* HF DNA polymerase, resulting *DpnI*-treated DNA was transformed into XL10-Gold ultracompetent cells (Agilent). This process was repeated on both genotypes (high and low risk) of the pGL3-P-*CLPTM1L* vector to obtain clones with all possible combinations of the 3 SNPs genotypes and validate in both directions (high-to-low risk and low-to-high risk) the effect of changing the SNPs' alleles. Correct sequences were verified by Sanger sequencing of the mutant clones. Resulting vectors were assessed for luciferase activity described above.

EMSA

Nuclear extracts from UM cell lines (MP41, OMM1, OMM2.5, Mel202 and MM66) were generated using the NE-PER Nuclear and Cytoplasmic Extraction kit (78833, Thermo Scientific) and their concentration was measured using BCA Protein Assay kit (23225, Pierce) according to the manufacturer's instructions. Forward and reverse 25bp oligos biotinylated at their 3' end, centered around either rs452384 or rs452932 and for both alleles (listed in Supplementary Table 2), were ordered from Eurofins Genomics along with unlabeled oligos with identical sequence (specific competitors). Forward and reverse primers were then annealed in TE buffer (10mM Tris, 1mM EDTA, 50mM NaCl) by heating at 95°C for 5 minutes in a thermoblock and letting it cool down to room temperature over 2-3 hours to create double stranded probes (3'biotin-labelled and unlabelled) harbouring the DNA sequence around both SNPs and for both alleles of these SNPs. EMSAs were performed using the LightShift Chemiluminescent EMSA Kit (20148 Thermo Scientific). Briefly, double stranded labelled DNA probes were incubated with 5µg of nuclear extracts for 30 minutes. Competition experiments were performed by adding 10 to 100-fold excess of unlabelled probe (specific competitor) compared to labelled probe. 1µg/µL poly dI/dC was used as a nonspecific competitor. EMSA reactions were ran on a 5% TBE polyacrylamide gel (Bio-Rad), transferred to Biodyne B Nylon membranes (Thermo Scientific), crosslinked at 120mJ/cm² using a Stratagene UV Stratalinker 1800, and detected by chemiluminescence using a Stabilized Streptavidin-Horseradish Peroxidase conjugate and a Luminol/Enhancer substrate solution. For supershift assays using FLAG-tagged proteins, MP41 cells were first transfected with pcDNA3.1(+)-C-(K)DYK (C-terminus FLAG tag) vectors expressing 4 candidate proteins (NKX2.4, GATA4, DLX6 or PITX2) or empty vector (ordered from GenScript). Protein overexpression was verified by Western Blotting using monoclonal anti-Flag M2 antibody (Sigma F1804), anti-Nkx2.4 antibody (abcam ab189202), anti-GATA4 antibody (Invitrogen MA5-15532), and anti-Histone H3 antibody (abcam ab1791) as a nuclear loading control. Cells were transfected using lipofectamine 2000 following manufacturer's instructions, and nuclear proteins were extracted

72 hours post-transfection. Supershift experiments were carried out by adding 2 μ g monoclonal anti-Flag M2 antibody (Sigma F1804), or 2 μ g normal mouse IgG (Santa Cruz sc2025) antibody as a negative control, to the nuclear extracts 10 minutes prior to adding labelled probes.

DNA pulldown assay and quantitative mass spectrometry

Sample preparation. To identify proteins that recognize and interact with a binding motif around rs452384, biotinylated-DNA pulldown assays were carried out using Dynabeads M-280 Streptavidin (11205D Invitrogen), following the manufacturer's instructions and as previously published⁴⁵. The same 25bp biotin-labeled forward primers described in the EMSA protocol (12bp on either side of rs452384-T or rs452384-T) were annealed to complementary unlabeled reverse primers as previously described, creating double stranded probes biotinylated on one end. A 3'biotin-labeled negative control probe that does not contain any known DNA binding motif was also included as a nonspecific binding probe (Supplementary Table 2). 150 μ g of streptavidin-coated beads were incubated with 2.25 μ g of double stranded DNA-biotin probes prior to adding freshly prepared MP41 nuclear extracts (1mg per reaction). DNA-protein complexes on beads were incubated for 45 minutes at room temperature in EMSA binding buffer supplemented with salt and spiked with poly(dI/dC) (50mM KCl, 10mM Tris, 0.1% NP-40, 1mM DTT, 1X protease and phosphatase inhibitors, 50mM NaCl, 10ng/ μ L poly(dI/dC)). Complexes on beads were washed 3 times in EMSA buffer followed by 3 times in 25mM Ammonium Bicarbonate (NH₄HCO₃), pulling down beads on a DynaMag Magnet each time.

Beads were resuspended in 100 μ L of 25 mM NH₄HCO₃ and digested by adding 0.2 μ g of trypsin-LysC (Promega) for 1 h at 37 °C. Samples were then loaded into custom-made C18 StageTips packed by stacking one AttractSPE® disk (#SPE-Disks-Bio-C18-100.47.20 Affinisep) and 2mg beads (#186004521 SepPak C18 Cartridge Waters) into a 200 μ L micropipette tip for desalting. Peptides were eluted using a ratio of 40:60 CH₃CN:H₂O + 0.1% formic acid and vacuum concentrated to dryness with a SpeedVac apparatus. Peptides were reconstituted in injection buffer in 0.3% trifluoroacetic acid (TFA) before liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis. Each of the 3 conditions (negative control probe NEG, rs452384-T probe, rs452384-C probe) was performed in 5 biological replicates.

LC-MS/MS Analysis. Online chromatography was performed using an RSLCnano system (Ultimate 3000, Thermo Fisher Scientific) coupled to an Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific). Peptides were trapped on a C18 column (75 μ m inner

diameter × 2 cm; nanoViper Acclaim PepMap 100, Thermo Fisher Scientific) with buffer A (2:98 CH₃CN:H₂O in 0.1% formic acid) at a flow rate of 3.0 µl/min over 4 min. Separation was performed using a 40 cm × 75 µm C18 column (Reprosil C18, 1.9 µm, 120 Å, Pepsep PN : PSC-40-75-1.9-UHP-nC), regulated to a temperature of 40 °C with a linear gradient of 3% to 32% buffer B (100% CH₃CN in 0.1% formic acid) at a flow rate of 150 nl min⁻¹ over 91 min. Full-scan MS was acquired using an Orbitrap Analyzer with the resolution set to 120,000, a scan range of 350-1550 m/z and ions from each full scan were fragmented by higher-energy C-trap collisional dissociation (HCD) mode and analyzed in the linear ion trap in rapid mode. We selected ions with charge state from 2+ to 6+ for screening. Normalized collision energy was set to 28, AGC target to 10,000 and the dynamic exclusion of 30s.

Data analysis. For identification, the data were searched against the Homo sapiens UP000005640 database (downloaded 12/2019 containing 20364 entries) using Sequest HT through Proteome Discoverer (version 2.4). Enzyme specificity was set to trypsin and a maximum of two missed cleavage sites were allowed. Oxidized methionine, N-terminal acetylation, methionine loss and methionine acetylation loss were set as variable modifications. Maximum allowed mass deviation was set to 10 ppm for monoisotopic precursor ions and 0.6 Da for MS/MS peaks. The resulting files were further processed using myProMS⁴⁶ v.3.9.2. False-discovery rate (FDR) was calculated using Percolator⁴⁷ and was set to 1% at the peptide level for the whole study. Label-free quantification was performed using peptide extracted ion chromatograms (XICs), computed with MassChroQ⁴⁸ v.2.2.21. For protein quantification, XICs from proteotypic peptides shared between compared conditions (TopN matching) with missed cleavages were used. Median and scale normalization at peptide level was applied on the total signal to correct the XICs for each biological replicate (N=5). To estimate the significance of the change in protein abundance, a linear model (adjusted on peptides and biological replicates) was performed, and p-values were adjusted using the Benjamini–Hochberg FDR procedure. For comparison of differential enrichment of proteins bound to NEG, rs452384-T or rs452384-C probes (3 datasets), proteins with at least 3 distinct peptides in all replicates (n=5), an adjusted p-value ≤ 0.05 and an 2-fold and 1.5-fold enrichment (fold change ≥ 2 for rs452384-C or -T vs. NEG probes; absolute fold change ≥ 1.5 for rs452384-T vs. -C probes for a more conservative approach) were considered significantly enriched in sample comparisons. Unique proteins were considered with at least three total peptides in all replicates. Proteins selected with these criteria were further analyzed; all proteins enriched over NEG dataset and preferentially bound to rs452384-C or -T are listed in Supplementary Table 3. For selection of functionally relevant proteins enriched in either rs452384-T or -C (Table 1), only transcription factors were kept for further functional analysis. The mass spectrometry proteomics data have been deposited to the ProteomeXchange

Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository.⁴⁹

ChIP

Chromatin immunoprecipitation (ChIP) experiments were carried out using the iDeal ChIP-Seq kit for Transcription Factors (Diagenode C01010055) according to the manufacturer's instructions, using MP41 cells (UM cell line with a rs452384 CT heterozygous genotype). Due to the absence of ChIP-grade antibody for NKX2.4, cells were first transfected with a pcDNA3.1(+)*NKX2.4*-HA expression vector (GenScript) using lipofectamine 2000 (ThermoFisher Scientific) and following the manufacturer's instructions. ChIP experiments were carried out using 12 million cells per condition, 48 hours after transfection for NKX2.4 ChIP. 4 µg of anti-HA tag antibody (ChIP grade, abcam ab9110), 4 µg of anti-GATA4 antibody (Invitrogen MA5-15532) or 4 µg of rabbit IgG as a nonspecific control (Diagenode C15410206) were used for individual ChIP experiments on sheared chromatin from 4 million cells. IP'd DNA was then analyzed by qPCR using SYBR Green PCR Master Mix (Applied Biosystems 4309155) and qPCR primers designed to amplify rs452384 target region compared to other sites within *CLPTM1L* gene (Supplementary Table 2). Enrichment was calculated relative to input DNA. Enrichment in NKX2.4-HA transfected samples was compared to empty vector and normalized on IgG levels. To quantify the enrichment of rs452384 C and T alleles in GATA4 and NKX2.4-HA immunoprecipitated DNA samples, a Taqman genotyping assay for rs452384 was used (C__1150760_10, ThermoFisher Scientific) in a droplet digital PCR (ddPCR) experiment using ddPCR Supermix for probes (No dUTP) (Bio-Rad) and following manufacturer's instructions (see end-point droplet-digital PCR below). Allele-specific enrichment ratio was determined from the ratio of concentration (copies/µL) of FAM (T) and VIC (C) positive droplets respectively, which was normalized to input chromatin DNA. ddPCR was repeated on three independent ChIP experiments. For H3K27ac ChIP on MP41 cell line, essentially the same protocol was used except that the iDeal ChIP-Seq kit for histones (Diagenode C01010051) was used, with 4 µg of H3K27ac antibody (Diagenode C15410196) and 4 µg of negative control IgG antibody described earlier).

Knockdown and overexpression studies of *NKX2.4* and *TERT* mRNA

For knockdown studies, ONTARGETplus SMARTpool siRNAs targeting *NKX2.4* (Cat# L-036505-01-0005), *GATA4* (Cat# L-008244-00-0005) and a negative control siRNA (Non-targeting control siRNA #2, Cat# D-001810-02-05) were purchased from Horizon Discovery. UM cell lines of the 3 genotypes for rs452384 (Mel202 TT, MP41 TC, OMM1 CC) were then transfected with siRNAs at a final concentration of 2.5nM with DharmaFECT 1 transfection

reagent (Horizon Discovery T-2001-02) as per manufacturer's specifications. For overexpression studies, pcDNA3.1(+)*NKX2.4*-HA and pcDNA3.1(+)*GATA4*-HA expression vectors (GenScript) were used to transfect above-mentioned cell lines with lipofectamine 2000 transfection reagent (ThermoFisher Scientific). Cells were collected 48h post-transfection and total RNA was extracted using RNeasy Plus Mini kit (Qiagen) following manufacturer's instructions. RNA was quantified and assessed for quality using a Nanodrop 2000 spectrophotometer subsequently reverse-transcribed to cDNA using the High-Capacity cDNA Reverse Transcription kit (Applied Biosystems), following manufacturer's protocol.

Real-time quantitative PCR and end-point droplet digital PCR

Following knockdown experiments, SYBR Green qPCR was used to assess knockdown efficiency of the two targeted genes was verified with primers targeting *NKX2.4* and *GATA4*, which were normalized to expression levels of two housekeeping genes *GAPDH* and *TBP* (Supplementary Table 2) and compared to negative control siRNA (siCTL). Due to the very low expression of *TERT* in UM, droplet digital PCR (ddPCR) was used to assess expression levels of *TERT* and *CLPTM1L* normalized to that of housekeeping gene *GUSB* using Bio-Rad ddPCR technology and following manufacturer's instructions. Taqman Gene Expression assays targeting either *TERT* (FAM-labelled, Cat# Hs00972650_m1) or *CLPTM1L* (FAM-labelled, Cat# Hs00363947_m1) in combination with *GUSB* (VIC-labelled, Cat# Hs00939627_m1) were used in PCR reactions containing ddPCR Supermix for probes (No dUTP) (Bio-Rad). Water-emulsion droplets were generated using Bio-Rad QX100 Droplet Generator, and resulting mixes were read in a QX100 Droplet Reader after PCR on a C1000 Touch thermocycler. ddPCR data was analyzed with Bio-Rad QuantaSoft™ software where fluorescence from channel 1 (FAM) was compared to that from channel 2 (VIC) based on the number of FAM- and VIC- single-positive droplets (fitted to a Poisson algorithm), representing starting concentration of target DNA molecule (in units of copies/μL input) for each fluorophore.

Telomere length measurement and genotype-telomere length association

Relative telomere length in germline DNA (from peripheral blood lymphocytes, PBL) of 480 UM patients and 120 healthy individuals (GWAS controls from the KIDRISK study) and in 72 tumor DNA samples (cohorts described in "Subjects") was measured by quantitative PCR using Cawthon's gold standard method⁵⁰ adapted to correct amplification efficiency using a method developed by Ningarhari and colleagues, TeloPCR.⁵¹ Primers used to amplify *TERT* and *HMBS* single copy reference gene are listed in Supplementary Table 2. Reactions were carried out with 20ng of genomic or tumor DNA, 0.1μM of each primer and SYBR Green Master Mix (Life Technologies) and ran on a 7900HT Fast Real Time PCR system (Life

Technologies). Cycling conditions were as follows: 10min at 95°C (denaturation), 40 cycles of 15s at 95°C and 1 min at 60°C. qPCR ΔC_t values were corrected for amplification efficiency using Ramakers *et al* method⁵² as described in Ningarhari *et al*. Samples for which C_t was undetermined for either telomere repeats or *HMBS* due to technical failure were removed from the analysis. Relative telomere length (fold change) was calculated relative to *HMBS* single copy gene using the $2^{-\Delta C_t}$ method. Median telomere length in the UM germline DNA group (n=453) was compared to that of UM tumor DNA cohort (n=67), using a non-parametric Mann-Whitney test. To further assess association between rs452384 genotype and relative telomere length (tQTL), samples were then stratified according to their rs452384 genotype (TT, CT, CC). Genotypes for rs452384 were retrieved from GWAS data (for UM germline cohort and KIDRISK controls) or were obtained using a Taqman genotyping assay (Cat#C__2396811_10) for the UM tumor cohort (72 tumor DNA samples). Because the Kidrisk control cohort (n=120) only contained germline PBL samples from male individuals (see patient characteristics in Supplementary Table 1), only male samples were selected in the germline PBL UM cohort to obtain homogeneous populations without gender confounding factor for telomere length comparison (n=326 PBL samples for UM and controls, Figure 5). Age-adjusted difference in telomere length between the 3 genotypes was then measured using a generalized linear additive model (GLM) adjusted with age (Figure 5; Supplementary Figure 7). To assess pairwise differences in telomere length and for all patients (male and female) non-parametric tests were also used: 3-way non-parametric Kruskal-Wallis Rank Sum test and pairwise Wilcoxon Rank Sum test (for pairwise comparison of CC vs TT, CT vs CC and CT vs TT) (Supplementary Figure 7).

Results

Characterization and regulatory activity of the chr5p15 *CLPTM1L*/*TERT* UM risk locus

Our previously published GWAS in uveal melanoma (UM) identified a susceptibility region on chromosome (chr) 5p15.33 *TERT*/*CLPTM1L* locus, led by two SNPs with the highest Odds Ratio (OR) and associated p-value, rs421284 (OR=1.95 [1.11–3.44], $p=7\times10^{-8}$) and rs452932 (OR=1.91 [1.10–3.30], $p=1.1\times10^{-7}$) (Figure 1A).¹⁵ These 2 SNPs are in high linkage disequilibrium (LD; $r^2 > 0.9$) with each other and with multiple surrounding SNPs within *CLPTM1L* introns, where the LD pattern is consistent with degradation of LD around rs421284 as distance from the center of the haploblock increases.¹⁵ The two other SNPs in highest LD with lead SNPs and with highest OR values are rs452384 and rs465498 (Figure 1A, SNPs in blue) while other associated SNPs in this region but with lower LD and OR are also present (Figure 1A, in grey). The overall UM risk region spans the entire length of the *CLPTM1L* gene, with lead SNPs in introns 8, 9 and 10, approximately 15kb downstream from the *CLPTM1L* promoter and 35kb upstream from the *TERT* promoter. To characterize this region and prioritize likely functional variants, we searched the regulatory marks surrounding UM risk SNPs annotated by the ENCODE project on the UCSC Genome Browser (Figure 1A). Interestingly, rs452932 and rs452384, UM risk variants among those with highest OR and p-values, are located in the middle of a gene regulatory region marked by H3K27ac and H3K4me1, major histone marks flanking enhancer elements. This region is also associated with DNase I hypersensitivity clusters, indicating a region of open chromatin accessible to transcription factors and other regulatory elements, in agreement with the binding of multiple transcription factors (ENCODE 3 ChIP-Seq clusters) at this site (Figure 1A). These observations prompted us to further characterize the *CLPTM1L* intronic region containing the strongest regulatory marks, which harbours UM risk SNPs rs452932 and rs452384 (region marked in red on Figure 1A).

We first sought to assess whether the region that overlaps active chromatin and histone modification marks in the ENCODE database could also be a gene regulatory region in UM models. To do so, we performed chromatin immunoprecipitation (ChIP) for H3K27ac mark in MP41 UM cell line followed by quantitative PCR (qPCR) on the region around rs452384, directly flanked by two strong H3K27ac signals and within a DNase I cluster in ENCODE database (Figure 1A). UM risk region centered around rs452384 displayed a 3.5– to 4–fold higher enrichment over input DNA compared to negative IgG control and compared to two regions within *CLPTM1L* outside of the mark displaying gene regulatory activity close to exons 3 and 4 (primers listed in Supplementary Table 2) (mean % enrichment $\pm$ SD

H3K27ac: rs452384, 2.803 ± 0.830 ; neg1 region: 0.690 ± 0.278 , neg2 region: 0.801 ± 0.166) (Figure 1B), suggesting that the *CLPTM1L* region may influence risk in UM through mechanisms involving gene regulation and modulation of enhancer activity. To address this hypothesis, we subcloned upstream of a pGL3-promoter reporter vector (pGL3-P) driving luciferase expression the intronic regulatory region of ~3kb (red mark on Figure 1A) of both the low-risk (LR) and high-risk (HR) haplotypes from MP41, a UM cell line heterozygous for rs452384 (low-risk T and high-risk C alleles) and other highly linked UM risk SNPs. The *CLPTM1L* regulatory region demonstrated a haplotype-specific increase in luciferase activity in UM cell lines, characterized in MP41 by a 7.8-fold ± 1.306 (range 5.38 – 8.44) mean $\pm$ SEM enrichment over pGL3-P empty vector for LR haplotype, and a 21-fold ± 1.932 (range 15.53 – 28.34) for HR haplotype over empty vector, and in OMM2.5 cell line a 3.3-fold ± 0.337 (range 1.88 – 4.38) in LR over empty vector, versus 6.4-fold ± 0.635 (range 3.91 – 8.52) for HR haplotype over empty vector (Figure 1C). Across all experiments, the average fold change of luciferase activity in the HR/LR haplotypes was 2.85 (range 2.05 – 3.66) in MP41 cells and 1.90 (range 1.72 – 2.13) in OMM2.5 (Figure 1C). Taken together, these results confirm that the *CLPTM1L* genomic region is part of an enhancer that induces transcriptional activity in a haplotype-specific way in UM, with a significant 2- to 3-fold higher increase in luciferase expression with HR haplotype compared to LR (MP41 mean $\pm$ SD: 2.65 ± 0.62 , unpaired two-tailed t-test $p_{LRvsHR} = 0.0002$; OMM2.5: 1.90 ± 0.18 , $p_{LRvsHR} = 0.002$).

Allele-specific regulatory effects are mediated by rs452384, which preferentially binds nuclear proteins

To narrow down the 3kb regulatory region showing haplotype-specific enhancer activity, we performed a deletion analysis to better define the enhancer region and to determine which enhancer sub-region confers the most differential gene regulation between the two haplotypes. To do so, we trimmed down the total 3kb region (“full vector” in Figure 1A) to smaller fragments ranging from ~250bp to 2.5kb starting from either extremities (5’ and 3’) of the full insert DNA sequence (regions A to G, Figure 1A bottom panel), we generated pGL3-P-luciferase vectors harboring the different fragments upstream of their promoter and determined the fold change luciferase activity between the high and low risk haplotypes (HR/LR relative luciferase activity). As shown, constructs E and F mediate the most haplotype-specific luciferase activity in OMM2.5 and MP41 cells respectively, with an average 5.88-times (range 2.70 – 7.54) higher expression with the HR haplotype compared to LR in MP41 cell line (construct F) and 2.25-times higher (range 1.61 – 3.39) in OMM2.5 (construct E), compared to construct D which induces gene expression equally in both haplotypes in both cell lines (Figure 1D). This finding is consistent with the loss of differential luciferase activity when trimming down construct A to construct B. These results suggest that important

A

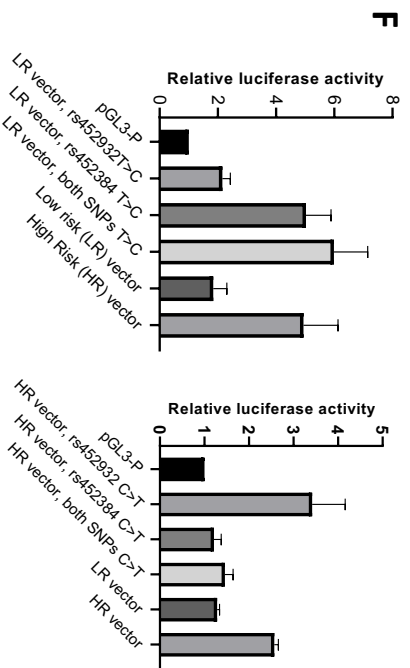
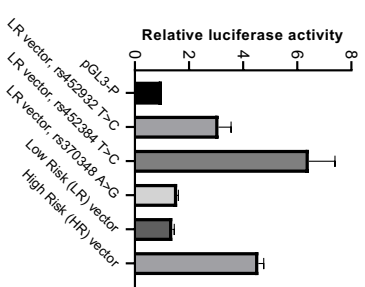
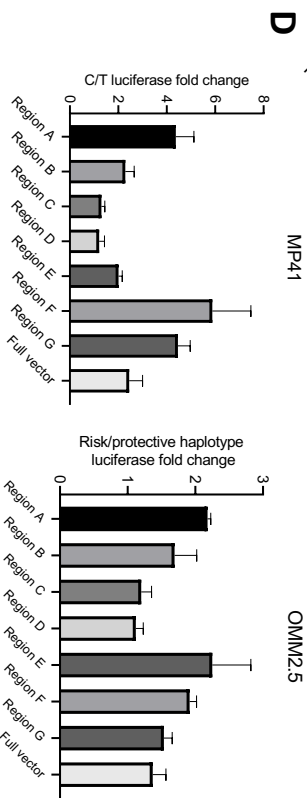
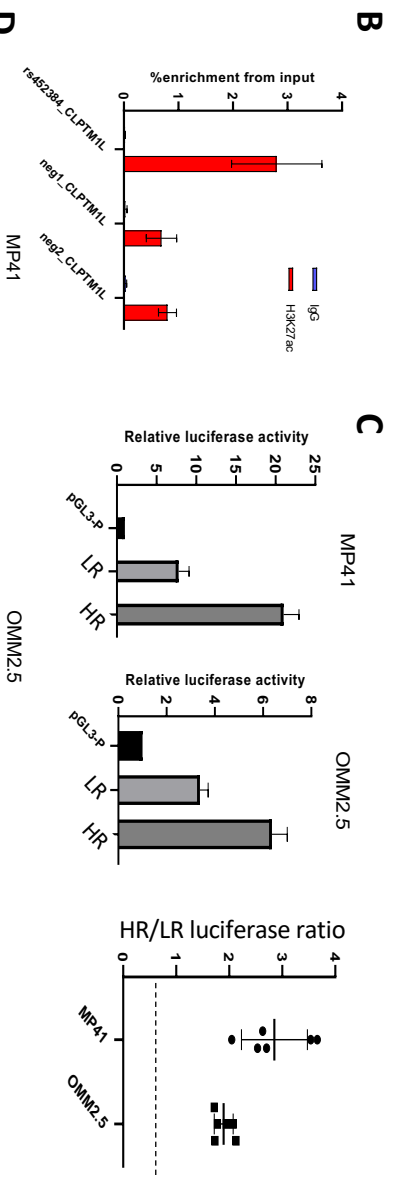
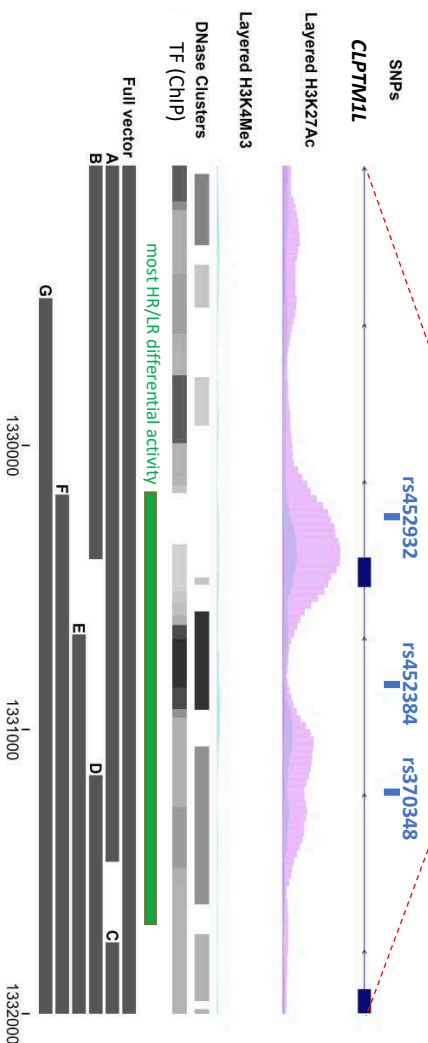
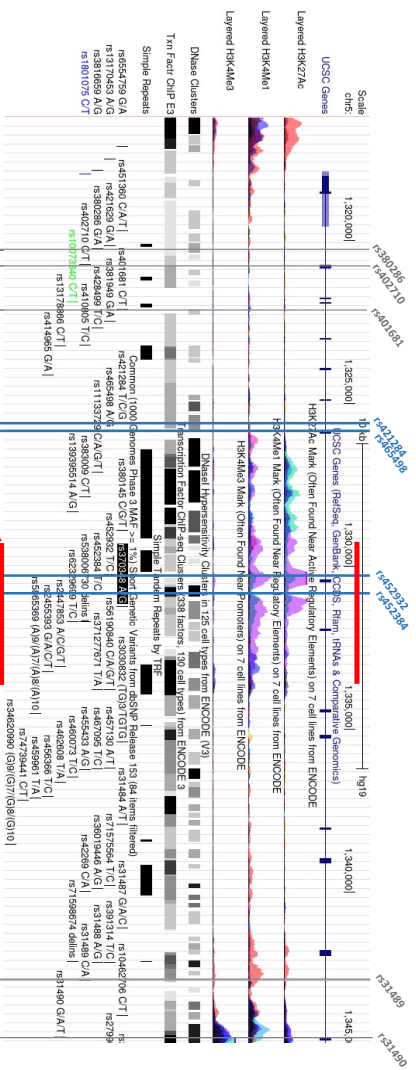


Figure 1. Allele-specific regulatory activity conferred by rs452384 at the *TERT/CLPTM1L* chr5p15.33 uveal melanoma risk locus.

(A) A sub-region at the *TERT/CLPTM1L* risk locus is enriched in marks of open chromatin and behaves as an enhancer. Top panel: Map of the *TERT/CLPTM1L* genomic region along with highly correlated variants associated with risk of uveal melanoma (UM) identified in our initial GWAS study, aligned on Hg19 genome and downloaded from UCSC genome browser. SNPs with odds ratio >1.8 appear in blue while the others appear in grey. Layered ENCODE chromatin marks (H3K27ac, H3K4me1, H3K4me3, DNase I hypersensitivity sequencing, transcription factor ChIP-seq and DNA repeated regions) are shown and were generated from ENCODE using sequencing data from all available cell lines (K562 in purple, HUVEC in light blue, GM12878 in red). Bottom panel: Fragment deletion analysis strategy to narrow down the sub-region from the full haploblock tested («full vector»; genomic region represented by a red rectangle) mediating the most differential luciferase activity between the high-risk and low-risk UM haplotypes. Full vector insert was trimmed from its 5'end or 3'end to generate smaller fragments A to G represented by grey bars aligned on Hg19 genome, subsequently used for luciferase assays in Figure 1D. UM risk variants within the region tested are shown in blue.

(B) Chromatin immunoprecipitation (ChIP)-qPCR using an antibody against H3K27ac histone mark in MP41 UM cell line, showing an enrichment (% from input DNA) of the chromatin mark at around rs452384 (rs452384_CLPTM1L) compared to two other amplicons with the *CLPTM1L* gene near exons 3 and 4 (neg1 and neg2), compared to background IgG levels. Mean enrichment $\pm$ standard deviation of 3 independent immunoprecipitations and PCRs are shown.

(C) Allele-specific luciferase activity of the high-risk (HR) and low-risk (LR) haploblocks of the *CLPTM1L* locus enriched in active histone marks (genomic region represented by the red fragment in Figure 1A) in two UM cell lines, MP41 and OMM2.5. The enhancer region containing either protective or risk variants associated with UM were cloned upstream of a pGL3-Promoter reporter vector, followed by transfection and luciferase assay. Luciferase activity was normalized to renilla luciferase levels and to empty vector. Mean relative luciferase activity $\pm$ standard error of the mean are shown; experiments were performed in triplicates and repeated at least 5 times. Mean HR/LR luciferase ratio $\pm$ standard deviation are also shown, where the hypothetical ratio of HR/LR=1 representing an equal luciferase activity between the two haplotypes is indicated by a black dashed line.

(D) Fragment deletion analysis by luciferase activity assay. Individual fragments A to G (genomic map on Figure 1A bottom panel) of both haplotypes (UM high-risk, HR, and low-risk, LR) were subcloned in pGL3-P luciferase reporter vectors and tested for luciferase activity, normalized to renilla luciferase levels, in UM cell lines MP41 and OMM2.5. Graphs represent the differential luciferase ratio observed with the HR vs. LR haplotypes in each fragment, compared to vector containing the full insert initially tested («full vector»). Most differential activity conferred by this region is represented by a green bar in Figure 1A bottom panel.

(E) Site-directed mutagenesis of rs452932, rs452384 and rs370348, the 3 variants associated with UM risk within the tested region, from their protective to risk allele in MP41. Resulting vectors were tested in luciferase activity assays as previously described and compared to vectors containing the full low-risk (LR) or high-risk (HR) haplotypes (i.e., harboring all 3 variants with either their protective allele or risk allele).

(F) Double site-directed mutagenesis at rs452932 and rs452384 positions (both SNPs T>C) in MP41, compared to « single » mutants described in Figure 1E by luciferase activity assays. The same process was subsequently repeated starting from the HR haplotype vector and mutating tested variants to their protective allele (C>T). For D, E and F, graphs represent mean $\pm$ standard error of the mean (SEM). Experiments were conducted in triplicates and carried out 3 independent times.

regulatory elements lie within the overlapping segment of constructs A and F (represented in green on Figure 1A), which also coincides with the highest peak of H3K27ac mark on ENCODE data in this region.

Among the multiple common variants (with a minor allele frequency [MAF] of >1% in dbSNP153 database) present within the *CLPTM1L* region (red squares, Figure 1A bottom panel), 3 SNPs are associated with UM risk (blue squares, Figure 1A): two previously identified in our first UM GWAS, rs452932T>C and rs452384T>C; and rs370348A>G, which was not genotyped on the Omni5 SNP array (used in the initial UM GWAS) but belongs to the same haploblock and was subsequently identified as lead SNP in our second GWAS recently published,¹³ in high LD with rs421284 ($r^2=0.98$). The genomic positions of these 3 SNPs were all within the identified sub-region conferring most allele-specific activity (green fragment, Figure 1A). We therefore tested the individual effect of these 3 SNPs on allele-specific luciferase activity by performing site-directed mutagenesis of the full length low-risk haplotype vector (LR vector) at the position of each of the 3 SNPs to its HR allele (rs452932-C, rs452384-C and rs370348-G). While luciferase activity levels in the rs370348A>G mutant vector did not change from LR vector, rs452932T>C change resulted in an increase in luciferase activity towards that of HR vector, and an even stronger one with rs452384T>C (Figure 1E; Supplementary Figure 1). To determine whether rs452932 and rs452384 acted in synergy or whether rs452384 was the single main SNP driving allele-specific luciferase expression in this region, a double mutagenesis was performed from LR vector (T>C double mutant). As shown, luciferase signals resulting from rs452384T>C single mutant and from T>C double mutant remained comparable while that of rs452932C>T was lower (mean luciferase activity: rs452932T>C vector: 2.16, range[1.33 – 2.23]; rs452384 T>C vector: 5.03, range [2.72 – 8.69]; both SNPs T>C: 5.99 range [2.25 – 10.76]; full LR vector: 1.85, range [1.22 – 3.19]; and full HR vector: 4.95, range[2.66 – 8.08]) (Figure 1F). To further validate these results, the same mutagenesis process was repeated starting from HR vector towards LR alleles of both SNPs. Consistent with the previous finding, swapping rs452384 HR (C) allele to LR (T) resulted in a sharp decrease in luciferase activity, comparable to the C>T double mutant, while rs452932C>T still demonstrated strong expression levels (Figure 1F). Taken together, these results indicate that rs452384 is the SNP in this UM risk region displaying the highest allele-specific regulatory effects, and that it may have a functional role in UM risk.

As rs452384 genotype influences gene expression in UM, we next aimed to determine whether this SNP and its proximal DNA sequence could influence the binding of nuclear proteins, by performing electrophoretic mobility shift assays (EMSA) in MP41 and OMM2.5 UM cell lines. In both cell lines, rs452384 exhibited allele-enriched binding of at least one

nuclear factor preferentially with the T protective allele (top arrow, Figure 2A), while one strong band and a lighter one with the rs452384-C probe either absent or lighter with rs452384-T probe indicated the preferential binding of one or two nuclear factors with risk allele C (middle and bottom arrows, Figure 2A). A consistent pattern of allele-specific binding was observed in 3 additional UM cell lines, OMM1, MM66 and Mel202 (Supplementary Figure 2A). Addition of C-allele-specific unlabelled competitor probe (used at 100X molar excess) resulted in a sharp decrease in band intensity of complexes preferentially bound to rs452384-C (Figure 2A), more so than upon addition of the same excess of T-allele-specific unlabelled competitor probe (Figure 2B), especially in MP41. The reverse was observed for the complex preferentially bound to rs452384-T, better inhibited by addition of T-allele-specific than C-allele-specific unlabelled probe (Figure 2A&B), more visible in OMM2.5 cell line due to a more intense banding of the T-enriched complex. Adding lower amounts of both unlabelled probes (10X and 100X excess) in MP41 and OMM2.5 cell lines confirmed the selective binding of a complex to rs452384-C and another to rs452384-T (Supplementary Figure 2B). These results suggest that allele-specific regulatory activity driven by rs452384 may act through different nuclear complexes that preferentially bind either protective T or risk C alleles, prompting us to further investigate the transcription factors potentially mediating these effects.

NKX2.4 transcription factor is a rs452384-T allele specific interactor and GATA4 preferentially binds rs452384-C

Identification of proteins bound to rs452384 by quantitative mass spectrometry

To test whether rs452384 differential regulatory activity could be linked to the allele-specific binding of transcription factors through the creation or disruption of potential binding motifs, we performed DNA-pulldown assays using biotin-labeled double stranded oligonucleotides (Supplementary Table 2) centered around rs452384-T (“probe T”) and -C (“probe C”) alleles, and using a nonrelevant biotinylated probe not known to contain any enhancer element⁵³ as a negative control (“NEG”), followed by quantitative mass spectrometry identification of proteins bound to these probes. Protein inclusion criteria for minimum peptide number, fold change and p-value are described in Methods. As shown on the volcano plot in Figure 3A, while most of the proteins identified equally bind C and T probes, multiple proteins (including 9 transcription factors) were found to be preferentially enriched with the T or C alleles, with NKX2.4 and GATA4 transcription factors (TFs) with substantially higher fold change (T-enriched and C-enriched respectively) and most significant p-values compared to other candidate TFs (NKX2.4: enrichment ratio_{T/C}= 13.05; p=3.1x10⁻³¹; GATA4: ratio_{C/T}= 2.98; p=2.0x10⁻²³) (Table 1; supplementary Table 3). Filtering out interactor proteins that were not significantly enriched against negative control probe (Supplementary Figure 3A), 6 distinct

Figure 2

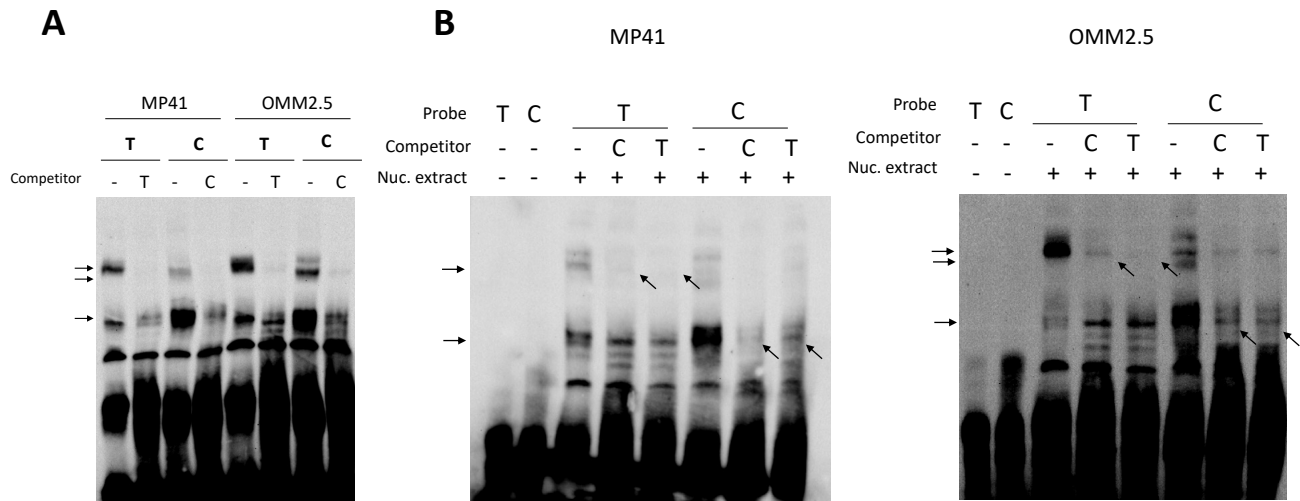


Figure 2. rs452384 preferentially binds nuclear proteins. Electrophoretic mobility shift assays (EMSA) using double-stranded biotinylated DNA oligonucleotides containing either rs452384-T (UM protective allele) or rs452384-C (UM risk allele) in MP41 and OMM2.5 UM cell lines. **(A)** Specific complexes preferentially bind either to the protective allele (T probe) or the risk allele (C probe), represented by black arrows. Upon addition of specific unlabeled competitor probes of the same allele in 100X molar excess, these complexes disappear or are strongly attenuated. **(B)** Nuclear complex preferentially bound to C probe (bottom horizontal arrow) is more strongly competed with unlabeled C probe compared to unlabeled T probe added in 100X excess (bottom angled arrows), and vice versa for complex preferentially bound to T probe, confirming complexes' allele specificities. Addition of unlabeled probes in 10X and 100X excess in Supplementary Figure 2B are consistent with these results and allow visualization of gradual decrease in complex formation upon addition of competitor.

TFs remained enriched in either one or both of rs452384 probes (Figure 3B, Supplementary Figure 3B): NKX2.4, HOXC9 and DLX6 (T-enriched), and GATA4, TRPS1 and PITX2 (C-enriched). A correlation plot comparing ratio_{T/NEG} to ratio_{C/NEG} of all proteins further allowed us to determine that while all other candidate TFs above-mentioned were enriched with either the C- or T- alleles (Figure 3C, top right square left and right respectively), only NKX2.4 displayed complete T-allele specificity (Figure 3C, top left square): both its T/C and NEG/C fold change ratio are significantly enriched, indicating that this protein selectively binds to rs452384-T whereas other candidates preferentially bind to one of the alleles, making it a top candidate for further analyses.

In order to only select the most relevant TFs to functionally assay from the 6 candidates, their predicted binding sites were searched in TF binding public databases JASPAR^{2020,54} and HOMOCOCO Motif collection (<http://hocomoco11.autosome.ru/>). GATA4, NKX2.4, HOXC9 and PITX2 were retained for further analyses as their predicted binding motif matched either perfectly or closely the genomic sequence centered around rs452384 (Supplementary Figure 4; red arrows representing correct expected position of rs452384 SNP on respective binding motifs).

Table 1. Protein candidates identified by quantitative mass spectrometry significantly enriched in rs452384-C or -T alleles. Applied criteria for selection of protein candidates are available in Methods and Figure 3A-C. NEG: negative control probe. Adj p-val: Benjamini-Hochberg adjusted p-value. Fold change ratio represents the relative enrichment of candidate proteins between the two conditions tested. "C/T enriched" proteins refer to proteins enriched in both (C vs NEG) and (T vs NEG) quantifications, that were more enriched in one allele over the other; "T-specific" proteins refer to proteins only enriched in (T vs NEG) quantifications but not enriched in (C vs NEG), suggesting clear specificity for T allele.

Protein	rs452384 allele specificity	T/C fold change ratio	Adj. p-val	T/NEG fold change ratio	Adj. p-val
NKX2.4	T-specific	13.052	3.1×10^{-31}	6.397	2.0×10^{-28}
DLX6	T-enriched	2.959	1.3×10^{-8}	8.885	4.7×10^{-19}

Protein	rs452384 allele specificity	C/T fold change ratio	Adj. p-val	C/NEG fold change ratio	Adj. p-val
GATA4	C-enriched	2.981	4.4×10^{-23}	36.316	3.5×10^{-18}
PITX2	C-enriched	1.727	4.4×10^{-5}	4.904	2.9×10^{-9}

Supershift EMSA for NKX2.4 and GATA4

We next sought to determine whether any of these four proteins preferentially bound the T allele (NKX2.4, DLX6) or C allele (GATA4, PITX2) of rs452384 by supershift EMSA. Due to lack of efficacy or absence of antibodies against the native form of these proteins, cells were first transfected with vectors encoding these proteins and harboring a C-terminal FLAG tag, and an anti-FLAG antibody was subsequently used in rs452384 EMSA. While DLX6 and PITX2 flagged proteins did not lead to any supershift (Supplementary Figure 5), addition of

Figure 3

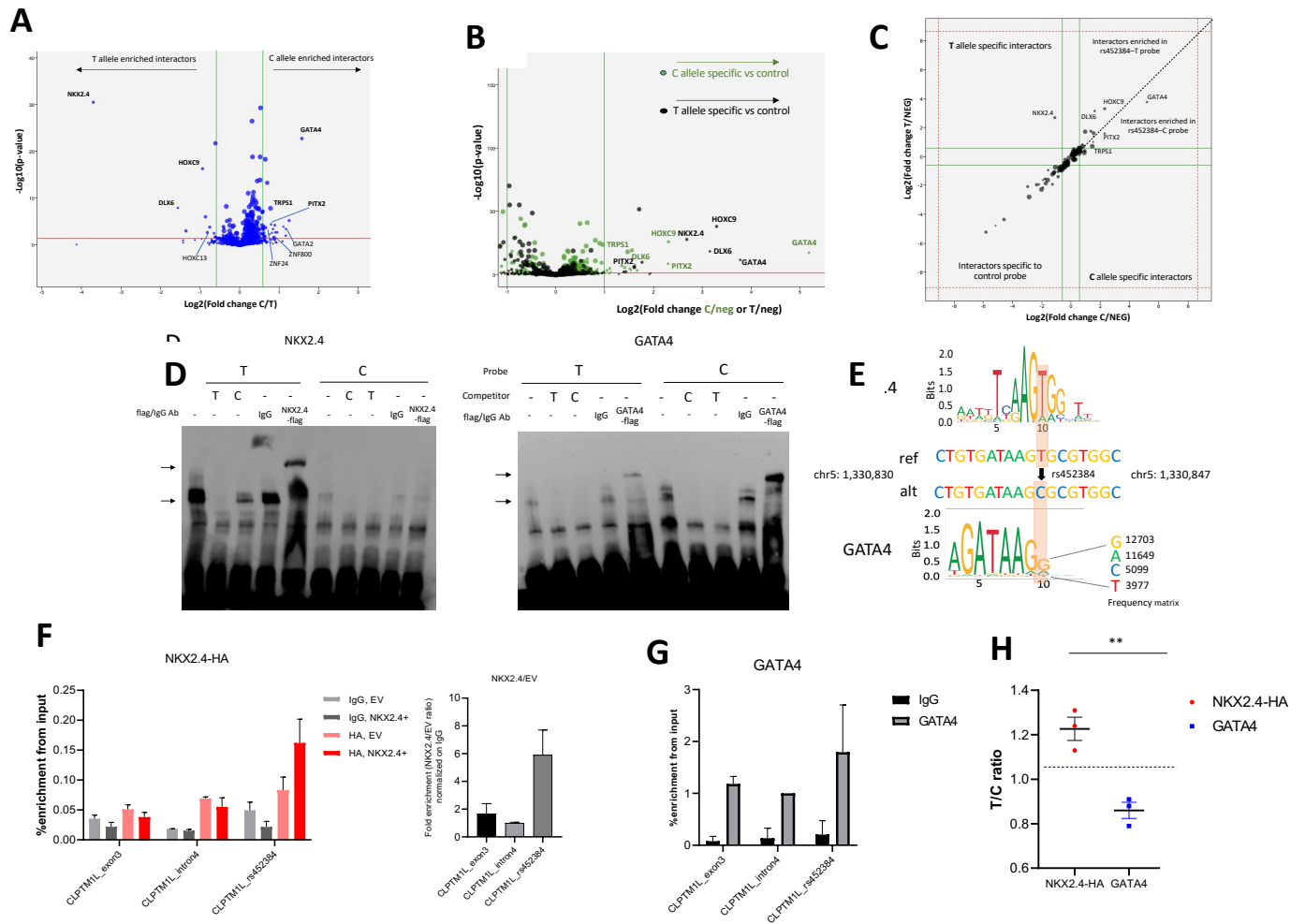


Figure 3. NKX2.4 is an allele-specific interactor of rs452384-T while GATA4 is enriched in rs452384-C.

(A, B, C) Quantitative mass spectrometry analysis after DNA pulldown using either rs452384-T or rs452384-C biotinylated oligonucleotides or a negative control probe, showing enrichment of proteins either with C or T allele compared to control. Each condition was performed in n=5 biological replicates and proteins with at least 3 distinct peptides and a Benjamini-Hochberg adjusted p-value of ≤ 0.05 (see Methods) were kept for analysis.

(A) Volcano plot of all proteins either enriched in T or C rs452384 alleles, denoted by arrows pointing left and right directions respectively, represented as $\log_2(\text{fold change C/T})$ on the x-axis and $-\log_{10}(\text{p-value})$ (y-axis). Horizontal red line represents significant p-value threshold ≤ 0.05 while the vertical green lines indicate the absolute fold change threshold ≥ 1.5 of C/T and T/C enrichments. Transcription factors significantly enriched are shown; only those written in bold were also significantly enriched against the negative control probe.

(B) Volcano plot of proteins enriched in C-allele (green) or T-allele (black) against negative control probe, using same representation and thresholds as in Figure 4A except that fold change threshold against negative probe (green lines) was set to ≥ 2 . Only transcription factors are listed.

(C) Correlation plot comparing enrichment ratios T/NEG (y-axis) to C/NEG (x-axis) represented on a $\log_2$ scale. Green lines indicate an absolute fold change threshold of ≥ 1.5 . Proteins that interact specifically with T or C alleles are found in the top left and bottom right squares respectively, while those preferentially enriched in either probes (but still detected in both) vs NEG are found in the top right corner, where the black dashed diagonal delimitates the T (left) or C (right)-enriched interactors.

(D) Supershift EMSAs using rs452384-T or -C probes, nuclear extracts of MP41 cells transfected with either NKX2.4-flag-tag (left) or GATA4-flag-tag (right) expression vectors, and 2 μ g of anti-flag antibody or negative IgG antibody control. Results show an allele-specific supershift (top arrows) upon NKX2.4 transfection with rs452384 T allele probe, while a supershift enriched in rs452384-C allele probe over T is seen with GATA4 transfection. Both complexes without addition of antibody (bottom arrows) disappear or greatly diminish upon addition of 100X excess of unlabeled competitor C and T probes.

(E) rs452384T>C on chr5p15.33 disrupts a predicted DNA binding motif for NKX2.4 (JASPAR database), while rs452384 C allele favors GATA4 binding compared to rs452384 T (the core binding motif itself remains unaltered but C is preferred over T at the 10th position).

(F,G,H) Chromatin immunoprecipitation (ChIP) experiments using an anti-HA antibody (after transfection of MP41 cells with NKX2.4-HA expression vector) **(F)** or an anti—GATA4 antibody **(G)** compared to background IgG levels. For NKX2.4-HA, % enrichment from input is compared to that of IP with empty vector (EV) and also plotted as an enrichment ratio normalized over EV representing the fold change from EV over three independent IP experiments. Enrichment at rs452384 genomic position was measured by qPCR compared to two other genomic regions within *CLPTM1L* exons 3 and intron 4. Graphs represent mean $\pm$ standard deviation for a representative experiment (3 independent experiments in total). In **(H)**, allele-specific enrichment of immunoprecipitated DNA at rs452384 was measured by digital droplet PCR (ddPCR) relative to input DNA allelic ratio. Graph represents mean T/C allelic ratio $\pm$ standard error of the mean, for three independent experiments; unpaired two-tailed t-test was used to compare T/C ratio between NKX2.4 and GATA4 immunoprecipitations.

exogenous NKX2.4 resulted in the formation of a nuclear complex and a supershift of this complex upon addition of FLAG antibody specifically with rs452384 T allele (Figure 3D). The reverse was observed with GATA4, for which a supershift associated with a decrease in intensity of C-allele specific banding was predominant with C allele over T, although a faint supershift band also appeared in rs452384-T (Figure 3D). While GATA4 C-bound complex was completely lost upon addition of either unlabeled C or T probes, NKX2.4 T-bound complex was better inhibited upon addition of T-allele specific competitor, compared to C. These results are consistent with previous quantitative mass spectrometry data that identified NKX2.4 as a rs452384-T specific interactor, while GATA4 binding was enriched in rs452384-C allele over T. Binding motif analyses of these two proteins further corroborated these findings (Figure 3E): on one hand, rs452384-C completely abolishes the NKX2.4 binding site sequence, in which the T nucleotide is part of the highly conserved core binding sequence TNAAGTG among NKX family members,^{55,54} on the other hand, the core GATA binding motif of GATA4 is not directly affected by rs452384 alleles, but the C base frequency at the 10th position of the binding motif is preferred over the T base (Figure 3E), in agreement with the fact that GATA4 is preferentially bound to rs452384-C allele but still present in both C- and T-allele pulldowns. Taken together, these results suggest that NKX2.4 is the nuclear factor that binds with strongest allele-specificity to rs452384, and GATA4 to a lesser extent, and that one or both these factors may play a role in allele-specific gene regulation at this genomic locus.

ChIP-qPCR and allele-specific ddPCR for NKX2.4 and GATA4

To further test genomic enrichment and allele-specificity of NKX2.4 and GATA4, we performed chromatin immunoprecipitation of NKX2.4-HA tagged protein or GATA4 in MP41 cell line (heterozygous for rs452384) using HA-tag and GATA4 antibodies, followed by qPCR targeting the genomic region of rs452384 and nearby other regions within *CLPTM1L* (Figure 3F&G, Supplementary Table 2). We noted an enrichment of NKX2.4-HA at the genomic position of rs452384 the two other regions on *CLPTM1L* (Figure 3F), corresponding to a 3.5– to 5.9–fold increase in enrichment over IgG normalized to empty vector (mean fold enrichment over IgG $\pm$ SEM: *CLPTM1L*_rs452384 = 5.91 $\pm$ 1.04, *CLPTM1L*_exon3 = 1.65 $\pm$ 0.43, *CLPTM1L*_intron 4= 1.00 $\pm$ 0.03) (Figure 3G). We observed a trend towards enrichment of GATA4 at rs452384 position, yet it was not as clear as for NKX2.4 and the increase in enrichment compared to other regions was relatively weak over three independent experiments (Figure 3G). We also assessed allele-specific enrichment of these two factors on immunoprecipitated DNA by droplet digital PCR (ddPCR) using an rs452384 Taqman genotyping probe, which showed a subtle increase and decrease in rs452384 T/C enrichment ratio (relative to input DNA) for NKX2.4 and GATA-4 respectively over 3

immunoprecipitations, resulting in a significant difference in allele preference for NKX2.4 and GATA4 (mean ratio_{T/C} ±SEM, NKX2.4: 1.23 ±0.05, GATA4: 0.86 ±0.04; two-tailed paired t-test $p_{(NKX2.4vsGATA4)} = 0.026$).

Taken together, these results suggest that NKX2.4 is the strongest interactor of rs452384, where it binds to rs452384-T preferentially. A hypothesis is that NKX2.4 and GATA4 compete for an overlapping binding site with an AAG sequence at its core (Figure 3E), where NKX2.4 T-allele binding displaces the binding site for GATA4.

NKX2.4 knockdown increases *TERT* and *CLPTM1L* mRNA levels

Given the genomic position of the enhancer element encompassing rs452384 within *CLPTM1L* intron and nearby *TERT* promoter, we sought to determine by siRNA-mediated post-transcriptional gene silencing whether *NKX2.4* depletion could result in a change of transcriptional levels of *CLPTM1L* and/or *TERT* in Mel202 UM cell line, homozygous TT for rs452384 due to NKX2.4 T-allele specific binding. Due to the very low expression levels of *TERT* in UM patients and in most UM cell lines, we assessed expression levels of *TERT* and *CLPTM1L* by droplet digital PCR (ddPCR, more sensitive than qPCR especially for low abundance targets) using Taqman gene expression assays of *CLPTM1L*, *TERT* or housekeeping control *GUSB*. siRNA-mediated knockdown of *NKX2.4* resulted in a small yet consistent increase in relative *CLPTM1L* expression, and a more pronounced increase in *TERT* mRNA levels over n=8 independent experiments compared to negative control siRNA ("siCTRL") (*CLPTM1L* relative expression, mean ±SD: siCTRL= 1.06 ±0.082, siNKX2.4= 1.27 ±0.167; *TERT* relative expression: siCTRL= 1.01 ±0.149, siNKX2.4= 1.72 ±0.23) (Figure 4), resulting in a significant fold change of 1.25 for *CLPTM1L* and 1.8 for *TERT* (*CLPTM1L* fold change_{siNKX2.4/siCTRL} = 1.25, unpaired t-test $p=0.006$; *TERT* fold change_{siNKX2.4/siCTRL} = 1.8, unpaired t-test $p=8.2 \times 10^{-6}$). *GATA4* knockdown, however, did not alter *CLPTM1L* nor *TERT* expression significantly, although there was a trend in the opposite direction from *NKX2.4* towards a decrease in *CLPTM1L* expression (*CLPTM1L* relative expression fold change_{siGATA4/siCTRL} = 0.8, unpaired t-test $p=0.237$; *TERT* relative expression fold change_{siNKX2.4/siCTRL} = 1.2, unpaired t-test $p=0.462$). Knockdown efficiencies are shown in Supplementary Figure 6A. These results suggest that NKX2.4 TF may act as a repressor of *CLPTM1L* and *TERT*, preferentially with rs452384-T compared to C allele. One could further hypothesize that GATA4 binding prevents binding of NKX2.4 and its associated repressor activity. Consistent with this hypothesis, there is a significant fold change in both *CLPTM1L* and *TERT* expression between cells transfected with siNKX2.4 and siGATA4 (Figure 4, Supplementary Figure 6B), whereas simultaneous dual knockdown of *NKX2.4* and *GATA4* in Mel202 attenuated the effects of individual knockdowns, particularly on *CLPTM1L* for which expression levels became comparable to cells treated with siCTRL (fold

Figure 4

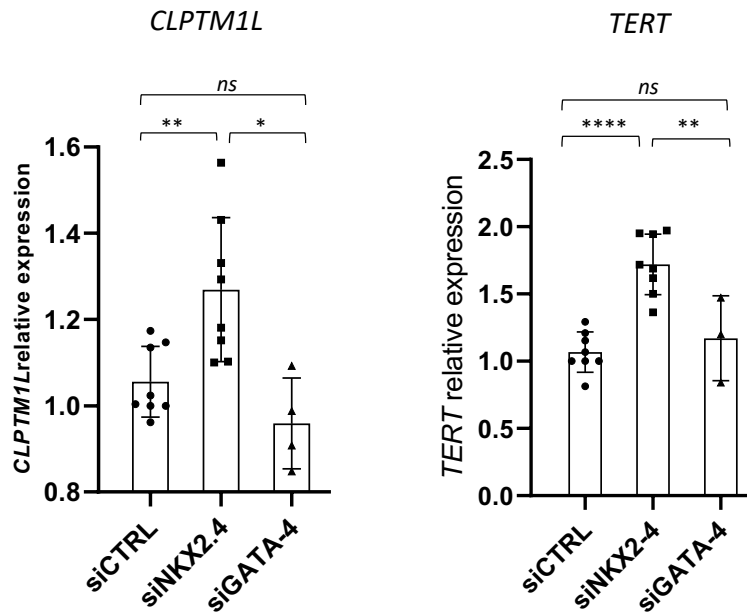


Figure 4 : knockdown of *NKX2.4* decreases *TERT* and *CLPTM1L* expression levels in UM rs452384-TT cell line. Mel202 UM cell line homozygous for rs452384 protective T allele were transfected with siRNAs targeting either *NKX2.4* or *GATA4*, or with a negative siRNA control (siCTRL). Knockdown efficiencies of both genes are shown in Supplementary Figure 6A. Resulting expression of *CLPTM1L* (left) and *TERT* (right) expression relative to that of *GUSB* housekeeping gene was measured by droplet digital PCR using Taqman gene expression assays, and compared to expression in the siCTRL condition. Experiments were performed at least n=7 times for siNKX2.4 and n=3 times for siGATA4; graphs represent mean expression ± standard deviation. T-tests were used to compare expression levels between conditions.

change_{(siGATA4+siNKX2.4)/siCTRL}=1.15, $p=0.184$) (Supplementary Figure 6B). Of note, neither overexpression of NKX2.4 nor that of GATA4 proteins led to changes in *CLPTM1L* or *TERT* expression compared to empty vector, in Mel202 (TT) and OMM1 (CC) UM cell lines (Supplementary Figure 6B and 6C); knockdown experiments in OMM1 cell line could not be interpreted due to poor knockdown efficiencies (Supplementary Figure 6A).

5p15.33 UM risk locus led by rs452384 genotype influences telomere length in peripheral blood lymphocytes of UM patients and European controls

We previously reported an expression quantitative trait loci (eQTL) analysis showing that rs421284 genotype at chr5p15.33 ($r^2>0.9$ with rs452384) is associated with *CLPTM1L* in two in-house series of 73 and 55 UM patients,¹⁵ with a positive correlation with the risk (C) allele, in agreement with our hypothesis that NKX2.4 preferentially represses *CLPTM1L* and *TERT* with the protective T allele of rs452384. *TERT* expression in UM tumor samples being barely detectable even by sensitive PCR techniques, making eQTL analyses challenging, we sought to determine whether differential levels of *TERT* expression, as seen in Mel202 cell line upon *NKX2.4* knockdown, could influence telomere length. We thus measured relative telomere length by qPCR (described in Methods) in (i) germline DNA samples obtained from peripheral blood lymphocytes (PBL) of 480 UM patients and 120 healthy controls (KIDRISK study) from our recently published GWAS in UM, and (ii) in another cohort of 72 UM tumor samples with that were genotyped for rs452384 (cohorts described in Methods, patient characteristics including age, sex and genotype described in Supplementary Table 1). In UM samples (UM PBL germline DNA vs. tumor DNA), telomeres were found to be considerably shorter in tumor samples (median[Q1-Q3] =4.64 [1.35 – 9.10]) than in PBL samples (median= 20.85 [10.10 – 31.47]), accounting for a significant 4.5-decrease in UM tumors compared to PBLs of UM patients (non-parametric Mann-Whitney test $p=3.4\times 10^{-14}$) (Figure 5A). To assess the potential correlation between rs452384 genotype and telomere length (tQTL) in the germline cohorts (PBL samples of UM patients and controls) and in the UM tumor cohort, we next split patients into 3 groups according to their rs452384 genotype. Since the 120 KIDRISK controls were all male, we restricted the analysis in Figure 5B&C to male individuals ($n=326$ PBLs and $n=24$ tumors) and used an age-adjusted generalized linear model (GLM) to compare telomere length without age and sex confounding factors. In the PBL cohort, we observed a significant increase in telomere length in the CC (risk) genotype group (median relative length =27.6), compared to CT (length =21.8) and TT (length =20.3) (GLM p -value= 9.78×10^{-3} , Figure 5B), consistent with our hypothesis that NKX2.4 preferential binding to rs452384 TT genotype results in lower *TERT* expression compared to CC genotype. No difference was found in the UM tumor cohort (Figure 5C, GLM p -value= 0.13), probably in part due to the small sample size and significantly shorter telomere length in this cohort. Pairwise

comparisons between genotypes in male-only individuals and in all individuals (male and female) within the PBL cohort also resulted in a significantly higher telomere length in CC vs. TT and CC vs. TC comparisons (pairwise Wilcoxon Rank Sum test; male-only analysis: $p_{CCvsTT} = 4.65 \times 10^{-3}$; $p_{CCvsTC} = 8.22 \times 10^{-3}$; all samples, male and female analysis: $p_{CCvsTT} = 0.01$; $p_{CCvsTC} = 0.55$) (Supplementary Figure 7). On the other hand, telomere length was not statistically different between TT and TC genotypes (male-only analysis: $p_{TCvsTT} = 0.62$; male and female analysis: $p_{TCvsTT} = 0.43$, Supplementary Figure 7), indicating that both rs452384 C alleles (UM risk allele) are required to significantly increase telomere length in PBL samples. Taken together, these differences in telomere length according to rs452384 genotype in PBL samples indicate that rs452384 influences telomere length, longer telomeres being associated with CC genotype (UM risk allele), and the fact that no difference is observed in UM tumors suggest that regulatory activity mediated by rs452384 on telomere length may arise prior to tumorigenesis in UM. These results suggest that rs452384 genotype influences telomere length in UM, with the T-allele associated with shorter telomeres than the C-allele. This is consistent with a decrease in *TERT* expression in rs452384-TT cell line, potentially regulated by NKX2.4 T-allele-specific interaction at this UM risk locus. Additional telomere measurement in more relevant tissues, such as non-transformed uveal melanocytes, will be required to confirm a role for rs452384 in telomere length regulation prior to UM tumorigenesis.

Figure 5

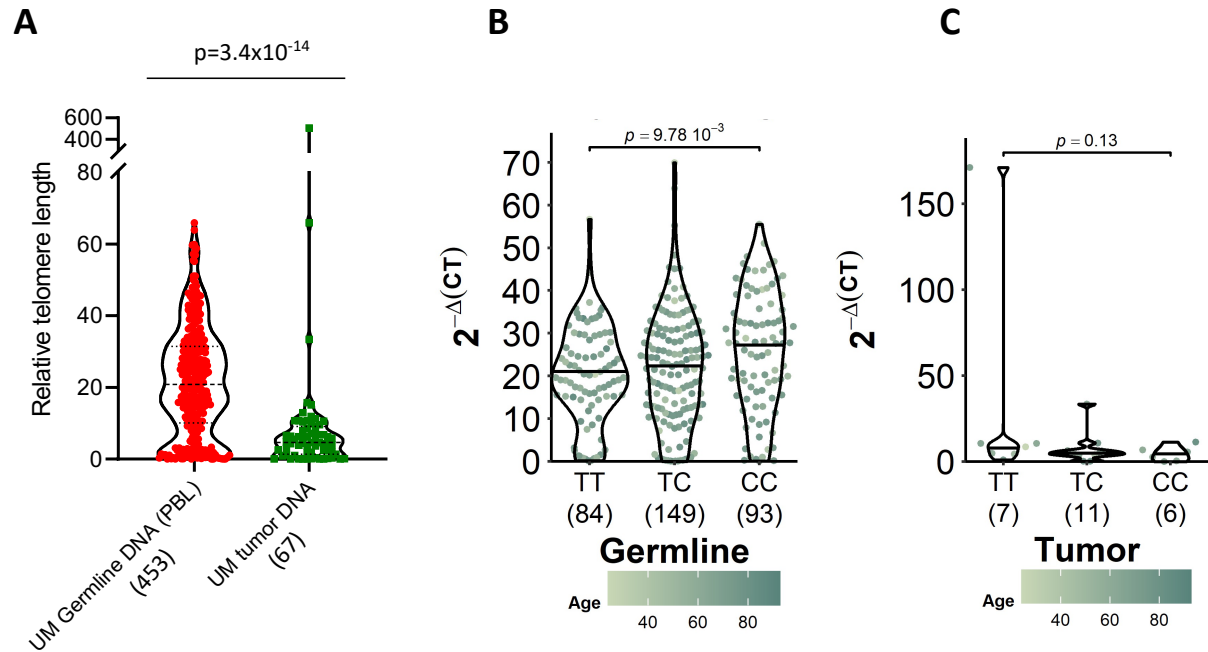


Figure 5: Relative telomere length in peripheral blood lymphocytes (PBL) of UM patients and controls, and in UM tumor samples according to rs452384 genotype (tQTL). DNA samples from PBLs of UM patients and controls from the GWAS cohort and from UM tumor DNA samples were used. Relative telomere length was measured by qPCR using primers for telomeric repeats and normalized to expression of single copy gene *HMBS*. **(A)** Relative telomere length between the UM PBL (germline, $n=453$) and tumor ($n=67$) DNA groups (comparable median age and gender proportion in the 2 cohorts) were compared using a non-parametric Mann-Whitney test p-value to assess statistical significance. Dots represent individual samples and show median telomere length, and Q1 and Q3 quartiles. **(B,C)** DNA samples from the PBL cohorts (B, germline DNA from both UM patients and healthy controls from GWAS) or tumor cohort (C) (male only, age-adjusted; B, $n=326$; C, $n=24$) were split according to their genotype at rs452384. Differences in telomere length between the 3 genotype subgroups were assessed using an age-adjusted generalized linear additive model (GLM) p-value. Because only males were present in the control cohort (KIDRISK), comparisons between genotypes was assessed in male samples only to ensure homogeneous population and avoid gender confounding factor in telomere length comparison. Differences in age are represented by green shadings. Pairwise comparisons of telomere length (between 2 genotypes) and for all patients (male and female) are shown in Supplementary Figure 7.

Discussion

In this study, we sought to functionally characterize the *CLPTM1L/TERT* locus identified in our GWAS studies as a predisposing genetic factor in uveal melanoma (UM), and to elucidate biological mechanisms underlying potentially causal variants within this genomic region ultimately favoring tumor development. The *CLPTM1L/TERT* locus on chr5p15.33 is a highly complex multi-cancer risk locus, with at least 10 independent association loci identified^{17,19} that are differentially associated with cancer depending on tumor type, particularly challenging to functionally characterize as some alleles have been associated with multiple tumor types with different directionality of effect,^{19,35} suggesting some tissue- and cancer-specific regulatory mechanisms.

In UM, a single *CLPTM1L/TERT* association locus has been identified, marked by rs370348 and rs421284.^{13,15} No other risk locus was identified in UM when conditioning on rs421284. UM risk alleles are part of a susceptibility locus previously reported as ‘Region 2’ initially marked by rs401681,^{18,19} and more recently ‘Region 10’ genetically linked to UM risk SNPs and shared with cutaneous melanoma, pancreatic and lung cancers (opposite directionality of effect with the latter, same directionality with the others).¹⁷

The current study aimed to functionally characterize the risk region that lies within intronic sequences spanning the entire length of the *CLPTM1L* gene (Figure 1A), with lead risk variants located approximately 30kb upstream of the promoter of *TERT*, which shares the same locus. Here, we focused on a regulatory region within *CLPTM1L* introns 6, 7 and 8, due to the presence of multiple active chromatin marks in the ENCODE database and subsequently found in UM. The region tested harbored multiple SNPs, including three of the most significantly associated with UM risk (rs421284, rs452932 and rs370348), and this work identified rs452384 as a functional SNP that explains some of the UM risk conferred by this locus, through biological mechanisms involving allele-specific gene regulation, differential protein binding and regulation of *TERT* and to some extent *CLPTM1L* expressions. We further show that genotype of this risk locus, and rs452384 more particularly, influences telomere length in peripheral blood lymphocytes (PBLs) of UM patients and healthy controls while telomeres in UM tumors are considerably shorter and are not impacted by rs452384 genotype. Of note, two other risk loci that are not part of the current study have also been identified as UM susceptibility regions, on chromosome 6 (*IRF4* locus marked by rs12203592) and on chromosome 15 (*HERC2/OCA2* locus, led by rs12913832) involving pigmentation traits.^{12,13} Within the *CLPTM1L/TERT* locus, although no other risk signal independent of the one marked by rs421284 have been identified, it should be noted that a ~4kb region, also harboring some H3K27ac marks and DNase I clusters (ENCODE database) lies downstream of the region tested in the present study, and includes two other UM risk SNPs with high Odds Ratio

(OR), rs421284 and rs465498, located just outside this regulatory mark (Figure 1A). Due to multiple DNA repeat elements at this genomic position (chr5:1,326,000 - 1,330,000 on hg19 genome build, Figure 1A), this tandem-repetitive minisatellite region proved very challenging to sequence and functionally assay in luciferase assays. We are thus currently investigating a potential implication of this genomic region in another study by long-range PCR, which could unravel as-of-yet unidentified risk variants, either linked to the current UM risk peak marked by rs421284 or as an independent risk locus. Our second (and larger) GWAS with imputation also identified additional risk SNPs linked to the rs421284 peak,¹³ including rs2447853 also implicated in cutaneous melanoma susceptibility and nevus count.³² While we cannot rule out that SNPs other than rs452384 also participate in mediating UM risk at this locus, we explored a functionally relevant region and provided biological grounds suggesting that rs452384 is a functional variant within the *TERT/CLPTM1L* locus, mediating UM risk through allele-specific gene regulation.

We found that at least two nuclear complexes preferentially bound rs452384-T and -C alleles respectively in EMSA experiments, subsequently confirmed by quantitative mass spectrometry, which enabled us to validate NKX2.4 transcription factor as a T-allele specific interactor (UM protective allele) and GATA4 as an interactor of both alleles, preferentially enriched with the risk C-allele. Allele-specific chromatin immunoprecipitation (ChIP) results confirmed NKX2.4 binding over rs452384-T and subtle preferential binding of NKX2.4 and GATA4 to T and C alleles respectively. SiRNA-mediated knockdown of *NKX2.4* but not *GATA4* led to a subtle but significant increase in *TERT* and *CLPTM1L*, which was attenuated by dual knockdown of *NKX2.4* and *GATA4* suggesting that these two factors compete for a binding site around rs452384 and that *NKX2.4* may act as a repressor at this genomic position in UM. NKX2.4 subtle preferential binding to rs452384-T over -C in ChIP experiments, along with weak increase in *CLPTM1L* and *TERT* expression (1.2-fold and 1.8-fold respectively) upon *NKX2.4* knockdown, is in accordance with the expected low penetrance conferred by such predisposing loci – characterized by low impact variants – over a person's lifetime.

NKX2.4 belongs to the NKX homeobox family of transcription factors, which recognizes the same core consensus DNA sequence 5'-TNAAGTG-3'.⁵⁶ There are at least 7 members of the NK2 family, with highly tissue-dependent expression patterns, where they play a critical role in organ development by regulating gene expression in a lineage-specific manner;⁵⁷⁻⁵⁹ and defects in these proteins lead to various diseases ranging from cancer to developmental defects. Some of the more extensively studied NK2 family members include NKX2.1, mostly driving expression of thyroid and lung-specific genes,⁶⁰ NKX2.2, implicated in the development of central nervous system and oligodendrocyte differentiation,⁵⁸ and NKX2.5 that regulates cardiac myogenesis and is involved in the development of heart and spleen.⁶¹

NKX2.4 is among the least extensively studied NK2 proteins; it has a high homology with NKX2.1 in DNA sequence (the latter mostly expressed in lung tissue), and its RNA expression level in normal tissue is very low except for testis (NCBI database), hypothalamus and pituitary gland (The Human Protein Atlas). However, all NK2 family members share highly conserved features, act through similar mechanisms and are mostly distinguished by their tissue expression, which could shed light on the function of NKX2.4. Their highly conserved homeodomain confers DNA binding specificity but also serves as a protein-protein interaction domain,⁵⁹ which also explains why NK family members can act both as activator and repressor depending on their binding partners. As such, NKX2.2 represses transcription by co-recruiting Groucho3 or 4, but also has an NK-shared transcriptional activation domain.⁵⁸ NKX2.5 in has dual activator and repressor role in cardiac development in zebrafish and mice, depending on developmental stages.⁶² Strikingly, a previous study has reported that NKX2.5 has a C-terminal inhibitory domain that is removed upon GATA4 co-binding, itself involved in heart formation, resulting in a sharp increase in transcriptional activity of genes regulating heart tissue differentiation.⁵⁵ GATA4 is also part of a family that share a highly conserved DNA binding domain, through two zinc fingers that recognize the sequence (A/T)GATA(A/G) and specify cell lineages.⁶³ Although NKX2.5 is not expressed in UM, one could hypothesize that NKX2.4 and GATA4 co-binding to a site near rs452384-T could result in a different gene regulation than GATA4 only binding on rs452384-C. NKX2.2 was recently shown to function as a tumor suppressor in cancer, where epigenetic silencing of NKX2.2 (which suppressed cell proliferation upon overexpression) via DNA hypermethylation was associated with colorectal cancer,⁵⁸ which would be in accordance with the apparent repressor activity of NKX2.4 in UM, although further studies are required. The prostate-specific homeoprotein NKX3.1 has also recently been shown to exert extranuclear functions, safeguarding against prostate cancer progression by localizing to the mitochondria and conferring protection against oxidative stress, and low NKX3.1 levels are associated with poor clinical outcome.^{64,65} Finally, a study aiming to identify functional lung adenocarcinoma (LUAD) SNPs among risk variants identified by GWAS studies reported rs452384 to bind to NKX2.1 as evidenced by ChIP-Seq data in a LUAD carrying the TT genotype, through a binding site predicted to be disrupted by the C alternative allele.⁶⁶ This study not only finds that rs452384 lies within an enhancer region, marked by H3K27ac and H3Kme1 and a sharp FAIRE (formaldehyde-assisted identification of regulatory elements) peak, but also corroborates our finding that rs452384 creates a binding site that is biologically recognized by NK2 family members, and validates rs452384 as a functional SNP driving LUAD where *NKX2.1*, which plays a key role in lung epithelial tissue differentiation, is significantly amplified.⁶⁶ Although *NKX2.4* is only weakly expressed in UM cell lines, expression of this gene in normal uveal melanocytes and/or in cohorts of UM patients has yet to be determined.

It should also be noted that NKX2.4 specifically binds to a DNA motif that is altered by rs452384T>C, while GATA4 binds slightly upstream at the 5p15.33 genomic sequence and is only marginally affected by rs452384 genotype (Figure 4E), which would argue that NKX2.4 is the main transcription factor regulating allele-specific expression conferred by rs452384. While a repressor activity for NKX2.4 binding on rs452384 protective T allele agrees with the observed increase in expression of *CLPTM1L* and *TERT* upon *NKX2.4* knockdown, the increase in luciferase activity observed with the C risk allele remains to be elucidated. Whether GATA4 and NKX2.4 can specifically coactivate a set of genes favouring UM tumorigenesis, in a similar manner to NKX2.5 and GATA4 co-activation of gene expression in heart tissues, has yet to be determined. 3C experiments and derived techniques (4C, 5C) could also shed light on gene targets other than *TERT* and *CLPTM1L* reported in the present study. Interestingly, it should also be noted that, although they do not function as transcription factors, proteins belonging to the high mobility group (HMG) superfamily were found to be significantly enriched with rs452384-C risk allele, especially HMGB2 and to some extent HMGA1 (Supplementary Table 3). These are non-histone chromatin proteins that mainly function as chromatin remodelers, inducing conformational changes and altering the transcription of genes.⁶⁷ They do not have specific DNA binding sites but recognize various motifs such as AT-hooks (for HMGAs) and HMG boxes composed of α -helices (for HMGBs). While these proteins are usually highly expressed during embryogenesis, they are down-regulated in adult tissues and abnormal changes in expression can lead to a variety of diseases including cancer.⁶⁷ Through their chromatin remodeling activity, HMG proteins have recently been linked to many cancer types, where they promote growth and aggressiveness. Accumulating studies have pointed towards a role for HMGBs in cancer progression (migration and metastasis) in tumors including hepatocellular carcinoma, colorectal, lung and breast cancer.⁶⁸⁻⁷⁰ Strikingly, HMGB2 is expressed mainly in thymus and testis,⁶⁸ the same pattern as NKX2.4 in cancer and healthy tissue. HMGAs can regulate the transcription of genes by either enhancing or suppressing transcription factors,⁷¹ and also have a demonstrated oncogenic activity where overexpression of HMGAs are constantly associated with malignant neoplasia such as in cervical cancer.^{71,72} While we focused in the present study on transcription factors recognizing a specific DNA binding motif altered by rs452384 alleles, other factors that have no clear DNA specificity such as these HMG proteins could participate in epigenetic regulation at the 5p region, as they are preferentially bound to rs452384-C UM risk allele, where they could play a role in recruiting co-factors, altering binding of other transcription factors or acting as chromatin remodelers, and thus participate in UM risk.

If other genes could also be implicated in UM tumorigenesis at the 5p risk locus, *TERT* and *CLPTM1L* remain the most plausible genes regulated by rs452384 downstream of NKX2.4 allele-specific binding. We show that knockdown of *NKX2.4* results in a subtle increase in *TERT* and *CLPTM1L* expression in Mel202 (rs452384-TT) cell line, suggesting that higher levels of these two genes are expected with UM risk allele C than with protective allele T of rs452384. This is consistent with (i) our luciferase reporter assays indicating a weaker gene expression induced by rs452384-T than by rs452384-C, (ii) expression quantitative trait loci (eQTL) analyses (GTEx database) showing association of *TERT* or *CLPTM1L* with rs452384 in numerous tissues (including stomach, oesophagus and skin), and (iii) studies showing implications of these two genes, especially *TERT*, in cancer. On the one hand, *CLPTM1L* expression is positively associated with rs421284 risk allele in UM and with rs465498 (in high LD of $r^2 > 0.9$ with rs452384) in cutaneous melanoma and in normal airway epithelium.¹⁵ The function of *CLPTM1L* and its implication in UM tumorigenesis has not yet been studied in UM to our knowledge, but the protein is overexpressed in pancreatic, ovarian and lung tumor cells, the latter in which it is thought to contribute to RAS-dependent transformation and tumorigenesis by interacting with PI3Ka.³⁹ In pancreatic adenocarcinoma, it relocates to the cell surface upon endoplasmic reticulum stress where it accumulates and interacts with GRP78,⁷³ and the protein promotes growth and aneuploidy in pancreatic cancer cells.⁴⁰ *CLPTM1L* has been shown in multiple tissues and tumor types to be upregulated in ER or genotoxic stress, overexpressed in malignant cells and therefore conferring resistance to chemotherapy.^{41,74} More studies (such as *CLPTM1L* targeted inhibition and its effect on cell growth and survival) would need to be carried out to demonstrate a tumorigenic role for *CLPTM1L* in UM, where the eQTL between rs452384 genotypes is subtle. On the other hand, *TERT*, which shares the same locus as *CLPTM1L*, has often been linked to cancer as it plays a major role in the elongation and maintenance of telomeres, which protect chromosome ends, prevent chromosomal fusion and rearrangements that could result in cell senescence. This mechanism can be overcome in cancer cells by upregulation of *TERT*, the catalytic component of the telomerase enzyme, which is the limiting factor for formation of the telomerase complex in cancer cells.^{75,76} In our previous study, we could not observe a correlation between rs421284 genotype and *TERT*, due to barely detectable expression levels in UM tumors by RNA-sequencing,¹⁵ in accordance with UM tumors being an outlier with the shortest telomeres (ratio of matching tumor and normal samples (T/N) among 31 cancer types studied.⁷⁷ However, several lines of evidence support a role of *TERT* in UM and more specifically in rs452384 allele-specific risk: (i) the GTEx database indicates a correlation between rs452384 genotypes and *TERT* expression, with decreased expression for individuals carrying the C allele in oesophagus; (ii) in lung carcinomas and others, SNPs in the same haplotype are

associated with *TERT* but not *CLPTM1L*,^{18,78} and (iii) *TERT* SNPs play a significant role in predisposition to many solid tumors including cutaneous melanoma.^{19,20,38} For all these reasons, the absence of detectable *TERT* expression in established UM does not rule out a potential role of the regulation of this gene by risk haplotype in UM development prior to tumorigenesis, such as in normal uveal melanocytes or during neural crest differentiation and migration. Here, we used droplet digital PCR (ddPCR) as a highly sensitive method to detect *TERT* expression in UM cell lines, with higher precision than qPCR, and could demonstrate a subtle (1.8-fold) yet consistent increase in expression in Mel202 UM cell line, homozygous for the protective allele (T) of rs452384, suggesting that rs452384 risk allele C may drive in an allele-specific way higher expression level of *TERT*. This slight increase in *TERT* expression may be enough to mediate changes in telomere length, as the telomerase enzyme is most usually found in very low abundance and its activity is tightly regulated, making telomere length sensible to even very subtle changes in *TERT* expression or function.⁷⁹ Once again, studying *TERT* differential expression levels in a tissue other than UM cell lines which are already transformed, and other than UM tumors which essentially do not express *TERT*, such as in normal uveal melanocytes, may reveal stronger allele-specific regulatory effects from rs452384 and higher implication of *TERT*.

To assess telomere regulation by rs452384, we measured relative telomere length in 480 germline DNA from UM patient's peripheral blood lymphocytes (PBLs) and 120 from European controls PBLs from our previous UM GWAS study, and in 72 tumor DNA from UM patients treated at Institut Curie, for which rs452384 genotypes were known. The sharp decrease in telomere length in tumor samples compared to PBL (4-fold) is of comparable magnitude to that of Barthel *et al*'s finding in TCGA UM dataset.⁷⁷ We found that in PBLs, UM patients and healthy controls with a rs452384 CC genotype (UM risk allele) had longer telomeres than those with TT genotype (UM protective allele), consistent with the expected differential *TERT* expression in the same direction suggested by lower levels of *TERT* upon knockdown of T-allele specific interactor NKX2.4. This genotype-telomere length correlation (tQTL) in PBLs of patients and controls corroborates our hypothesis that rs452384 alleles are associated with differential telomere length, longer telomeres being seen with the C allele, conferring risk in UM. UM tumors and cancer cell lines may not be the most fitted model to investigate the role of *TERT* and telomere length in (and prior to) UM tumorigenesis, but suggest a role for a differential telomere regulation conferred by rs452384 alleles to explain the cancer risk associated with this UM susceptibility region. Studies of an association between telomere length and rs452384 genotype in a more appropriate tissue such as uveal melanocytes may shed light on a differential regulation prior to tumorigenesis in UM. Whether the risk conferred by the *CLPTM1L/TERT* risk locus is caused by *CLPTM1L*, *TERT*, or a combination of both

has yet to be determined, although differential telomere length in PBLs points towards a predominant role for *TERT*. Furthermore, it is not impossible that *TERT* is involved in UM tumorigenesis through some of its telomere-independent functions such as gene expression regulation, RNA polymerase activity or regulation of cell survival which may also favor tumorigenesis.⁸⁰⁻⁸²

In summary, the aim of this work was to elucidate some of the biological mechanisms underlying the 5p15.33 *CLPTM1L/TERT* susceptibility region in UM, one of the three independent risk loci identified by UM GWAS along with *IRF4* (chr6) and *HERC2/OCA2* (chr15). We identified rs452384 as a functional variant among the polymorphisms part of a risk region marked by active chromatin marks, where this SNP mediates its risk via allele-specific gene regulation and binding to nuclear factors. We identified NKX2.4 as a protective T-allele specific interactor and suggest that risk mediated by rs452384-NKX2.4 involves differential regulation of *CLPTM1L* and *TERT* in UM. Rescue experiments would be required to validate these preliminary results. We also observe an increased telomere length in PBLs of individuals carrying the rs452384-CC genotype compared to TT (respectively risk and protective alleles in UM predisposition), highlighting the need for future studies to assess whether telomere length is the biological mechanism mediating risk at the 5p15.33 locus in UM. In addition, other factors could participate in mediating allele-specific effects conferred by rs452384, including binding of other transcription factors individually or cooperatively, and other genes affected by this differential regulation. In a similar way, additional risk variants outside the tested region may act in synergy or exert other biological functions also required for UM tumorigenesis. We here unraveled some of the complex regulation mechanisms at chr5p15.33 in UM, encouraging future studies to establish a fully causal link between *TERT/CLPTM1L* risk variants and UM oncogenesis.

Data availability

Uveal melanoma GWAS data is available from EGA under accession number EGAS00001002334 (UM cases from dataset 1), EGAS0000100233 (UM cases from datasets 2 and 3), and from dbGAP under study accession numbers phs001271.v1.p1 (European controls from KIDRISK study). Expression data from tumor UM cohort used for eQTL analyses is accessible on EGA (EGAS00001002334). Proteomics data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository.⁴⁹

Acknowledgments

A.-C. D. was supported by the Horizon 2020 program and innovation program under the

Marie Skłodowska-Curie grant agreement No 666003 and the Ligue Nationale Contre le Cancer. L.M. was supported by the Horizon 2020 program UM Cure (No 667787). D.L. was supported by “Région Ile-de-France” and Fondation pour la Recherche Médicale. This study was funded by the Institut National de la Santé et de la Recherche Médicale (INSERM), the Institut Curie and the Ligue Nationale Contre le Cancer (Labellisation).

References

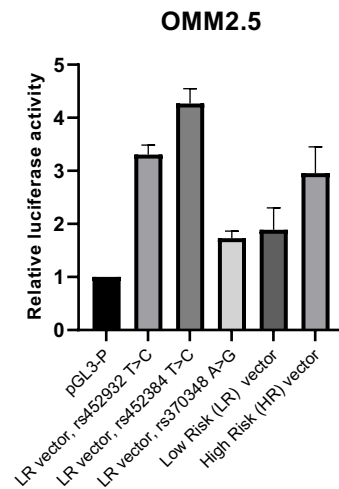
1. Singh, M., Durairaj, P. & Yeung, J. Uveal Melanoma: A Review of the Literature. *Oncol Ther* **6**, 87-104 (2018).
2. Jager, M.J. *et al.* Uveal melanoma. *Nat Rev Dis Primers* **6**, 24 (2020).
3. Van Raamsdonk, C.D. *et al.* Frequent somatic mutations of GNAQ in uveal melanoma and blue naevi. *Nature* **457**, 599-602 (2009).
4. Johansson, P. *et al.* Deep sequencing of uveal melanoma identifies a recurrent mutation in PLCB4. *Oncotarget* **7**, 4624-31 (2016).
5. Harbour, J.W. *et al.* Frequent mutation of BAP1 in metastasizing uveal melanomas. *Science* **330**, 1410-3 (2010).
6. Furney, S.J. *et al.* SF3B1 mutations are associated with alternative splicing in uveal melanoma. *Cancer Discov* **3**, 1122-1129 (2013).
7. Martin, M. *et al.* Exome sequencing identifies recurrent somatic mutations in EIF1AX and SF3B1 in uveal melanoma with disomy 3. *Nat Genet* **45**, 933-6 (2013).
8. Rodrigues, M. *et al.* So Close, yet so Far: Discrepancies between Uveal and Other Melanomas. A Position Paper from UM Cure 2020. *Cancers (Basel)* **11**, 1032 (2019).
9. Aronow, M.E., Topham, A.K. & Singh, A.D. Uveal Melanoma: 5-Year Update on Incidence, Treatment, and Survival (SEER 1973-2013). *Ocul Oncol Pathol* **4**, 145-151 (2018).
10. Bishop, K.D. & Olszewski, A.J. Epidemiology and survival outcomes of ocular and mucosal melanomas: a population-based analysis. *Int J Cancer* **134**, 2961-71 (2014).
11. Weis, E., Shah, C.P., Lajous, M., Shields, J.A. & Shields, C.L. The association between host susceptibility factors and uveal melanoma: a meta-analysis. *Arch Ophthalmol* **124**, 54-60 (2006).
12. Ferguson, R. *et al.* Genetic markers of pigmentation are novel risk loci for uveal melanoma. *Sci Rep* **6**, 31191 (2016).
13. Mobuchon, L. *et al.* Different Pigmentation Risk Loci for High-Risk Monosomy 3 and Low-Risk Disomy 3 Uveal Melanomas. *J Natl Cancer Inst* (2021).
14. Strickland, D. & Lee, J.A. Melanomas of eye: stability of rates. *Am J Epidemiol* **113**, 700-2 (1981).
15. Mobuchon, L. *et al.* A GWAS in uveal melanoma identifies risk polymorphisms in the CLPTM1L locus. *NPJ Genom Med* **2**(2017).
16. Thomsen, H. *et al.* Genome-wide study on uveal melanoma patients finds association to DNA repair gene TDP1. *Melanoma Res* (2019).
17. Chen, H.M., A.; *et al.* Large-scale cross-cancer fine-mapping of the 5p15.33 region reveals multiple independent signals. *Human Genetics and Genomics Advances* **2**(2021).
18. Fang, J. *et al.* Functional characterization of a multi-cancer risk locus on chr5p15.33 reveals regulation of TERT by ZNF148. *Nat Commun* **8**, 15034 (2017).
19. Wang, Z. *et al.* Imputation and subset-based association analysis across different cancer types identifies multiple independent risk loci in the TERT-CLPTM1L region on chromosome 5p15.33. *Hum Mol Genet* **23**, 6616-33 (2014).
20. Rafnar, T. *et al.* Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. *Nat Genet* **41**, 221-7 (2009).
21. Michailidou, K. *et al.* Association analysis identifies 65 new breast cancer risk loci. *Nature* **551**, 92-94 (2017).
22. Huyghe, J.R. *et al.* Discovery of common and rare genetic risk variants for colorectal cancer. *Nat Genet* **51**, 76-87 (2019).
23. McKay, J.D. *et al.* Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. *Nat Genet* **49**, 1126-1132 (2017).

24. Klein, A.P. *et al.* Genome-wide meta-analysis identifies five new susceptibility loci for pancreatic cancer. *Nat Commun* **9**, 556 (2018).
25. Schumacher, F.R. *et al.* Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat Genet* **50**, 928-936 (2018).
26. Scelo, G. *et al.* Genome-wide association study identifies multiple risk loci for renal cell carcinoma. *Nat Commun* **8**, 15724 (2017).
27. Phelan, C.M. *et al.* Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. *Nat Genet* (2017).
28. Lesseur, C. *et al.* Genome-wide association analyses identify new susceptibility loci for oral cavity and pharyngeal cancer. *Nat Genet* **48**, 1544-1550 (2016).
29. Gharahkhani, P. *et al.* Genome-wide association studies in oesophageal adenocarcinoma and Barrett's oesophagus: a large-scale meta-analysis. *Lancet Oncol* **17**, 1363-1373 (2016).
30. O'Mara, T.A. *et al.* Identification of nine new susceptibility loci for endometrial cancer. *Nat Commun* **9**, 3166 (2018).
31. Chung, C.C. *et al.* Meta-analysis identifies four new loci associated with testicular germ cell tumor. *Nat Genet* **45**, 680-5 (2013).
32. Landi, M.T. *et al.* Genome-wide association meta-analyses combining multiple risk phenotypes provide insights into the genetic architecture of cutaneous melanoma susceptibility. *Nat Genet* **52**, 494-504 (2020).
33. Law, M.H. *et al.* Genome-wide meta-analysis identifies five new susceptibility loci for cutaneous malignant melanoma. *Nat Genet* **47**, 987-95 (2015).
34. Melin, B.S. *et al.* Genome-wide association study of glioma subtypes identifies specific differences in genetic susceptibility to glioblastoma and non-glioblastoma tumors. *Nat Genet* **49**, 789-794 (2017).
35. Bojesen, S.E. *et al.* Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. *Nat Genet* **45**, 371-84, 384e1-2 (2013).
36. Yin, J. *et al.* TERT-CLPTM1L polymorphism rs401681 contributes to cancers risk: evidence from a meta-analysis based on 29 publications. *PLoS One* **7**, e50650 (2012).
37. Barrett, J.H. *et al.* Genome-wide association study identifies three new melanoma susceptibility loci. *Nat Genet* **43**, 1108-13 (2011).
38. Huang, F.W. *et al.* Highly recurrent TERT promoter mutations in human melanoma. *Science* **339**, 957-9 (2013).
39. James, M.A., Vikis, H.G., Tate, E., Rymaszewski, A.L. & You, M. CRR9/CLPTM1L regulates cell survival signaling and is required for Ras transformation and lung tumorigenesis. *Cancer Res* **74**, 1116-27 (2014).
40. Jia, J. *et al.* CLPTM1L promotes growth and enhances aneuploidy in pancreatic cancer cells. *Cancer Res* **74**, 2785-95 (2014).
41. James, M.A. *et al.* Functional characterization of CLPTM1L as a lung cancer risk candidate gene in the 5p15.33 locus. *PLoS One* **7**, e36116 (2012).
42. Cheung, A.L. & Deng, W. Telomere dysfunction, genome instability and cancer. *Front Biosci* **13**, 2075-90 (2008).
43. Alsafadi, S. *et al.* Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. *Nat Commun* **7**, 10615 (2016).
44. Amirouchene-Angelozi, N. *et al.* Establishment of novel cell lines recapitulating the genetic landscape of uveal melanoma and preclinical validation of mTOR as a therapeutic target. *Mol Oncol* **8**, 1508-20 (2014).
45. Singh, B. & Nath, S.K. Identification of Proteins Interacting with Single Nucleotide Polymorphisms (SNPs) by DNA Pull-Down Assay. *Methods Mol Biol* **1855**, 355-362 (2019).
46. Poulet, P., Carpentier, S. & Barillot, E. myProMS, a web server for management and validation of mass spectrometry-based proteomic data. *Proteomics* **7**, 2553-6 (2007).
47. The, M., MacCoss, M.J., Noble, W.S. & Kall, L. Fast and Accurate Protein False Discovery Rates on Large-Scale Proteomics Data Sets with Percolator 3.0. *J Am Soc Mass Spectrom* **27**, 1719-1727 (2016).
48. Valot, B., Langella, O., Nano, E. & Zivy, M. MassChroQ: a versatile tool for mass spectrometry quantification. *Proteomics* **11**, 3572-7 (2011).
49. Perez-Riverol, Y. *et al.* The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* **47**, D442-D450 (2019).
50. Cawthon, R.M. Telomere measurement by quantitative PCR. *Nucleic Acids Res* **30**, e47 (2002).

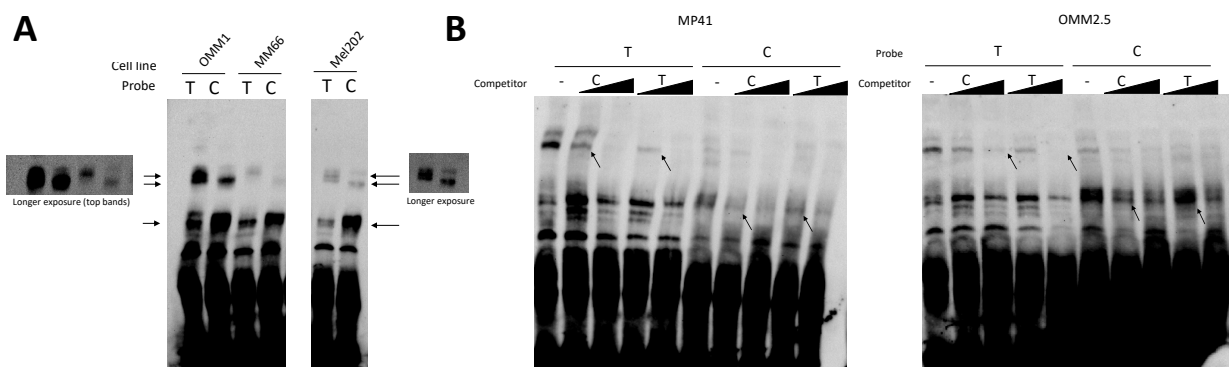
51. Ningarhari, M. *et al.* Telomere length is key to hepatocellular carcinoma diversity and telomerase addiction is an actionable therapeutic target. *J Hepatol* **74**, 1155-1166 (2021).
52. Ramakers, C., Ruijter, J.M., Deprez, R.H. & Moorman, A.F. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci Lett* **339**, 62-6 (2003).
53. Redondo-Munoz, J. *et al.* Matrix metalloproteinase-9 promotes chronic lymphocytic leukemia b cell survival through its hemopexin domain. *Cancer Cell* **17**, 160-72 (2010).
54. Fornes, O. *et al.* JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **48**, D87-D92 (2020).
55. Sepulveda, J.L. *et al.* GATA-4 and Nkx-2.5 coactivate Nkx-2 DNA binding targets: role for regulating early cardiac gene expression. *Mol Cell Biol* **18**, 3405-15 (1998).
56. Watada, H., Mirmira, R.G., Kalamaras, J. & German, M.S. Intramolecular control of transcriptional activity by the NK2-specific domain in NK-2 homeodomain proteins. *Proc Natl Acad Sci U S A* **97**, 9443-8 (2000).
57. Stanfel, M.N., Moses, K.A., Schwartz, R.J. & Zimmer, W.E. Regulation of organ development by the NKX-homeodomain factors: an NKX code. *Cell Mol Biol (Noisy-le-grand)* **Suppl 51**, OL785-99 (2005).
58. He, Y. *et al.* NK homeobox 2.2 functions as tumor suppressor in colorectal cancer due to DNA methylation. *J Cancer* **11**, 4791-4800 (2020).
59. Gorski, B.e.a. Nkx2.5-dependent alterations of the embryonic heart DNA methylome identify novel cis-regulatory elements in cardiac development. *Bio R Xiv* (2017).
60. Boggaram, V. Thyroid transcription factor-1 (TTF-1/Nkx2.1/TITF1) gene regulation in the lung. *Clin Sci (Lond)* **116**, 27-35 (2009).
61. Benson, D.W. *et al.* Mutations in the cardiac transcription factor NKX2.5 affect diverse cardiac developmental pathways. *J Clin Invest* **104**, 1567-73 (1999).
62. Targoff, K.L., Schell, T. & Yelon, D. Nkx genes regulate heart tube extension and exert differential effects on ventricular and atrial cell number. *Dev Biol* **322**, 314-21 (2008).
63. Ko, L.J. & Engel, J.D. DNA-binding specificities of the GATA transcription factor family. *Mol Cell Biol* **13**, 4011-22 (1993).
64. Papachristodoulou, A. *et al.* NKX3.1 Localization to Mitochondria Suppresses Prostate Cancer Initiation. *Cancer Discov* **11**, 2316-2333 (2021).
65. Finch, A.J. & Baena, E. Spatiofunctional Dynamics of NKX3.1 to Safeguard the Prostate from Cancer. *Cancer Discov* **11**, 2132-2134 (2021).
66. Yang, C. *et al.* Positional integration of lung adenocarcinoma susceptibility loci with primary human alveolar epithelial cell epigenomes. *Epigenomics* **10**, 1167-1187 (2018).
67. Hock, R., Furusawa, T., Ueda, T. & Bustin, M. HMGB chromosomal proteins in development and disease. *Trends Cell Biol* **17**, 72-9 (2007).
68. Niu, L. *et al.* Biological functions and theranostic potential of HMGB family members in human cancers. *Ther Adv Med Oncol* **12**, 1758835920970850 (2020).
69. Kwon, J.H. *et al.* Overexpression of high-mobility group box 2 is associated with tumor aggressiveness and prognosis of hepatocellular carcinoma. *Clin Cancer Res* **16**, 5511-21 (2010).
70. Fu, D. *et al.* HMGB2 is associated with malignancy and regulates Warburg effect by targeting LDHB and FBP1 in breast cancer. *Cell Commun Signal* **16**, 8 (2018).
71. Fusco, A. & Fedele, M. Roles of HMGA proteins in cancer. *Nat Rev Cancer* **7**, 899-910 (2007).
72. Fu, F. *et al.* HMGA1 exacerbates tumor growth through regulating the cell cycle and accelerates migration/invasion via targeting miR-221/222 in cervical cancer. *Cell Death Dis* **9**, 594 (2018).
73. Clarke, W.R., Amundadottir, L. & James, M.A. CLPTM1L/CRR9 ectodomain interaction with GRP78 at the cell surface signals for survival and chemoresistance upon ER stress in pancreatic adenocarcinoma cells. *Int J Cancer* **144**, 1367-1378 (2019).
74. Parashar, D. *et al.* Targeted biologic inhibition of both tumor cell-intrinsic and intercellular CLPTM1L/CRR9-mediated chemotherapeutic drug resistance. *NPJ Precis Oncol* **5**, 16 (2021).
75. Dratwa, M., Wysoczanska, B., Lacina, P., Kubik, T. & Bogunia-Kubik, K. TERT-Regulation and Roles in Cancer Formation. *Front Immunol* **11**, 589929 (2020).
76. Daniel, M., Peek, G.W. & Tollefsbol, T.O. Regulation of the human catalytic subunit of telomerase (hTERT). *Gene* **498**, 135-46 (2012).
77. Barthel, F.P. *et al.* Systematic analysis of telomere length and somatic alterations in 31 cancer types. *Nat Genet* **49**, 349-357 (2017).
78. Yang, Y.C. *et al.* rs401681 and rs402710 confer lung cancer susceptibility by regulating TERT expression instead of CLPTM1L in East Asian populations. *Carcinogenesis* **39**, 1216-1221 (2018).

79. McNally, E.J., Luncsford, P.J. & Armanios, M. Long telomeres and cancer risk: the price of cellular immortality. *J Clin Invest* **129**, 3474-3481 (2019).
80. Cao, Y., Li, H., Deb, S. & Liu, J.P. TERT regulates cell survival independent of telomerase enzymatic activity. *Oncogene* **21**, 3130-8 (2002).
81. Chiodi, I. & Mondello, C. Telomere-independent functions of telomerase in nuclei, cytoplasm, and mitochondria. *Front Oncol* **2**, 133 (2012).
82. Martinez, P. & Blasco, M.A. Telomeric and extra-telomeric roles for telomerase and the telomere-binding proteins. *Nat Rev Cancer* **11**, 161-76 (2011).

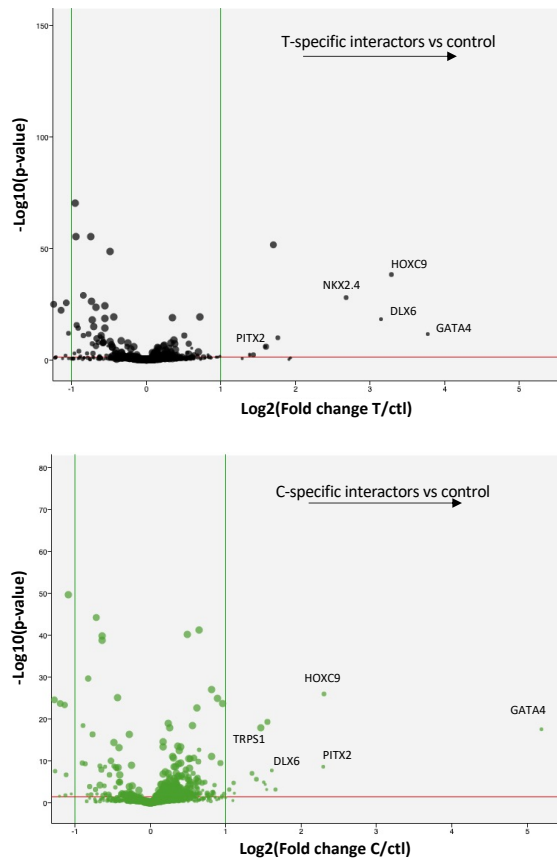
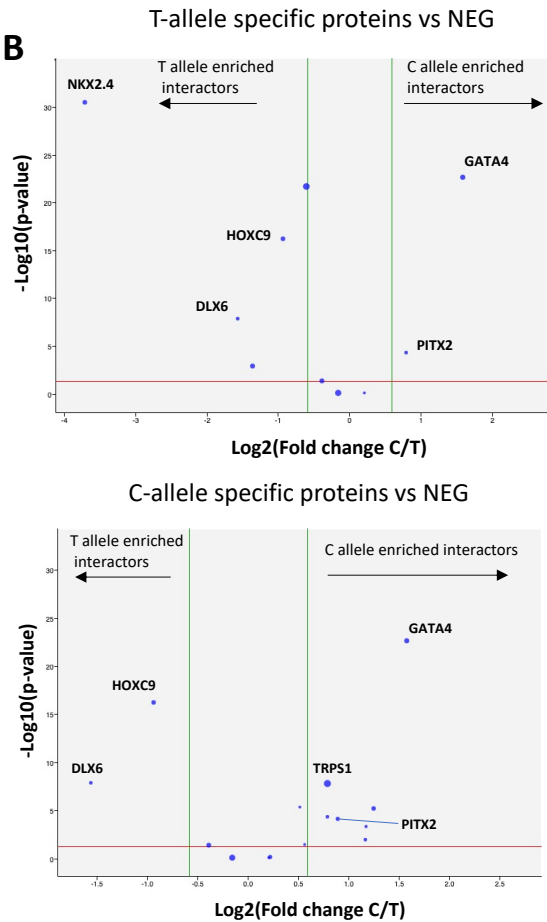
Supplementary Figures 1–7



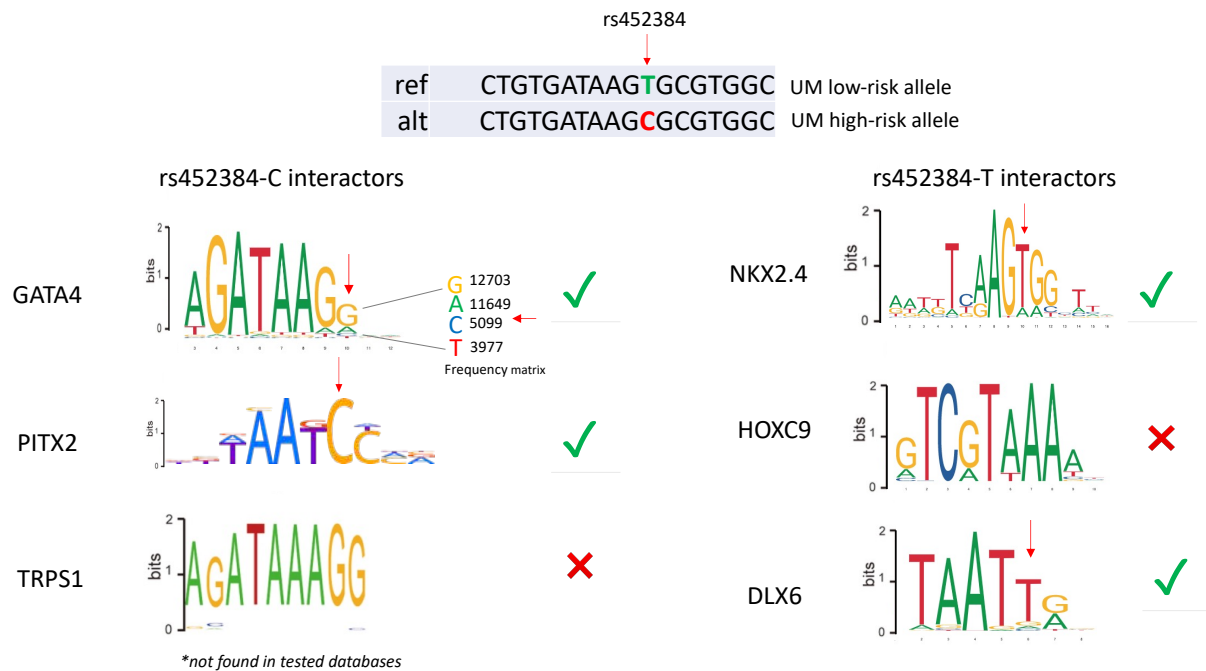
Supplementary Figure 1. Site-directed mutagenesis of the three candidate SNPs and luciferase activity assay in OMM2.5 UM cell line. Site-directed mutagenesis of rs452932, rs452384 and rs370348, the three variants associated with UM risk within the tested region on chr5p15.33, from their protective to risk allele in OMM2.5 cell line. Resulting vectors were tested in luciferase activity assays as previously described and compared to vectors containing the full low-risk (LR) or high-risk (HR) haplotypes (ie, all three variants associated with UM with either their protective allele or risk allele). Graph represents mean $\pm$ standard error of the mean (SEM) from 3 independent experiments.



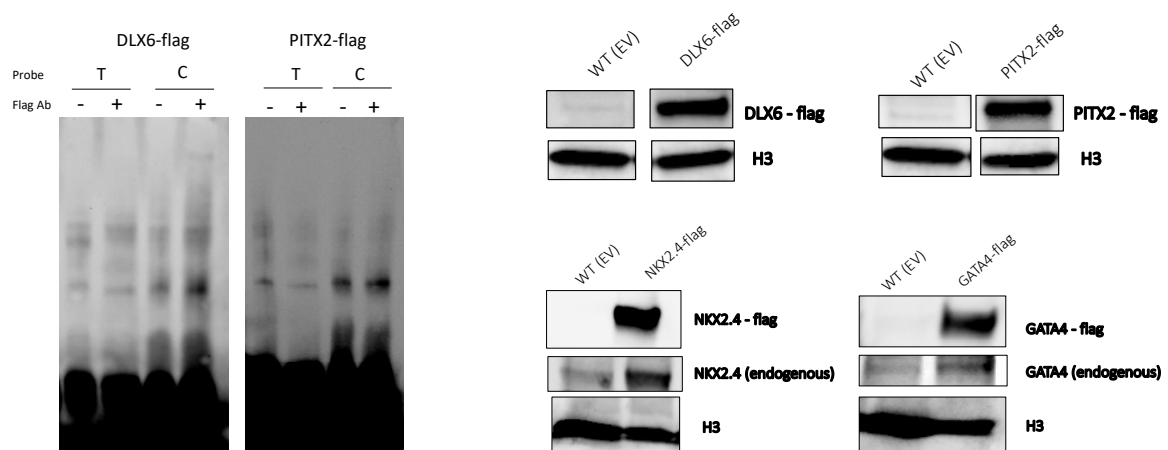
Supplementary Figure 2. rs452384 electrophoretic mobility shift assay in UM cell lines. Electrophoretic mobility shift assays (EMSA) using double-stranded biotinylated DNA oligonucleotides containing either rs452384-T (UM protective allele) or rs452384-C (UM risk allele) in OMM1, MM66, Mel202 (**A**) and MP41 and OMM2.5 (**B**) UM cell lines. In (**A**), images with longer exposure times are also shown to better visualize allele-specific complexes (top arrows) in UM cell lines where lower amounts of nuclear extract was added relative to OMM1. In (**B**), unlabelled specific competitor probes containing either C or T alleles of rs452384 were added at 10X and 100X molar excess (represented by black gradient) to probes-nuclear extract complexes, showing allele-specific decrease of signal (angled arrows).

A**B**

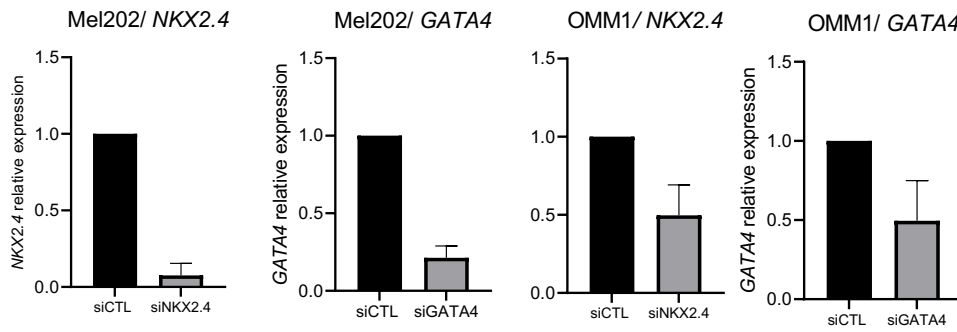
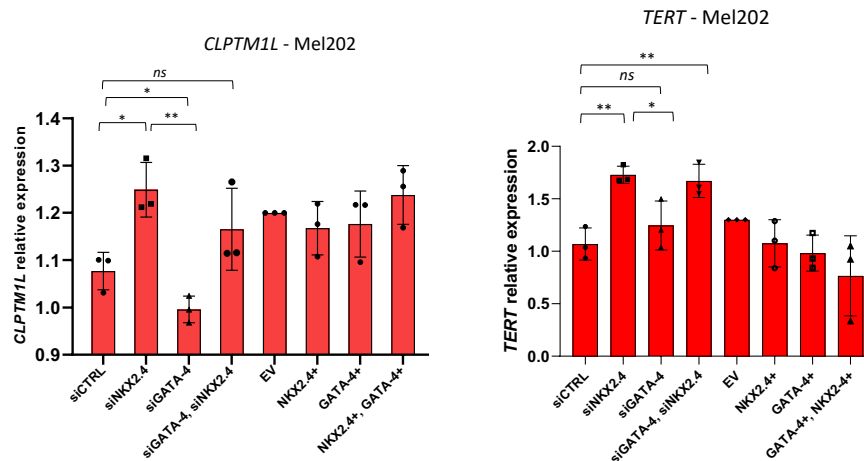
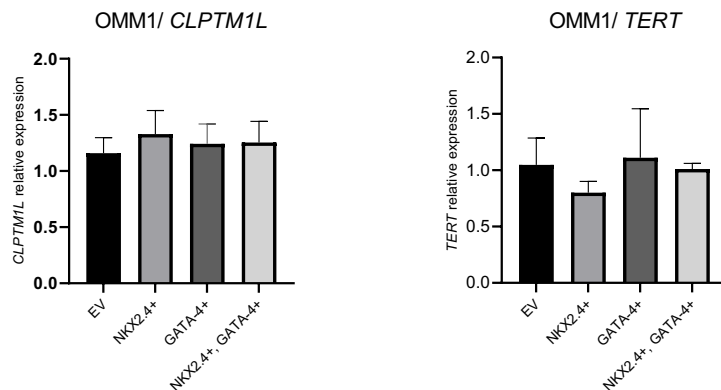
Supplementary Figure 3. Volcano plots of rs452384-C- and -T enriched proteins versus negative control probe (A) and of T-enriched versus C-enriched interactors after filtering out proteins not significantly enriched versus negative control probe (B). Quantitative mass spectrometry was performed after DNA pulldown using either rs452384-T or rs452384-C biotinylated oligonucleotides or a negative control probe, showing enrichment of proteins either with C or T allele compared to control (NEG). Each condition was performed in $n=5$ biological replicates and proteins with at least 3 distinct peptides and a Benjamini-Hochberg adjusted p-value of ≤ 0.05 (see Methods) were kept for analysis. Green lines represent fold change threshold, set at ≥ 2 against NEG probe (A) and an absolute fold change of ≥ 1.5 for C vs T and T vs C comparisons (B). In (A), the top volcano plot represents quantification of T-enriched interactors vs. NEG, while the bottom volcano plot represents C vs NEG quantification. In (B), proteins not significantly enriched vs. NEG have been filtered out, and the remaining proteins enriched in either T vs NEG quantification (top) or C vs neg quantification (bottom) are tested for C vs T enrichment. Only proteins that are transcription factors are listed.



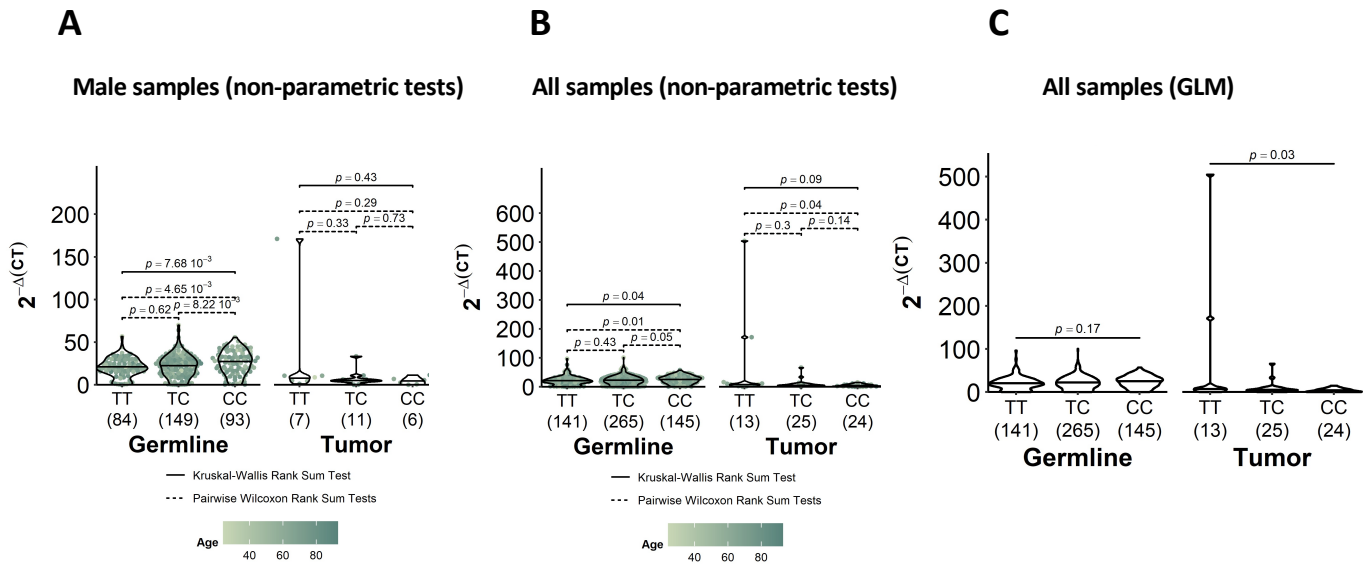
Supplementary Figure 4. Candidate transcription factors binding site analysis using JASPAR and HOMOCOCO binding motif databases, compared to rs452384 and surrounding genomic sequence on chromosome 5p15.33 *TERT/CLPTM1L* locus. The genomic sequence surrounding rs452384 is shown, with rs452384 UM protective allele T in green and risk allele C in red. Only transcription factors for which the binding site matches perfectly or closely the sequence surrounding rs452384, with rs452384 polymorphism either creating or disrupting the binding site, were kept for future analyses (green check marks).



Supplementary Figure 5. (A) Supershift EMSAs for rs452384-C/T using MP41 nuclear extracts transfected with either DLX6-flag or PITX2-flag expression vectors, and using 2ug of anti-flag antibody; **(B)** Over-expression of the four candidate transcription factors NKX2.4, GATA4, PITX2 and DLX6 visualized by Western blots with anti-GATA4 and anti-NKX2.4 (for endogenous protein), anti-flag for exogenous proteins, and anti-histone 3 (H3) as a nuclear protein for loading loading

A**B****C**

Supplementary Figure 6. siRNA-mediated knockdown of *GATA4* and *NKX2.4* in UM cell lines. (A) Knockdown efficiencies for *NKX2.4* and *GATA4* in Mel202 (rs452384-TT) and OMM1 (rs452384-CC) UM cell lines measured by qPCR using primers for *GATA4* and *NKX2.4* amplicons and normalized on GAPDH and TBP (Supplementary Table 2). Expression levels were normalized to that of transfection with a nontargeting siRNA control (siCTL). Graphs represent mean $\pm$ standard deviation. (B,C) Expression levels of *CLPTM1L* and *TERT* in Mel202 TT cell line after knockdown or overexpression of *NKX2.4* or *GATA4* (B) or in OMM1 CC cell line after overexpression (knockdown efficiencies were insufficient to assess expression levels in these conditions- see A). Expression levels were assessed by droplet digital PCR (ddPCR) and normalized to *GUSB* housekeeping gene. Experiments were carried out in 3 independent experiments; graphs represent mean expression level $\pm$ standard deviation. Unpaired t-tests were used to assess statistical significance between expression levels in each condition.



Supplementary Figure 7. Relative telomere length according to rs452384 genotype (tQTL) in peripheral blood lymphocytes (PBL) and in UM tumors. Because only males were present in the KIDRISK control cohort within the PBL group (n=120 males, n=0 female), the tQTLs in the different cohorts (germline and tumor) in Figure 5 were performed in using male samples only. **(A,B)** Pairwise and three-way comparisons of telomere length between the 3 rs452384 genotypes (TT, TC, CC) using non-parametric pairwise Wilcoxon Rank sum test and Kruskal-Wallis Rank Sum test p-values, respectively, in male subjects (A) and in all subjects (male and female, B) of the germline PBL cohort (UM patients and controls) and of the tumor UM cohort. Populations were homogeneous in age distribution (green shadings). Number of samples in each group is shown below the genotype. **(C)** Age- and gender-adjusted generalized linear model (GLM) comparing telomere length according to rs452384 genotype for all samples (male and female) in the germline (PBL from UM patients and controls) and in the UM tumor cohorts.

Supplementary Table 1: Characteristics of individuals included in the telomere length measurement study. After filtering out samples that technically failed during qPCR telomere length measurement, 433 germline DNA from cohort of UM patients, (UM_GL_1-433), 118 germline DNA from GWAS European controls (Kidrisk_1-118) and 62 tumor DNA samples (TUM_1-62) from UM tumor cohort were kept for analysis.

ID	Sex	Age	Cohort	rs452384 genotype
UM_GL_1	Female	46	UM germline	TC
UM_GL_2	Male	79	UM germline	TC
UM_GL_3	Female	79	UM germline	TT
UM_GL_4	Female	31	UM germline	TC
UM_GL_5	Female	40	UM germline	TC
UM_GL_6	Male	70	UM germline	CC
UM_GL_7	Female	76	UM germline	TT
UM_GL_8	Male	29	UM germline	TT
UM_GL_9	Male	78	UM germline	CC
UM_GL_10	Female	66	UM germline	TT
UM_GL_11	Male	72	UM germline	CC
UM_GL_12	Male	59	UM germline	CC
UM_GL_13	Female	64	UM germline	TC
UM_GL_14	Male	52	UM germline	CC
UM_GL_15	Male	72	UM germline	CC
UM_GL_16	Male	62	UM germline	CC
UM_GL_17	Female	84	UM germline	TT
UM_GL_18	Female	85	UM germline	TC
UM_GL_19	Female	75	UM germline	CC
UM_GL_20	Female	77	UM germline	TC
UM_GL_21	Female	25	UM germline	TC
UM_GL_22	Female	84	UM germline	CC
UM_GL_23	Female	72	UM germline	TC
UM_GL_24	Male	77	UM germline	TC
UM_GL_25	Female	65	UM germline	TT
UM_GL_26	Male	65	UM germline	TT
UM_GL_27	Female	61	UM germline	TT
UM_GL_28	Male	36	UM germline	TC
UM_GL_29	Male	73	UM germline	TC
UM_GL_30	Male	70	UM germline	TC
UM_GL_31	Male	85	UM germline	TC
UM_GL_32	Female	44	UM germline	CC
UM_GL_33	Female	59	UM germline	TC
UM_GL_34	Male	54	UM germline	TC
UM_GL_35	Male	79	UM germline	CC
UM_GL_36	Female	53	UM germline	TC
UM_GL_37	Female	78	UM germline	TC
UM_GL_38	Male	78	UM germline	CC
UM_GL_39	Male	44	UM germline	TC
UM_GL_40	Female	72	UM germline	TC
UM_GL_41	Male	67	UM germline	TC
UM_GL_42	Female	75	UM germline	CC
UM_GL_43	Male	74	UM germline	TC
UM_GL_44	Female	83	UM germline	TC
UM_GL_45	Male	44	UM germline	TC
UM_GL_46	Male	72	UM germline	CC

UM_GL_47	Male	66	UM germline	CC
UM_GL_48	Female	83	UM germline	TC
UM_GL_49	Female	49	UM germline	CC
UM_GL_50	Female	72	UM germline	CC
UM_GL_51	Female	74	UM germline	TT
UM_GL_52	Female	63	UM germline	CC
UM_GL_53	Female	29	UM germline	TT
UM_GL_54	Male	68	UM germline	CC
UM_GL_55	Male	46	UM germline	CC
UM_GL_56	Male	60	UM germline	CC
UM_GL_57	Male	31	UM germline	TC
UM_GL_58	Male	55	UM germline	TC
UM_GL_59	Male	70	UM germline	CC
UM_GL_60	Female	71	UM germline	TC
UM_GL_61	Male	36	UM germline	CC
UM_GL_62	Male	77	UM germline	TT
UM_GL_63	Female	86	UM germline	CC
UM_GL_64	Female	24	UM germline	CC
UM_GL_65	Male	75	UM germline	CC
UM_GL_66	Female	73	UM germline	CC
UM_GL_67	Female	61	UM germline	TT
UM_GL_68	Female	65	UM germline	CC
UM_GL_69	Female	49	UM germline	TC
UM_GL_70	Female	81	UM germline	TT
UM_GL_71	Male	58	UM germline	CC
UM_GL_72	Female	60	UM germline	TC
UM_GL_73	Female	47	UM germline	TC
UM_GL_74	Female	64	UM germline	TT
UM_GL_75	Female	55	UM germline	CC
UM_GL_76	Male	51	UM germline	CC
UM_GL_77	Male	44	UM germline	TC
UM_GL_78	Female	73	UM germline	TC
UM_GL_79	Female	84	UM germline	TT
UM_GL_80	Male	41	UM germline	TC
UM_GL_81	Male	72	UM germline	TT
UM_GL_82	Male	74	UM germline	TC
UM_GL_83	Female	77	UM germline	CC
UM_GL_84	Female	78	UM germline	TC
UM_GL_85	Female	73	UM germline	TC
UM_GL_86	Female	77	UM germline	TC
UM_GL_87	Female	46	UM germline	TC
UM_GL_88	Male	50	UM germline	TC
UM_GL_89	Male	82	UM germline	TC
UM_GL_90	Female	83	UM germline	TC
UM_GL_91	Female	69	UM germline	TT
UM_GL_92	Male	59	UM germline	CC
UM_GL_93	Male	54	UM germline	TC
UM_GL_94	Male	74	UM germline	CC
UM_GL_95	Female	82	UM germline	TC
UM_GL_96	Male	55	UM germline	TT
UM_GL_97	Female	67	UM germline	TC
UM_GL_98	Female	57	UM germline	CC
UM_GL_99	Female	56	UM germline	TT
UM_GL_100	Male	65	UM germline	TT
UM_GL_101	Male	82	UM germline	TC
UM_GL_102	Male	77	UM germline	TT

UM_GL_103	Female	59	UM germline	CC
UM_GL_104	Male	76	UM germline	TT
UM_GL_105	Male	70	UM germline	TT
UM_GL_106	Male	47	UM germline	TC
UM_GL_107	Male	65	UM germline	TT
UM_GL_108	Male	33	UM germline	CC
UM_GL_109	Female	35	UM germline	TC
UM_GL_110	Female	79	UM germline	TT
UM_GL_111	Male	60	UM germline	TC
UM_GL_112	Female	79	UM germline	TC
UM_GL_113	Male	67	UM germline	CC
UM_GL_114	Female	53	UM germline	TC
UM_GL_115	Male	70	UM germline	CC
UM_GL_116	Male	63	UM germline	TC
UM_GL_117	Male	80	UM germline	TC
UM_GL_118	Male	34	UM germline	TC
UM_GL_119	Female	85	UM germline	TC
UM_GL_120	Female	60	UM germline	TC
UM_GL_121	Male	69	UM germline	TT
UM_GL_122	Male	58	UM germline	TC
UM_GL_123	Female	52	UM germline	TT
UM_GL_124	Male	78	UM germline	TC
UM_GL_125	Female	78	UM germline	TT
UM_GL_126	Female	63	UM germline	TC
UM_GL_127	Female	63	UM germline	TC
UM_GL_128	Female	62	UM germline	TT
UM_GL_129	Male	66	UM germline	TC
UM_GL_130	Male	53	UM germline	CC
UM_GL_131	Male	70	UM germline	TC
UM_GL_132	Male	62	UM germline	TT
UM_GL_133	Male	68	UM germline	CC
UM_GL_134	Female	75	UM germline	TC
UM_GL_135	Male	60	UM germline	TT
UM_GL_136	Female	70	UM germline	CC
UM_GL_137	Female	74	UM germline	TC
UM_GL_138	Female	53	UM germline	CC
UM_GL_139	Female	53	UM germline	TC
UM_GL_140	Male	76	UM germline	TC
UM_GL_141	Female	70	UM germline	TC
UM_GL_142	Female	77	UM germline	TT
UM_GL_143	Male	64	UM germline	CC
UM_GL_144	Male	73	UM germline	CC
UM_GL_145	Female	63	UM germline	CC
UM_GL_146	Female	67	UM germline	TC
UM_GL_147	Female	60	UM germline	TC
UM_GL_148	Male	68	UM germline	TC
UM_GL_149	Male	76	UM germline	CC
UM_GL_150	Female	87	UM germline	CC
UM_GL_151	Female	59	UM germline	TC
UM_GL_152	Female	75	UM germline	TC
UM_GL_153	Male	69	UM germline	TC
UM_GL_154	Male	70	UM germline	TC
UM_GL_155	Female	64	UM germline	CC
UM_GL_156	Male	79	UM germline	TC
UM_GL_157	Female	79	UM germline	TC
UM_GL_158	Female	72	UM germline	TC

UM_GL_159	Female	84	UM germline	TC
UM_GL_160	Female	84	UM germline	TT
UM_GL_161	Male	77	UM germline	CC
UM_GL_162	Female	54	UM germline	CC
UM_GL_163	Male	37	UM germline	CC
UM_GL_164	Female	50	UM germline	TC
UM_GL_165	Male	67	UM germline	TT
UM_GL_166	Female	64	UM germline	TT
UM_GL_167	Female	55	UM germline	TC
UM_GL_168	Female	79	UM germline	TT
UM_GL_169	Male	73	UM germline	CC
UM_GL_170	Female	58	UM germline	CC
UM_GL_171	Male	51	UM germline	TC
UM_GL_172	Female	72	UM germline	CC
UM_GL_173	Male	50	UM germline	TT
UM_GL_174	Female	79	UM germline	CC
UM_GL_175	Male	78	UM germline	TC
UM_GL_176	Female	66	UM germline	TC
UM_GL_177	Female	63	UM germline	TT
UM_GL_178	Male	54	UM germline	CC
UM_GL_179	Male	46	UM germline	TC
UM_GL_180	Male	66	UM germline	TC
UM_GL_181	Female	60	UM germline	TT
UM_GL_182	Male	57	UM germline	CC
UM_GL_183	Male	68	UM germline	TC
UM_GL_184	Male	46	UM germline	CC
UM_GL_185	Female	69	UM germline	CC
UM_GL_186	Female	48	UM germline	TC
UM_GL_187	Female	65	UM germline	TC
UM_GL_188	Female	67	UM germline	TC
UM_GL_189	Female	65	UM germline	CC
UM_GL_190	Male	48	UM germline	TC
UM_GL_191	Female	33	UM germline	TC
UM_GL_192	Female	45	UM germline	TC
UM_GL_193	Male	60	UM germline	TC
UM_GL_194	Male	59	UM germline	TC
UM_GL_195	Female	45	UM germline	TT
UM_GL_196	Female	76	UM germline	CC
UM_GL_197	Male	85	UM germline	TC
UM_GL_198	Female	59	UM germline	TC
UM_GL_199	Male	59	UM germline	TC
UM_GL_200	Female	52	UM germline	CC
UM_GL_201	Male	59	UM germline	TC
UM_GL_202	Male	53	UM germline	TT
UM_GL_203	Male	71	UM germline	CC
UM_GL_204	Female	81	UM germline	CC
UM_GL_205	Male	71	UM germline	TC
UM_GL_206	Female	84	UM germline	TC
UM_GL_207	Female	57	UM germline	TC
UM_GL_208	Male	80	UM germline	TT
UM_GL_209	Female	57	UM germline	TT
UM_GL_210	Female	57	UM germline	TT
UM_GL_211	Female	65	UM germline	CC
UM_GL_212	Male	83	UM germline	TC
UM_GL_213	Male	45	UM germline	TC
UM_GL_214	Male	76	UM germline	TC

UM_GL_215	Female	66	UM germline	CC
UM_GL_216	Female	69	UM germline	TT
UM_GL_217	Female	72	UM germline	TC
UM_GL_218	Female	73	UM germline	TC
UM_GL_219	Female	46	UM germline	CC
UM_GL_220	Female	36	UM germline	TT
UM_GL_221	Male	37	UM germline	TC
UM_GL_222	Male	86	UM germline	TC
UM_GL_223	Male	70	UM germline	TT
UM_GL_224	Male	67	UM germline	TT
UM_GL_225	Male	65	UM germline	TC
UM_GL_226	Female	67	UM germline	TC
UM_GL_227	Female	65	UM germline	CC
UM_GL_228	Female	64	UM germline	CC
UM_GL_229	Female	80	UM germline	TT
UM_GL_230	Male	64	UM germline	TT
UM_GL_231	Female	54	UM germline	TC
UM_GL_232	Female	34	UM germline	TT
UM_GL_233	Male	52	UM germline	TC
UM_GL_234	Female	68	UM germline	TC
UM_GL_235	Male	64	UM germline	TC
UM_GL_236	Male	79	UM germline	TC
UM_GL_237	Female	60	UM germline	TC
UM_GL_238	Male	62	UM germline	TC
UM_GL_239	Male	67	UM germline	TC
UM_GL_240	Male	70	UM germline	TC
UM_GL_241	Male	31	UM germline	CC
UM_GL_242	Female	59	UM germline	TC
UM_GL_243	Male	71	UM germline	CC
UM_GL_244	Male	66	UM germline	TC
UM_GL_245	Female	59	UM germline	TT
UM_GL_246	Female	62	UM germline	TC
UM_GL_247	Female	55	UM germline	TC
UM_GL_248	Male	30	UM germline	TC
UM_GL_249	Male	63	UM germline	TC
UM_GL_250	Female	40	UM germline	TC
UM_GL_251	Female	61	UM germline	TC
UM_GL_252	Female	61	UM germline	TT
UM_GL_253	Female	71	UM germline	TC
UM_GL_254	Female	82	UM germline	CC
UM_GL_255	Female	33	UM germline	TC
UM_GL_256	Female	80	UM germline	TC
UM_GL_257	Male	26	UM germline	TT
UM_GL_258	Female	67	UM germline	TT
UM_GL_259	Female	70	UM germline	TT
UM_GL_260	Male	57	UM germline	TC
UM_GL_261	Male	76	UM germline	CC
UM_GL_262	Male	74	UM germline	CC
UM_GL_263	Male	59	UM germline	TC
UM_GL_264	Female	85	UM germline	TT
UM_GL_265	Male	55	UM germline	TT
UM_GL_266	Female	66	UM germline	TC
UM_GL_267	Female	86	UM germline	TT
UM_GL_268	Male	73	UM germline	CC
UM_GL_269	Female	42	UM germline	CC
UM_GL_270	Female	73	UM germline	CC

UM_GL_271	Female	72	UM germline	TC
UM_GL_272	Male	67	UM germline	CC
UM_GL_273	Female	83	UM germline	TT
UM_GL_274	Female	57	UM germline	CC
UM_GL_275	Male	38	UM germline	CC
UM_GL_276	Female	75	UM germline	CC
UM_GL_277	Female	47	UM germline	TC
UM_GL_278	Male	50	UM germline	TT
UM_GL_279	Female	55	UM germline	TC
UM_GL_280	Male	70	UM germline	TC
UM_GL_281	Male	49	UM germline	CC
UM_GL_282	Female	81	UM germline	CC
UM_GL_283	Female	71	UM germline	TC
UM_GL_284	Female	65	UM germline	TT
UM_GL_285	Male	52	UM germline	CC
UM_GL_286	Female	86	UM germline	TC
UM_GL_287	Male	77	UM germline	TC
UM_GL_288	Female	54	UM germline	TC
UM_GL_289	Female	82	UM germline	TT
UM_GL_290	Male	50	UM germline	CC
UM_GL_291	Male	65	UM germline	TC
UM_GL_292	Male	47	UM germline	TC
UM_GL_293	Female	67	UM germline	TC
UM_GL_294	Male	66	UM germline	TC
UM_GL_295	Male	47	UM germline	TT
UM_GL_296	Male	68	UM germline	CC
UM_GL_297	Male	46	UM germline	CC
UM_GL_298	Female	32	UM germline	TC
UM_GL_299	Male	66	UM germline	TC
UM_GL_300	Female	32	UM germline	TC
UM_GL_301	Male	42	UM germline	CC
UM_GL_302	Female	87	UM germline	TT
UM_GL_303	Female	59	UM germline	TT
UM_GL_304	Female	67	UM germline	CC
UM_GL_305	Female	85	UM germline	TC
UM_GL_306	Female	46	UM germline	TC
UM_GL_307	Female	48	UM germline	CC
UM_GL_308	Female	78	UM germline	TT
UM_GL_309	Female	67	UM germline	TC
UM_GL_310	Male	59	UM germline	TT
UM_GL_311	Male	62	UM germline	TC
UM_GL_312	Male	50	UM germline	CC
UM_GL_313	Female	53	UM germline	TC
UM_GL_314	Female	57	UM germline	TC
UM_GL_315	Female	65	UM germline	TC
UM_GL_316	Female	62	UM germline	CC
UM_GL_317	Female	67	UM germline	TC
UM_GL_318	Male	52	UM germline	TC
UM_GL_319	Male	48	UM germline	CC
UM_GL_320	Female	63	UM germline	TT
UM_GL_321	Female	83	UM germline	TC
UM_GL_322	Female	73	UM germline	TT
UM_GL_323	Male	75	UM germline	TC
UM_GL_324	Male	62	UM germline	TT
UM_GL_325	Female	56	UM germline	CC
UM_GL_326	Male	75	UM germline	CC

UM_GL_327	Male	68	UM germline	TC
UM_GL_328	Female	70	UM germline	TC
UM_GL_329	Female	64	UM germline	TC
UM_GL_330	Female	70	UM germline	TC
UM_GL_331	Female	60	UM germline	TT
UM_GL_332	Male	65	UM germline	TT
UM_GL_333	Male	57	UM germline	CC
UM_GL_334	Male	75	UM germline	TT
UM_GL_335	Female	71	UM germline	TT
UM_GL_336	Female	43	UM germline	TC
UM_GL_337	Female	74	UM germline	TC
UM_GL_338	Male	64	UM germline	CC
UM_GL_339	Male	72	UM germline	TT
UM_GL_340	Male	51	UM germline	TT
UM_GL_341	Female	76	UM germline	TC
UM_GL_342	Male	71	UM germline	CC
UM_GL_343	Female	64	UM germline	TT
UM_GL_344	Female	29	UM germline	TC
UM_GL_345	Female	58	UM germline	CC
UM_GL_346	Male	65	UM germline	CC
UM_GL_347	Male	63	UM germline	TT
UM_GL_348	Female	56	UM germline	TC
UM_GL_349	Female	50	UM germline	TC
UM_GL_350	Male	52	UM germline	TC
UM_GL_351	Female	59	UM germline	TC
UM_GL_352	Male	78	UM germline	TC
UM_GL_353	Male	72	UM germline	TC
UM_GL_354	Female	50	UM germline	TC
UM_GL_355	Female	57	UM germline	CC
UM_GL_356	Male	64	UM germline	TC
UM_GL_357	Female	49	UM germline	TC
UM_GL_358	Male	84	UM germline	TT
UM_GL_359	Male	63	UM germline	TC
UM_GL_360	Female	84	UM germline	TT
UM_GL_361	Male	62	UM germline	TT
UM_GL_362	Male	35	UM germline	CC
UM_GL_363	Male	63	UM germline	TC
UM_GL_364	Male	68	UM germline	CC
UM_GL_365	Male	64	UM germline	TC
UM_GL_366	Male	45	UM germline	TT
UM_GL_367	Female	66	UM germline	TT
UM_GL_368	Male	70	UM germline	TT
UM_GL_369	Female	40	UM germline	TC
UM_GL_370	Male	79	UM germline	TC
UM_GL_371	Female	74	UM germline	TC
UM_GL_372	Male	64	UM germline	TC
UM_GL_373	Male	87	UM germline	TT
UM_GL_374	Female	43	UM germline	TT
UM_GL_375	Female	83	UM germline	CC
UM_GL_376	Male	72	UM germline	TC
UM_GL_377	Female	55	UM germline	TC
UM_GL_378	Male	52	UM germline	CC
UM_GL_379	Male	64	UM germline	TC
UM_GL_380	Female	50	UM germline	CC
UM_GL_381	Female	60	UM germline	TC
UM_GL_382	Male	74	UM germline	CC

UM_GL_383	Female	62	UM germline	TT
UM_GL_384	Male	80	UM germline	TC
UM_GL_385	Male	62	UM germline	TT
UM_GL_386	Male	68	UM germline	TT
UM_GL_387	Male	84	UM germline	TC
UM_GL_388	Female	38	UM germline	TC
UM_GL_389	Female	74	UM germline	TC
UM_GL_390	Female	67	UM germline	TT
UM_GL_391	Male	51	UM germline	TC
UM_GL_392	Female	83	UM germline	CC
UM_GL_393	Male	71	UM germline	CC
UM_GL_394	Female	72	UM germline	TC
UM_GL_395	Male	60	UM germline	TT
UM_GL_396	Male	67	UM germline	CC
UM_GL_397	Male	71	UM germline	TC
UM_GL_398	Female	49	UM germline	TC
UM_GL_399	Male	86	UM germline	TC
UM_GL_400	Male	78	UM germline	TT
UM_GL_401	Male	69	UM germline	TC
UM_GL_402	Female	76	UM germline	TC
UM_GL_403	Female	64	UM germline	TC
UM_GL_404	Male	63	UM germline	CC
UM_GL_405	Female	70	UM germline	CC
UM_GL_406	Male	70	UM germline	TC
UM_GL_407	Female	51	UM germline	CC
UM_GL_408	Male	74	UM germline	TC
UM_GL_409	Male	62	UM germline	TC
UM_GL_410	Female	48	UM germline	CC
UM_GL_411	Female	43	UM germline	TC
UM_GL_412	Male	58	UM germline	TC
UM_GL_413	Female	73	UM germline	TC
UM_GL_414	Male	52	UM germline	TT
UM_GL_415	Male	57	UM germline	TC
UM_GL_416	Female	65	UM germline	TT
UM_GL_417	Male	64	UM germline	TC
UM_GL_418	Female	57	UM germline	TC
UM_GL_419	Female	79	UM germline	TT
UM_GL_420	Female	87	UM germline	TC
UM_GL_421	Female	64	UM germline	TT
UM_GL_422	Female	72	UM germline	TC
UM_GL_423	Female	71	UM germline	TC
UM_GL_424	Female	73	UM germline	TC
UM_GL_425	Male	51	UM germline	TC
UM_GL_426	Male	54	UM germline	TC
UM_GL_427	Male	50	UM germline	TC
UM_GL_428	Male	71	UM germline	TT
UM_GL_429	Male	49	UM germline	TT
UM_GL_430	Male	61	UM germline	CC
UM_GL_431	Female	50	UM germline	TT
UM_GL_432	Male	74	UM germline	CC
UM_GL_433	Male	58	UM germline	CC
Kidrisk_1	Male	72	Control germline	TC
Kidrisk_2	Male	74	Control germline	TT
Kidrisk_3	Male	65	Control germline	TT
Kidrisk_4	Male	73	Control germline	TC
Kidrisk_5	Male	73	Control germline	TT

Kidrisk_6	Male	64	Control germline	TT
Kidrisk_7	Male	70	Control germline	TC
Kidrisk_8	Male	59	Control germline	TC
Kidrisk_9	Male	74	Control germline	TC
Kidrisk_10	Male	74	Control germline	TC
Kidrisk_11	Male	63	Control germline	TT
Kidrisk_12	Male	70	Control germline	TT
Kidrisk_13	Male	70	Control germline	TC
Kidrisk_14	Male	75	Control germline	TC
Kidrisk_15	Male	74	Control germline	TT
Kidrisk_16	Male	72	Control germline	TT
Kidrisk_17	Male	72	Control germline	TC
Kidrisk_18	Male	68	Control germline	TC
Kidrisk_19	Male	69	Control germline	TT
Kidrisk_20	Male	70	Control germline	CC
Kidrisk_21	Male	74	Control germline	TT
Kidrisk_22	Male	68	Control germline	TC
Kidrisk_23	Male	70	Control germline	TC
Kidrisk_24	Male	63	Control germline	TC
Kidrisk_25	Male	77	Control germline	TT
Kidrisk_26	Male	70	Control germline	TT
Kidrisk_27	Male	65	Control germline	CC
Kidrisk_28	Male	65	Control germline	CC
Kidrisk_29	Male	70	Control germline	CC
Kidrisk_30	Male	75	Control germline	TT
Kidrisk_31	Male	75	Control germline	TC
Kidrisk_32	Male	74	Control germline	TT
Kidrisk_33	Male	73	Control germline	CC
Kidrisk_34	Male	73	Control germline	TC
Kidrisk_35	Male	69	Control germline	TC
Kidrisk_36	Male	71	Control germline	CC
Kidrisk_37	Male	68	Control germline	TT
Kidrisk_38	Male	73	Control germline	TC
Kidrisk_39	Male	73	Control germline	TT
Kidrisk_40	Male	72	Control germline	TC
Kidrisk_41	Male	75	Control germline	CC
Kidrisk_42	Male	65	Control germline	TC
Kidrisk_43	Male	71	Control germline	CC
Kidrisk_44	Male	72	Control germline	TC
Kidrisk_45	Male	73	Control germline	TC
Kidrisk_46	Male	72	Control germline	TC
Kidrisk_47	Male	68	Control germline	TT
Kidrisk_48	Male	71	Control germline	TT
Kidrisk_49	Male	75	Control germline	TC
Kidrisk_50	Male	71	Control germline	TC
Kidrisk_51	Male	77	Control germline	TC
Kidrisk_52	Male	80	Control germline	TT
Kidrisk_53	Male	77	Control germline	TC
Kidrisk_54	Male	76	Control germline	CC
Kidrisk_55	Male	77	Control germline	CC
Kidrisk_56	Male	80	Control germline	TT
Kidrisk_57	Male	69	Control germline	TT
Kidrisk_58	Male	75	Control germline	TC
Kidrisk_59	Male	71	Control germline	TT
Kidrisk_60	Male	53	Control germline	CC
Kidrisk_61	Male	60	Control germline	CC

Kidrisk_62	Male	62	Control germline	CC
Kidrisk_63	Male	64	Control germline	TT
Kidrisk_64	Male	69	Control germline	TC
Kidrisk_65	Male	69	Control germline	CC
Kidrisk_66	Male	65	Control germline	TC
Kidrisk_67	Male	72	Control germline	CC
Kidrisk_68	Male	71	Control germline	TC
Kidrisk_69	Male	71	Control germline	TT
Kidrisk_70	Male	71	Control germline	CC
Kidrisk_71	Male	45	Control germline	CC
Kidrisk_72	Male	64	Control germline	TC
Kidrisk_73	Male	59	Control germline	TC
Kidrisk_74	Male	50	Control germline	TT
Kidrisk_75	Male	56	Control germline	TC
Kidrisk_76	Male	63	Control germline	TC
Kidrisk_77	Male	63	Control germline	TT
Kidrisk_78	Male	68	Control germline	TT
Kidrisk_79	Male	68	Control germline	TC
Kidrisk_80	Male	77	Control germline	TC
Kidrisk_81	Male	57	Control germline	TC
Kidrisk_82	Male	54	Control germline	CC
Kidrisk_83	Male	76	Control germline	TC
Kidrisk_84	Male	60	Control germline	TT
Kidrisk_85	Male	77	Control germline	CC
Kidrisk_86	Male	62	Control germline	TC
Kidrisk_87	Male	62	Control germline	TT
Kidrisk_88	Male	53	Control germline	TC
Kidrisk_89	Male	60	Control germline	TT
Kidrisk_90	Male	62	Control germline	TT
Kidrisk_91	Male	60	Control germline	TC
Kidrisk_92	Male	66	Control germline	CC
Kidrisk_93	Male	67	Control germline	CC
Kidrisk_94	Male	62	Control germline	TT
Kidrisk_95	Male	62	Control germline	TT
Kidrisk_96	Male	71	Control germline	TC
Kidrisk_97	Male	80	Control germline	TC
Kidrisk_98	Male	64	Control germline	TC
Kidrisk_99	Male	67	Control germline	TC
Kidrisk_100	Male	75	Control germline	TC
Kidrisk_101	Male	62	Control germline	TT
Kidrisk_102	Male	66	Control germline	TC
Kidrisk_103	Male	71	Control germline	TT
Kidrisk_104	Male	62	Control germline	CC
Kidrisk_105	Male	84	Control germline	TT
Kidrisk_106	Male	51	Control germline	CC
Kidrisk_107	Male	60	Control germline	TT
Kidrisk_108	Male	60	Control germline	TC
Kidrisk_109	Male	64	Control germline	TC
Kidrisk_110	Male	81	Control germline	TT
Kidrisk_111	Male	45	Control germline	TT
Kidrisk_112	Male	50	Control germline	CC
Kidrisk_113	Male	62	Control germline	TC
Kidrisk_114	Male	49	Control germline	TT
Kidrisk_115	Male	44	Control germline	CC
Kidrisk_116	Male	53	Control germline	CC
Kidrisk_117	Male	46	Control germline	CC

Kidrisk_118	Male	54	Control germline	TT
TUM_1	Female	65	UM tumor DNA	TC
TUM_2	Female	67	UM tumor DNA	CC
TUM_3	Female	31	UM tumor DNA	TC
TUM_4	Male	78	UM tumor DNA	TC
TUM_5	Female	59	UM tumor DNA	CC
TUM_6	Female	50	UM tumor DNA	TC
TUM_7	Female	68	UM tumor DNA	CC
TUM_8	Female	90	UM tumor DNA	CC
TUM_9	Male	80	UM tumor DNA	CC
TUM_10	Female	55	UM tumor DNA	TC
TUM_11	Male	25	UM tumor DNA	TT
TUM_12	Male	64	UM tumor DNA	TT
TUM_13	Female	73	UM tumor DNA	TC
TUM_14	Male	60	UM tumor DNA	TT
TUM_15	Male	77	UM tumor DNA	TC
TUM_16	Female	73	UM tumor DNA	TC
TUM_17	Male	56	UM tumor DNA	CC
TUM_18	Female	83	UM tumor DNA	TC
TUM_19	Male	46	UM tumor DNA	TC
TUM_20	Female	54	UM tumor DNA	TC
TUM_21	Female	66	UM tumor DNA	CC
TUM_22	Male	64	UM tumor DNA	TC
TUM_23	Female	63	UM tumor DNA	TC
TUM_24	Male	75	UM tumor DNA	TC
TUM_25	Female	51	UM tumor DNA	TT
TUM_26	Male	69	UM tumor DNA	CC
TUM_27	Female	75	UM tumor DNA	TC
TUM_28	Male	81	UM tumor DNA	TC
TUM_29	Male	69	UM tumor DNA	TC
TUM_30	Female	61	UM tumor DNA	CC
TUM_31	Female	93	UM tumor DNA	TT
TUM_32	Male	47	UM tumor DNA	TT
TUM_33	Male	58	UM tumor DNA	CC
TUM_34	Female	60	UM tumor DNA	CC
TUM_35	Male	59	UM tumor DNA	CC
TUM_36	Female	87	UM tumor DNA	CC
TUM_37	Female	49	UM tumor DNA	TT
TUM_38	Male	69	UM tumor DNA	TT
TUM_39	Male	54	UM tumor DNA	TT
TUM_40	Female	54	UM tumor DNA	TT
TUM_41	Female	68	UM tumor DNA	TC
TUM_42	Female	78	UM tumor DNA	CC
TUM_43	Female	63	UM tumor DNA	CC
TUM_44	Female	69	UM tumor DNA	TT
TUM_45	Female	48	UM tumor DNA	TC
TUM_46	Female	48	UM tumor DNA	TC
TUM_47	Female	85	UM tumor DNA	CC
TUM_48	Female	51	UM tumor DNA	CC
TUM_49	Female	64	UM tumor DNA	TT
TUM_50	Male	63	UM tumor DNA	CC
TUM_51	Male	61	UM tumor DNA	TC
TUM_52	Female	43	UM tumor DNA	CC
TUM_53	Female	60	UM tumor DNA	TC
TUM_54	Female	55	UM tumor DNA	CC
TUM_55	Male	38	UM tumor DNA	TC

TUM_56	Female	41	UM tumor DNA	CC
TUM_57	Male	73	UM tumor DNA	TT
TUM_58	Female	56	UM tumor DNA	CC
TUM_59	Male	63	UM tumor DNA	TC
TUM_60	Male	53	UM tumor DNA	TC
TUM_61	Female	57	UM tumor DNA	CC
TUM_62	Female	79	UM tumor DNA	CC

Supplementary Table 2: list of all primer pairs (forward and reverse, 5' to 3' orientation) used for luciferase assays, site-directed mutagenesis, EMSA, mass spectrometry and qPCR experiments

Experiment	PCR amplicon	Forward primer (5'-3')	Reverse primer (5'-3')
CLPTM1L insert cloning into pGL3-P vector (In-Fusion cloning primers)	Full-length insert (3,044 bp)	CCCGGGCTCGAGATCTTCTACAGACACATTTTCATCC	TGCAGATCGCAGATCTAGCTTCTCTTTGATTTCTG
	Fragment A (2,509 bp)	CCCGGGCTCGAGATCTTCTACAGACACATTTTCATCC	TGCAGATCGCAGATCTAGTTAATAGGCCGTGGTG
	Fragment B (1,444 bp)	CCCGGGCTCGAGATCTTCTACAGACACATTTTCATCC	TGCAGATCGCAGATCTATTGAGGTGAGTTCGGGC
	Fragment C (251 bp)	CCCGGGCTCGAGATCTCTCCCTGGGCTTTACGGTG	TGCAGATCGCAGATCTAGCTTCTCTTTGATTTCTG
	Fragment D (840 bp)	CCCGGGCTCGAGATCTTGCTCTCACTCTAACCACAG	TGCAGATCGCAGATCTAGCTTCTCTTTGATTTCTG
	Fragment E (1,335 bp)	CCCGGGCTCGAGATCTGCAGCTAGGAAAGTGTGGAAAG	TGCAGATCGCAGATCTAGCTTCTCTTTGATTTCTG
	Fragment F (1,827 bp)	CCCGGGCTCGAGATCTTCATGAGCTGATACCACCG	TGCAGATCGCAGATCTAGCTTCTCTTTGATTTCTG
Site-directed mutagenesis	Fragment G (2,519 bp)	CCCGGGCTCGAGATCTTGGGATTTCTCAAGCACATGAAAAC	TGCAGATCGCAGATCTAGCTTCTCTTTGATTTCTG
	rs452932 T to C	CACCACTGCTGCCATCCACAGCTTCCAGAACACTGAGG	CCTCAGTGTTCTGGAAGCTGTGGATGGCAGGCAGTGGGTG
	rs452384 T to C	CTTATGCCAGTTCTGTGATAAGCGCGTGGCTTTTCAAGTCTGG	CCAGACTTGAAAAGCCACGCGCTTATCACAGAAGTGGCATAAG
	rs370348 G to A	CCTCACACCATCAACCGCTCGCCCTGTCCAGAC	GTCTGGACAGGGGCGAGCGGTGATGGGTGTGAGG
	rs452932 C to T	CACCACTGCTGCCATCCACAGCTTCCAGAACACTGAGG	CCTCAGTGTTCTGGAAGCTGTAGGATGGCAGGCAGTGGGTG
EMSA probes	rs452384 C to T	CTTATGCCAGTTCTGTGATAAGTGCCTGGCTTTTCAAGTCTGG	CCAGACTTGAAAAGCCACGCACTTATCACAGAAGTGGCATAAG
	rs452384-3'biotin (low-risk T allele)	TTCTGTGATAAGTGCCTGGCTTTT[BIO]	GAAAAGCCACGCACTTATCACAGAA[BIO]
	rs452932-3'biotin (low-risk T allele)	TGCCTGCCATCTACAGCTTCCAGA[BIO]	TCTGGAAGCTGTAGGATGGCAGGCA[BIO]
	rs452384-3'biotin (high-risk C allele)	TTCTGTGATAAGCGCGTGGCTTTT[BIO]	GAAAAGCCACGCGCTTATCACAGAA[BIO]
	rs452932-3'biotin (high-risk C allele)	TGCCTGCCATCCACAGCTTCCAGA[BIO]	TCTGGAAGCTGTGGGATGGCAGGCA[BIO]
	rs452384-unlabeled (low-risk T allele)	TTCTGTGATAAGTGCCTGGCTTTT	GAAAAGCCACGCACTTATCACAGAA
	rs452932-unlabeled (low-risk T allele)	TGCCTGCCATCTACAGCTTCCAGA	TCTGGAAGCTGTAGGATGGCAGGCA
DNA pulldown / mass spectrometry probes	rs452384-unlabeled (high-risk C allele)	TTCTGTGATAAGCGCGTGGCTTTT	GAAAAGCCACGCGCTTATCACAGAA
	rs452932-unlabeled (high-risk C allele)	TGCCTGCCATCCACAGCTTCCAGA	TCTGGAAGCTGTGGGATGGCAGGCA
	rs452384-T biotin probe	[BIO]TTCTGTGATAAGTGCCTGGCTTTT	GAAAAGCCACGCACTTATCACAGAA
	rs452384-C biotin probe	[BIO]TTCTGTGATAAGCGCGTGGCTTTT	GAAAAGCCACGCGCTTATCACAGAA
ChIP-qPCR	negative control probe	[BIO]AGAGTGGTCACTACCCCTCTG	CAGAGGGGGTAGTGACCACTCT
	CLPTM1L rs452384	CTTCCGGCGGTGTCTTAT	GGGAGGAAACAAATCCAGAC
	CLPTM1L exon3 (neg1)	CCCAAGCCAGAAGAAATCAA	GCCTGTTATCAGTAACCCAT
	CLPTM1L intron4	AGCCTGGTAAACATGGTGAA	GATTCAAGCGATTCTCGTTC
qPCR primers	CLPTM1L exon4 (neg2)	GAGGACCCGTCAAAGACAAA	TCTGTTTTCTCACTTAACA
	NXK2.4	CATGGGCAGCTACTGCAAC	TGAACCTGGAGATTGACGAG
	GATA4	CACAAGATGAACGGCATCAA	CGTGGAGCTTCATGTAGAGG
	TBP	CTGGCCCATAGTGATCTTT	GCTGGAAGTCTCTCATATTTC
Telomere length qPCR	GAPDH	CTCCTGCACCACCACTGCT	GGGCCATCCACAGTCTTCTG
	Telomeric repeats	CGGTTTGTGGTTTGGGTTTGGGTTTGGGTT	GGCTTGCCTTACCCCTTACCCCTACCCCTACCCCT
	HMBS single copy gene	GTCCAGCCTGTTGGTGAG	GGCCACAATTCAGATCTTCTTA

Supplementary Table 3: List of all proteins enriched in either rs452384-T or rs452384-C probe among proteins enriched over negative control probe (NEG) by quantitative mass spectrometry. Filtering criteria for protein selection were as follows: proteins with at least 3 distinct peptides, a fold change ≥ 2 compared to NEG and a Benjamini-Hochberg adjusted p-value of ≤ 0.05 (see Methods) were kept for analysis. This was done for both (T vs NEG) and (C vs NEG) quantifications and other proteins were filtering out. From those, only proteins with an absolute fold change ≥ 1.5 in the (C vs T) quantification, an associated pvalue ≤ 0.05 and ≥ 3 peptides were retained. This was done on proteins enriched in (T vs NEG) quantification (Table 3A) and in proteins enriched in (C vs NEG) quantification (Table 3B). Proteins are ranked by pvalue in the (T vs C) quantification, from most to least significant. Transcription factors are highlighted in bold. MW: molecular weight in kilo daltons (kDa).

Table 3A: (T vs C) quantification in proteins enriched in T vs Neg.

T-allele enriched proteins:

Petptide ID	Protein	T/C				T/Neg				MW (kDa)	Protein description
		Ratio T/C	Log2 (ratio)	Adj. P-value	Number of peptides	Ratio C/neg	Log2 (ratio)	Adj. P-value	Number of peptides		
Q9H2Z4	NKX2-4,NKX2D	13.05	3.71	3.10E-31	8	6.40	2.68	2.02E-28	8	36.2	Homeobox protein Nkx-2.4
Q99729	HNRNPAB,ABBP1,HNRPAB	1.53	0.61	1.81E-22	17	3.27	1.71	2.01E-52	17	36.2	Heterogeneous nuclear
P31274	HOXC9,HOX3B	1.92	0.94	5.45E-17	8	9.78	3.29	7.45E-39	8	29.2	Homeobox protein Hox-C9
P56179	DLX6	2.96	1.56	1.27E-08	5	8.89	3.15	4.67E-19	5	19.7	Homeobox protein DLX-6
Q04695	KRT17	2.57	1.36	0.001	9	2.70	1.43	4.96E-03	9	48.1	Keratin, type I cytoskeletal 17

C-allele enriched proteins:

Petptide ID	Protein	T/C				T/Neg				MW (kDa)	Protein description
		Ratio T/C	Log2 (ratio)	Adj. P-value	Number of peptides	Ratio C/neg	Log2 (ratio)	Adj. P-value	Number of peptides		
P43694	GATA4	0.34	-1.58	2.03E-23	10	13.69	3.77	3.73E-12	5	44.6	Transcription factor GATA-4
Q99697	PITX2,ARP1,RGS,RIEG,RIEG	0.58	-0.79	4.38E-05	5	3.03	1.60	1.69E-06	5	35.4	Pituitary homeobox 2

Table 3B: (C vs T) quantification in proteins enriched in C vs Neg.

C-allele enriched proteins:

Petptide ID	Protein	C/T				C/Neg				MW (kDa)	Protein description
		Ratio C/T	Log2 (ratio)	Adj. P-value	Number of peptides	Ratio C/neg	Log2 (ratio)	Adj. P-value	Number of peptides		
P43694	GATA4	2.98	1.58	2.03E-23	10	36.32	5.18	3.53E-18	5	44.6	Transcription factor GATA-4
Q9UHF7	TRPS1	1.72	0.79	1.60E-08	20	2.76	1.47	1.40E-18	19	141.5	Zinc finger transcription factor Trps1
P26583	HMGB2,HMG2	2.37	1.25	6.19E-06	8	2.66	1.41	3.02E-06	8	24.0	High mobility group protein B2
Q99697	PITX2,ARP1,RGS,RIEG,RIEG	1.73	0.79	4.38E-05	5	4.90	2.29	2.85E-09	5	35.4	Pituitary homeobox 2
P0DP25	CALM3,CALML2,CAM3,CAM3	1.85	0.89	6.50E-05	7	2.15	1.11	2.10E-05	7	16.8	Calmodulin-3
P17096	HMGAI1,HMGIY	2.24	1.17	4.08E-04	4	2.84	1.50	1.09E-05	4	11.7	High mobility group protein HMG-I/HMG-Y
O15347	HMGB3,HMG2A,HMG4	2.24	1.16	9.88E-03	5	3.17	1.66	8.03E-04	6	23.0	High mobility group protein B3

T-allele enriched proteins:

Petptide ID	Protein	C/T				C/Neg				MW (kDa)	Protein description
		Ratio C/T	Log2 (ratio)	Adj. P-value	Number of peptides	Ratio C/neg	Log2 (ratio)	Adj. P-value	Number of peptides		
P31274	HOXC9,HOX3B	0.52	-0.94	5.45E-17	8	4.94	2.30	1.29E-26	8	29.2	Homeobox protein Hox-C9
P56179	DLX6	0.34	-1.56	1.27E-08	5	3.06	1.61	2.07E-08	5	19.7	Homeobox protein DLX-6

CONCLUSION TO ARTICLE 3

FUNCTIONAL CHARACTERIZATION OF CLPTM1L/TERT 5p15.33 RISK LOCUS IN UVEAL MELANOMA IDENTIFIES RS452384 AS A FUNCTIONAL VARIANT REGULATING CLPTM1L AND TERT EXPRESSION THROUGH NKX2.4 ALLELE- SPECIFIC BINDING

This work represents the first steps of a functional genomics study of genetic risk factors in UM, aiming to characterize the biological mechanisms underlying the *TERT/CLPTM1L* susceptibility region. Our main findings are the identification of rs452384 functional variant within *CLPTM1L* intron 8, part of an enhancer regulatory region. This SNP mediates allele-specific transcriptional regulation via binding of transcription factor NKX2.4 to the protective allele T of rs452384; siRNA-mediated *NKX2.4* silencing results in increased *CLPTM1L* and *TERT* expression, making these two genes – known for their implication in multiple other tumor types – plausible candidates in UM tumorigenesis processes. Strikingly, rs452384 is also a likely functional variant in lung adenocarcinoma (LUAD), for which *TERT/CLPTM1L* is also another important susceptibility region, where the LUAD risk allele T (opposite to UM) binds transcription factor NKX2.1 in an allele-specific manner.⁶³⁷

Although we identify rs452384 a functional SNP at the 5p15 region in UM, future work is needed to fully establish causality between rs452384, NKX2.4 specific binding to this SNP, and regulation of *TERT* and/or *CLPTM1L*. Targeting the genomic region surrounding rs452384 by CRISPR-derived methods, such as CRISPR interference (CRISPRi, leading to targeted transcriptional repression) and CRISPR activation (CRISPRa, leading to RNA-mediated transcriptional activation), are complementary methods that would allow us to validate the regulatory effect of rs452384 on gene expression and particularly that of *CLPTM1L* and *TERT*. Considering the low expression of *TERT*, even in most UM cell lines, CRISPRa may prove more successful to observe the modulatory effect of rs452384 on its expression, and potentially telomere length. To confirm the biological role between NKX2.4 transcription factor binding and transcriptional regulation of *TERT* and/or *CLPTM1L*, re-expression of exogenous *NKX2.4* (resistant to siRNA treatment) after siRNA-mediated knockdown of endogenous *NKX2.4* could

be useful to test whether this re-expression can rescue and restore initial *CLPTM1L* and *TERT* expression levels, validating them as specific target genes of NKX2.4.

We hypothesize that rs452384 mediates some of its function via the regulation of telomere length, as the risk C allele is associated with longer telomeres in peripheral blood lymphocytes (PBL) of UM patients and of European controls from the GWAS, consistent with our finding of *TERT* as a target gene in the regulation mediated by rs452384-NKX2.4. The link between 5p15.33 genotypes and *TERT*/telomeres (eQTL and tQTL) is difficult to address in UM tumors and in UM models/cell lines, as *TERT* is only marginally expressed in tumors and UM cell lines (as determined by in-house RNA-seq and qPCR data), and previous studies identified UM as an outlier among 31 tumor types with shortest telomeres,⁶³⁸ which we recently confirmed as telomeres were significantly shorter in UM tumors compared to (unmatched) PBL samples of UM patients.

We are planning to validate our telomere length study (and correlation with rs452384 genotype) by whole-genome sequencing of UM samples. As we hypothesize that rs452384 may regulate its target genes and/or telomere length prior to tumorigenesis, we wish to assess telomere length and *TERT* expression levels in normal uveal melanocytes, although this represents a major challenge as these cells are difficult to extract and access to human eye tissue is very limited.

CLPTM1L is another key candidate in UM tumorigenic process linked to 5p15.33 susceptibility region, given its reported implication in other tumor types such as lung, pancreatic and ovarian cancers, where it is often overexpressed at the tumor cell surface, acts as an anti-apoptotic factor inducing cell survival, is upregulated upon endoplasmic reticular or genotoxic stress and has been shown to confer resistance to multiple chemotherapeutic agents.^{538,541,543,544,639} However, there is no apparent overexpression of *CLPTM1L* in UM tumors, and gain of chromosome 5p is not a frequent cytogenetic event in UM; our preliminary studies of *CLPTM1L* siRNA-mediated silencing and overexpression also do not indicate a role for *CLPTM1L* in cell survival or proliferation so far. Additional studies are required to validate or invalidate both *CLPTM1L* and *TERT* gene candidates in UM to unravel the biological mechanisms linked to rs452384 and its associated risk region in UM. Long-range chromatin interaction studies, including chromosome conformation capture assays (3C) and

their higher order counterparts 4C and 5C derived techniques, will help validate the regulatory effect of rs452384 genomic region on target genes. Hypothesis-driven 3C would indeed allow us to confirm the interaction between rs452384 and its environment to the promoter elements of *CLPTM1L* and *TERT*, while 4C results could potentially point towards additional cis-regulated genes, further away from the *TERT/CLPTM1L* locus, also impacted by the rs452384 genetic environment.

CONCLUSIONS AND PERSPECTIVES

Taken together, our work contributed to the identification and characterization of moderate and low penetrance genetic risk factors in UM, extending our knowledge on genetic susceptibility to this rare disease and generating additional points of discussion and research to further refine UM tumorigenesis. The main messages of this work are the implication of *MBD4* in UM predisposition with moderate penetrance, and the characterization of low penetrance alleles in UM. By conducting our second and largest GWAS in UM, we confirmed *TERT/CLPTM1L* susceptibility region in UM and identified two additional risk loci, pigmentation markers *HERC2/OCA2* on chromosome 15 and *IRF4* on chromosome 6. We further could show by stratifying UM patients based on their chromosome 3 status that *HERC2/OCA2* risk alleles are specifically associated with high-metastatic risk (monosomy 3) UM subgroup while *IRF4* predisposes to low-metastatic risk (disomy 3) UM. We further characterized risk alleles at the *TERT/CLPTM1L* locus on chromosome 5p15.33 and highlighted the role of rs452384 as a functional variant mediating some of the risk conferred by this genomic locus in UM through allele-specific gene regulation and binding to NKX2.4 transcription factor.

We evidenced the role of *MBD4* pathogenic germline variants in UM predisposition, specifically associated with the high-metastatic risk UM subgroup and resulting in a dramatic change of UM tumor characteristics, with a CpG>TpG hypermutator phenotype radically different from the usually very low tumor mutation burden in UM. We demonstrated that this gene is associated with an almost 10-fold relative risk of developing UM compared to *MBD4*-wild-type carriers and estimated the frequency of germline *MBD4* mutations in UM based on an unselected, large consecutive series of over 1,000 patients at 0.7%, slightly lower than that of *BAP1* predisposition in UM. An important application of this work, in continuation with previous work carried out by the team,⁵⁵³ is the integration of *MBD4* in a clinical genetic setting to better recognize patients affected by *MBD4* predisposition, as the tumor spectrum affected by this predisposition syndrome is likely to expand. More systematic analyses of *MBD4* will allow to better define the clinical histories of germline *MBD4* mutant carriers as so far, the low number of UM patients identified may explain the lack of familial cancer history

observed and prevents robust studies on age of onset and survival analyses in *MBD4* deficiency context, although preliminary results indicate a lack of early UM cancer onset in *MBD4* deficiency context (unlike in acute myeloid leukemia patients affected by germline biallelic inactivation of *MBD4*).⁶²³ In addition, the identification of identical (hotspot) *MBD4* germline mutations in multiple unrelated patients, either within our UM cohorts or in patients affected by other cancers such as acute myeloid leukemia (AML)⁶²³ and glioblastoma,⁵⁵³ suggest that these mutations may have a founder origin and may shed light on some common, pan-cancer underlying mechanisms. We are currently collaborating with Dr Ahmed Idbaih's team to assess *MBD4* predisposition in gliomas in a large genetic screen, which could identify additional mutation carriers and refine the clinical phenotype of germline *MBD4* mutant carriers. As multiple unrelated UM patients also harbor the same *MBD4* mutation, we are also planning to evaluate by Identity by State (IBS) analysis common genetic segments between these patients that share the same mutation, to demonstrate their founder origin.

Another important application of this work is that patients affected by *MBD4* deficiencies may be offered additional treatment options due to their high mutation burden and CpG>TpG phenotype, including immunotherapy such as PD1 inhibitors. In this respect, *MBD4* deficiency in tumor could represent a biomarker for immunotherapy. Since there is an absence of specific *MBD4* antibody to use in clinical setting, we are developing a collaboration with Sandrine Moutel's team (Institut Curie therapeutic and recombinant antibodies platform) to screen and produce nanobodies targeting *MBD4*, using the human recombinant protein produced in our work. The fact that the first UM patient identified with an *MBD4* germline mutation initially responded to anti-PD1 treatment provides grounds to analyze anti-PD1 responders within UM patients treated at Institut Curie and look at their *MBD4* status. So far, the anti-PD1 response rate is estimated at 60% for UM patients with *MBD4* deficiency (*MBD4*def), and 5% without (Saint Ghislain & Rodrigues, manuscript in preparation).⁴⁶⁵ Retrospective analyses of large cohorts will be required to fully evaluate the role of *MBD4* in response to anti-PD1.

While these clinical applications are of utmost importance, much has yet to be unraveled to understand the biological mechanisms underlying *MBD4* deficiency in oncogenic processes. Common mutations found in an *MBD4*def context, such as *BAP1* inactivation in all UM *MBD4*-carriers, *ATRX* mutations in glioblastoma,⁶²⁷ *DNMT3A* mutations in AML patients⁶²³ suggest a

common role for *MBD4* in cancer progression, accelerating tumorigenic processes and potentially regulating major epigenetic processes, consistent with its critical role in maintenance of DNA methylation patterns. Sometimes referred to as the “guardian of the epigenetic galaxy”,⁶⁴⁰ *MBD4* may play critical roles in maintaining 5-meCpG sites at promoter sequences within noncoding regions, and its inactivation may result in altered DNA methylation landscapes, aberrant demethylation of promoters or changes in noncoding RNAs and/or repetitive elements. We are currently investigating at the transcriptional and epigenetic levels such potential consequences of *MBD4* deficiency in UM patients.

In UM, *MBD4*def has been shown to act as a genetic clock,⁴¹⁷ characterized by a mutational signature reminiscent of SBS1 ageing signature, which may allow to track clonal evolution events and unravel new aspects of tumorigenesis, especially late metastatic events. To study *MBD4*-specific mutational processes, we have generated by CRISPR *MBD4* knockout (*MBD4*-ko) isogenic cell line models and we are further planning to develop an inducible degradable *MBD4* cell line. An important application to such cell line models is that they will allow us to investigate biological pathways specific to *MBD4*def context, which could be targeted in new strategies for cancer therapies. We have initiated a synthetic lethality drug screening on our *MBD4*-ko models, exploiting the defect in me-CpG>TpG DNA repair resulting from *MBD4* inactivation, and are currently testing various drugs in our HAP1 and Mel202 models. Although these tests will require further optimization, *MBD4*-ko HAP1 model shows promising sensitivity to some drugs compared to its wild-type counterpart. In addition, we have initiated a collaboration to perform a high-throughput drug screening using *MBD4*-ko models, which could further identify treatment options in this context; this may be of particular importance as UM patients are expected to present secondary resistances to anti-PD1 therapy, as a consequence of the continuous mutagenic process resulting from *MBD4* inactivation.

Finally, while in our work we have functionally characterized by *in vitro* glycosylase assay *MBD4* mutations located within the glycosylase domain of the protein, (i) we have not investigated the kinetics underlying these mutations, which may not be equivalent; and (ii) we have not looked at mutations located elsewhere on the protein, identified within some samples from our cohort of UM patients. Additional works testing these mutants in biochemical assays will be required to functionally assess their effects and may shed light on

MBD4 mechanisms of action. The team is currently developing an improved fluorescence-based glycosylase activity assay, which will allow to further study the kinetics related to MBD4 activity, and the importance of its methyl-binding domain for enzymatic activity.

The second, and main, objective of this work was to characterize common genetic variation in UM susceptibility, to explain part of the heritability that is not accounted for by other known – and rare – predisposing genes in UM, such as well-characterized highly penetrant *BAP1* germline mutations, or infrequent mutations of *MLH1*, *PALB2*, *MSH6*, *CDKN2A* and few others that require independent validation in epidemiologic studies. Through a large GWAS (given the rarity of the disease) in UM and a functional genomics study, we confirmed the association of three main genomic risk loci in UM, on chromosome 5 (*TERT/CLPTM1L*), chromosome 6 (*IRF4*, rs12203592 single SNP) and chromosome 15 (*HERC2/OCA2*). Within the multi-cancer 5p15.33 *TERT/CLPTM1L* region harboring more than 10 independent association loci, the single UM association locus identified shares some common alleles with melanoma, pancreatic and lung cancers.⁵²⁹ Strikingly, rs452384 was also identified as a functional SNP in lung cancer,⁶³⁷ although the 5p locus in this tumor type has an opposite directionality of effect than in pancreatic cancer and melanoma (and UM). Importantly, our combined and imputed GWAS allowed us to identify additional SNPs within the 5p peak initially marked by rs421284 in UM in the first GWAS carried out by the team,⁵²⁶ including multiple SNPs that are also highly associated cutaneous melanoma (CM) as evidenced in the most recent GWAS study and multiple previous ones,^{171,535} including rs401681 associated with multiple cancers, rs465498 which was shown to associate with telomere length, and rs2447853 which was previously not associated with UM but is a major CM risk SNP, also strongly implicated in nevus development.^{171,529}

These results strongly support important shared etiological factors between UM and CM more than any other cancer type, not only at this locus but also corroborated by risk alleles of pigmentation markers *IRF4* and *HERC2/OCA2*, which play a role in both these cancers. Yet, this generates additional questions regarding the genetics and genetic susceptibility in UM and CM, given the radical differences in genetic and genomic driver events in these two cancers, in their clinical aspects, and to some extent their highly penetrant predisposing genes (*BAP1* much more frequent in UM than CM, *CDKN2A* and *CDK4* essentially exclusive to CM).

Similarly, and despite the common epidemiology of UM and CM characterized by a prevalence of individuals with fair skin and light-colored irises, while the role ultra-violet exposure (UVR) is demonstrated in CM risk, where darker pigmentation protects against UVR, it is much less obvious and UM (excepted for iris melanoma), raising additional questions regarding the role of their shared susceptibility loci on *IRF4* and *HERC2/OCA2*. A plausible explanation that was suggested is the fact that, despite their shared cell of origin – melanocytes – and common neural crest embryonic origin, minor location differences within the neural crest anterior-posterior positioning (nonepithelial melanocytes, such as ocular melanocytes, predominantly from the cranial neural crest, while epithelial melanocytes mostly arise from the trunk),^{641,642} and differences in cell environment (skin melanocytes interacting with epithelial cells, in contrast to ocular melanocytes with mesodermal stroma), may result in tissue-specific cell-cell contacts, proliferation, and distinct selection of mutational driver events depending on the environment, resulting in two clearly distinct diseases.^{226,417}

Another plausible explanation is that *HERC2/OCA2* and/or *IRF4* risk alleles may play additional roles outside from pigmentation (and potentially shared between UM and CM tumorigenesis), which we suggest in UM in our GWAS work (Article 2), especially for *IRF4*, by conditioning on eye color, although a clear limitation to our study is the circularity of the analysis due to the fact that eye color is predicted from known pigmentation SNPs, including UM risk alleles rs12203592 (*IRF4*) and rs12913832 (*HERC2*), making it challenging to derive causal statements. To overcome this limitation, we are planning to validate our results in an independent cohort, involving individuals for which eye color phenotype is known (this information is accessible for UM patients treated at Institut Curie, and we are currently aiming to have access to the Rotterdam Study as a control population). Regardless, multiple factors support for a role for *IRF4* rs12203592 SNP outside pigmentation. In addition to its implication, via upregulation of *IRF4* expression levels with its own C (UM protective) allele and along with MITF, in the synthesis of tyrosinase enzyme critical to melanogenesis,^{631,643} *IRF4* is also known for its immune functions;^{644,645} the same SNP has also been identified in an eQTL study on primary melanocytes to be a significant *trans*-eQTL for four genes, *PLA1A*, *NEO1*, *TMEM140* and *MIR3681HG*, whose expressions were at least in part linked to *IRF4* expression levels and the transcription factor it encodes,¹³⁴ making them potential direct

targets downstream IRF4. Notably, NEO1 has been associated with cell proliferation and migration favoring cancer,⁶⁴⁶ and *TMEM140* is an interferon-stimulated gene.⁶⁴⁷ Functional assays of these target genes could shed light on mechanisms downstream of UM and CM *IRF4* susceptibility locus.

To an even higher extent than *IRF4*, *HERC2/OCA2* is a major determinant of eye color, where *HERC2* rs12913832 alleles play a key role in blue or brown eye phenotypes. UM risk allele of rs12913832 prevents the formation of an activation loop with the *OCA2* promoter, resulting in decreased *OCA2* synthesis and consequently low synthesis of eumelanin darker pigment.²⁴⁰ Not only does this explain UM epidemiology with a prevalence of individuals with light eye color, low *OCA2* levels are also thought to result in accumulation of DHICA intermediate due to lower amounts of tyrosinase being transferred from the endoplasmic reticulum to melanosomes, resulting in the generation of reactive oxygen species (ROS) (Figure 20).²²³ This pigmentation locus could then also mediate risk via some UVR-independent functions. In addition, we show that *IRF4* and *HERC2/OCA2* UM susceptibility loci are respectively associated with low-metastatic risk, disomy 3 (D3) UM, and high-metastatic risk, monosomy 3 (M3) UM. In this respect, the higher amount of ROS generated with *HERC2/OCA2* risk alleles (and/or lower protective mechanisms from ROS) could participate in the distinction between M3 and D3 UM. This hypothesis is consistent with results from another UM GWAS study⁶⁴⁸ – although not confirmed by our study – that identified DNA repair enzyme TDP1 as a risk locus preferentially associated with poor-prognosis UM (epithelioid cells, chromosome 3 loss, 8q gain), relevant in this context as TDP1 is mainly known for its repair of oxidative DNA lesions, which further corroborates a hypothetic role of susceptibility to oxidative damage in UM and particularly poor prognosis UM.

The identification of distinct genetic risk factors for high-risk M3 and low-risk D3 UM is perhaps the most exciting result in our UM GWAS, and adds to the list of differences between M3 and D3 UMs on many aspects, including genetic, genomic, transcriptomic, epigenetic and clinical levels.^{260,261,297,409} We are thus planning to replicate and extend these findings in a larger cohort of UM patients with known chromosome 3 status that were are currently accruing, and to perform a M3 vs. D3 GWAS in UM, which could further refine the distinct etiological factors between the two UM subtypes and possibly point towards distinct

biological mechanisms in UM tumor progression, conferring differential risks of metastatic evolution.

Unlike pigmentation loci, the *CLPTM1L/TERT* locus on chromosome 5p15.33 is equally associated with M3 and D3 UM subgroups, indicating a shared oncogenic mechanism conferred by this locus in UM. Due to the multiple independent loci and various directionality of association of this locus across cancer types, a tissue- and cell-specific regulation is expected, making it challenging to address the functional consequences of this complex locus, highlighting the need for cancer-specific studies. Through functional characterization of a sub-region at 5p15.33 plausible to be functionally relevant due to the presence of active chromatin marks as shown in ENCODE database (including H3K27ac, DNase I clusters, TF binding), we have identified a functional variant, rs452384, part of the association signal marked by rs421284 and rs370348. Although we cannot rule out the existence of additional functional variants at this genomic locus in UM, the strong allele-specific regulation at the specific genomic position surrounding rs452384, added to the absence of additional risk signal independent from rs421284/rs452384 peak as evidenced by conditional analyses support an important role for rs452384 in gene regulation and in UM risk. However, our second GWAS identified additional 5p15.33 risk alleles, within a genomic region directly downstream rs452384 (3–5kb away) also marked by H3K27ac signal, which we could not assess by classical cloning and luciferase strategies due to the highly repetitive nature of this genomic minisatellite segment. We therefore recently sequenced this 5kb fragment in 96 UM patients by PacBio long-read sequencing, for which genotyping information of tested UM SNPs outside the minisatellite region is known, and we are currently performing the bioinformatics analysis of this region to determine whether additional risk SNPs lie within this region, and whether these are independently associated with UM or part of the same risk signal.

Additional work remains to be done to fully determine the function of rs452384 in UM tumorigenesis. Our data strongly supports a role for NKX2.4 transcription factor preferential binding to rs452384-T, UM protective allele. In this respect, one would expect NKX2.4 to either upregulate genes with tumor suppressor activity or to downregulate genes with oncogenic properties, such as cell survival or proliferation. The decrease in *TERT* and *CLPTM1L* expression upon *NKX2.4* knockdown in UM rs452384-TT cell line argues for a role of NKX2.4

as a transcriptional repressor, and for *TERT* and *CLPTM1L* as the primary target genes downstream the 5p15.33 risk locus, as it is the case in multiple other cancers associated with this locus (albeit through other functional SNPs), including lung, pancreatic and ovarian cancers.^{183,541,544,546,636} We are planning a functional study of *CLPTM1L* in UM cell survival, proliferation and resistance to apoptosis upon ER-mediated stress, as there is a subtle yet significant eQTL for *CLPTM1L* with rs452384 and rs421284 in UM tumor samples. *TERT* and its regulatory effects on telomere length are also a major candidate for mediating UM risk at this locus, however we face the difficulty that this gene is essentially not expressed in UM tumors, making it challenging to assess for genotypic correlation in eQTL studies, in line with the dramatic decrease in telomere length in UM tumor compared to PBL samples, supporting previous findings in pan-cancer studies of telomere length that found UM to be an outlier tumor with the shortest telomeres.⁶³⁸ Hypothesizing that rs452384 may exert effects on telomeres in UM predisposition prior to tumorigenesis, supported by the telomere QTL (tQTL) were observed between rs452384 and telomere length in PBL samples of UM patients and GWAS controls, we are currently developing a protocol to extract tumor uveal melanocytes (from enucleation samples of UM patients) and surrounding healthy tissue (matching primary uveal melanocytes), which would allow us to compare telomere length in both contexts. Generally speaking, *TERT* and *CLPTM1L* functional studies in non-transformed, primary uveal melanocyte cell lines could reveal biological mechanisms underlying the 5p15.33 normal biology to explain risk at this region, and may prove more successful in the identification of potential oncogenic pathways.

A TWAS in cutaneous melanoma (CM), imputing eQTL data from primary melanocytes in CM GWAS summary statistics, revealed four novel susceptibility loci, highlighting the need to select the appropriate tissue and cell type.¹³⁴ In a similar way to the eQTL available in the GTEx database on rs452384, which shows numerous inverse directionalities of effect on both *TERT* and *CLPTM1L* according to tissue type (and, strikingly, inverse association between UM and skin tissues), Zhang and colleagues observe major differences in eQTL directionalities from skin tissues, which contain multiple cell types, and primary melanocytes, cells of origin for CM.¹³⁴ This highlights the complexity of identifying potential downstream target genes based on tissue expression data, especially in UM.

In addition to *CLPMT1L* and *TERT*, the 5p15.33 locus could also regulate other target genes. Performing a TWAS with our GWAS data and eQTL information of primary melanocytes could potentially identify relevant transcriptional targets downstream rs452384, and other UM risk loci. In addition, transcriptional analyses by RNA-seq of UM cell lines in NKX2.4 wild-type and knock-out contexts would allow to identify genes differentially expressed and linked to NKX2.4 regulatory activity. In the same way, we are planning SNP editing via CRISPR, to obtain isogenic cell line for rs452384 risk and protective alleles, which will provide a better model to functionally test the effect of rs452384 and its downstream target genes and pathways. Finally, chromosome conformation capture (3C) and derived techniques such as 4C and 5C would not only allow to confirm the interaction of rs452384 with the promoter of *CLPMT1L* and/or *TERT*, providing additional grounds for these genes in mediating UM risk at the 5p locus, but could also identify additional *cis*-regulated targets in an unbiased approach.

Altogether, our action plan following this work is to continue to characterize the genetic risk factors identified in UM, to refine UM tumorigenesis and ultimately to provide strategies for earlier diagnosis of UM and/or potential targets for therapy.

RÉSUMÉ DE LA THÈSE VERSION FRANÇAISE

Introduction et objectifs

Génétique du mélanome uvéal

Le mélanome uvéal (MU) est un cancer rare, mais reste la tumeur intraoculaire la plus fréquente chez l'adulte avec ~500 à 600 nouveaux cas par an en France. Le MU est lié à la transformation maligne des mélanocytes de l'uvée, composée de la choroïde (90% des cas), du corps ciliaire (6%) et de l'iris (4%). Sur le plan génétique, le MU a une très faible charge mutationnelle somatique, et comporte deux événements oncogéniques majeurs impliqués dans la transformation maligne: l'activation constitutive de la voie de signalisation Gαq à travers des mutations activatrices mutuellement exclusives de *GNAQ* ou *GNA11*^{268,269} (ou, dans de rares cas, *PLCB4*²⁵⁸ ou *CYSLTR2*²⁷⁸), et une seconde étape dans la transformation maligne, caractérisée elle aussi par des mutations majoritairement mutuellement exclusives impactant le risque métastatique : les mutations de *BAP1* (associées à une inactivation bi-allélique du gène par perte de l'allèle sauvage par monosomie 3, M3) constituent le MU à haut risque métastatique et mauvais pronostic,^{260,297,302} tandis que les mutations de *SF3B1*^{259,301} ou *EIF1AX*³⁰⁰ (associées à la disomie 3, D3) confèrent, respectivement, un risque intermédiaire (et survenant plus tard)³³⁸ et faible de métastases. Le très mauvais pronostic du MU métastatique, hépatique dans la majorité des cas, se caractérise par une survie médiane de moins de 13,5 mois.⁴⁷³ Alors que la tumeur primaire est majoritairement traitée avec succès par irradiation ou énucléation, la progression métastatique du MU qui survient dans 30 à 50% des cas⁴⁷⁶ est effectivement réfractaire au traitement et, jusqu'à très récemment, aucune thérapie n'a prouvé son efficacité dans la prolongation de la survie.

Epidémiologie du MU

Le MU est associé aux individus à la peau et aux yeux clairs,⁴⁸³ comme le mélanome cutané. Par ailleurs, les populations de souche européenne ont un risque relatif (RR) 10 à 20 fois plus élevé de développer un MU que les individus d'origines africaine ou asiatique,^{476,481} et il existe également un gradient croissant Sud-Nord en Europe avec une incidence de <2 cas par million

par an dans les pays d'Europe du Sud (Italie, Espagne), et jusqu'à 4 fois plus dans les pays scandinaves (Norvège, Danemark, Irlande).^{223,479,480} Cependant, (i) l'absence de signature UV dans les tumeurs de MU (à l'exception notable du mélanome de l'iris),²⁵⁹⁻²⁶¹ (ii) leur faible taux mutationnel,²⁵⁹ et (iii) l'incidence stable du MU ces dernières décennies malgré l'augmentation de l'incidence du mélanome cutané liée à l'augmentation de l'exposition aux rayonnements ultra-violet (UV), excluent le rôle des différences de pigmentation dans la protection contre les UV pour expliquer cette épidémiologie remarquable.

Notre hypothèse pour expliquer la prévalence du MU dans les populations européennes est l'existence de facteurs de prédisposition génétique au MU présents dans ces populations, ou inversement la présence d'allèles protecteurs dans les populations asiatique et africaine. Par ailleurs, les rares formes familiales du MU (survenant chez au moins deux membres d'une même famille) surviennent dans environ 1% des cas,^{503,506} or, les mutations germinales du gène suppresseur de tumeur *BAP1*, seul facteur de prédisposition génétique à forte pénétrance connu dans le MU,^{307,310} n'expliquent qu'une partie des cas familiaux, suggérant l'existence d'autres facteurs de susceptibilité.

Etude d'association pan-génomique (GWAS) dans le MU

Afin d'identifier des variants génétiques de risque du MU dans la population européenne, l'équipe a réalisé en 2017 une étude d'association pan-génomique (*genome-wide association study*, GWAS) incluant 259 patients de MU et 401 contrôles sains d'origine européenne, permettant d'identifier une première région de susceptibilité au MU sur le chromosome 5p15.33, au locus *TERT/CLPTM1L* (*telomerase reverse transcriptase* et *cleft lip and palate trans-membrane 1 like*), avec deux polymorphismes nucléotidiques (*single-nucleotide polymorphism*, SNP) atteignant le seuil de signification statistique, rs42124 et rs452932.⁵²⁶ Ces deux SNPs sont génétiquement liés par fort déséquilibre de liaison (LD), entre eux et avec d'autres SNPs (appartenant au même haploblock) situés dans des régions introniques à *CLPTM1L*, ouvrant un champ de recherche vers une étude de caractérisation fonctionnelle de la région génomique de risque du MU au 5p15.33. Cependant, la fréquence comparable de l'allèle mineur (MAF) de ces SNPs dans les populations africaine, asiatique et européenne⁵²⁶ suggère l'existence d'autres loci de prédisposition pour expliquer la prévalence européenne de la maladie.

MBD4

En parallèle de ce GWAS, notre équipe de recherche a identifié en 2018 une mutation constitutionnelle délétère du gène *MBD4* (*methyl-CpG binding domain 4*), localisé sur le chromosome 3, chez une patiente de MU dont la tumeur a répondu de manière exceptionnelle à un traitement anti-PD1 (point de contrôle immunitaire) d'immunothérapie, caractérisée par un phénotype hypermuté majoritairement constitué de transitions de type CpG>TpG.⁵⁵³ Cette mutation germinale était accompagnée d'une perte de l'allèle sauvage dans la tumeur par monosomie 3, amenant à l'inactivation bi-allélique de ce gène et au phénotype hypermutateur. *MBD4* code pour une glycosylase impliquée dans la première étape du mécanisme de réparation par excision de bases (*base excision repair, BER*). Cette enzyme reconnaît notamment, grâce à son *methyl-binding domain*, les lésions de type CpG>TpG provenant de la désamination spontanée des 5-méthylcytosines (5-mC), et son domaine C-terminal lui confère son activité glycosylase permettant de cliver la base altérée afin d'initier le processus de réparation de l'ADN. La déficience de *MBD4* engendre ainsi l'absence de réparation des lésions CpG>TpG, résultant en un taux élevé de mutations ayant ce profil.

La mutation germinale de cette patiente était la seule altération délétère de *MBD4* parmi une série de 42 patients de MU métastatique traités à l'Institut Curie,⁵⁵³ de telles mutations étant par ailleurs très rares dans la population générale. En investiguant les données publiques de tumeurs parmi la cohorte du *Cancer Genome Atlas (TCGA)*, l'équipe a pu identifier deux autres mutations germinales délétères de *MBD4*, dans une tumeur de MU et dans un glioblastome, elles aussi associées avec une inactivation bi-allélique du gène et un phénotype hypermuté de type CpG>TpG.⁵⁵³ Alors que ces résultats suggéraient une implication potentielle des mutations germinales délétères de *MBD4* dans la prédisposition au cancer, 3 autres mutations constitutionnelles bi-alléliques de ce gène ont été reportées dans des cas de leucémie aiguë myéloblastique,⁶²³ ainsi que des cas ponctuels dans d'autres types de tumeurs. Ces découvertes nous ont poussé à investiguer le rôle potentiel de *MBD4* dans la prédisposition au MU, en déterminant la fréquence de mutations germinales dans une large cohorte de MU et la comparant à leur fréquence dans la population générale.

Objectifs de la thèse

Les objectifs de cette thèse étaient d'identifier et de caractériser fonctionnellement les facteurs de prédisposition génétique du MU. Deux approches ont été menées en parallèle afin d'investiguer la prédisposition au MU: (i) une stratégie gène candidat, étudiant le rôle potentiel de *MBD4* dans la prédisposition au MU à travers un criblage génétique dans une large cohorte consécutive de patients du MU; et (ii) une étude de GWAS, visant à l'identification de polymorphismes fréquents associés au risque du MU avec une faible pénétrance, et à leur caractérisation biologique par une étude de génomique fonctionnelle, dite « *post-GWAS* », afin d'explorer le rôle des allèles de susceptibilité du MU.

Résultats

Article 1: Germline *MBD4* mutations and predisposition to uveal melanoma

Cet article, publié dans le *Journal of National Cancer Institute*, présente nos travaux démontrant que les mutations constitutionnelles délétères de *MBD4* prédisposent au MU, avec un risque relatif (RR) de 9.15 (9.15 fois plus de chances de développer un MU chez un sujet comportant une mutation *MBD4* par rapport à un individu n'en ayant pas).⁶⁴⁹

Comme mentionné précédemment, la découverte de mutations constitutionnelles de *MBD4* dans des tumeurs de MU, glioblastome et de leucémie aiguë myéloblastique suggéraient un rôle de ce gène dans la prédisposition au cancer, et nous ont poussé à investiguer la fréquence de mutations germinales délétères de *MBD4* dans le MU. Pour ce faire, nous avons accru une série consécutive de plus de 1000 patients de MU, traités à l'Institut Curie de 2013 à 2018, lesquels avaient préalablement consenti à la recherche de facteurs de prédisposition génétique dans leur ADN constitutionnel (issu du sang) dans le cadre de notre étude de GWAS mentionnée par la suite. De même, nous avons accru une deuxième cohorte d'ADNs tumoraux de 192 patients de MU dont la tumeur comportait une monosomie du chromosome 3 (M3), le gène *MBD4* étant localisé sur ce-dernier; les patients de MU comportant une M3 dans leur tumeur sont associés au plus haut risque métastatique. A travers un criblage génétique de *MBD4* par séquençage des exons (*targeted sequencing*) dans ces deux cohortes et à travers une étude fonctionnelle (test d'activité glycosylase *in vitro*) des mutations au caractère pathogène incertain (*variants of unknown significance, VUS*), nous avons identifié 8 mutations pathogènes de ce gène dans la série consécutive d'ADNs constitutionnels,

correspondant à une incidence 9.15 fois plus élevée de mutations délétères que dans la population générale (*Genome aggregation database*, gnomAD v2.1), et 4 mutations délétères supplémentaires dans la cohorte tumorale M3. L'analyse par *Whole exome sequencing* (WES) des tumeurs de ces patients nous ont permis de démontrer le caractère hypermuté de ces tumeurs par rapport à une précédente cohorte tumorale de MU comportant la version *wild-type* de *MBD4*,⁴¹⁷ et la large prévalence des mutations de type CpG>TpG dans le contexte *MBD4*-déficient. De plus, la totalité des patients porteurs de mutation délétères présentait une monosomie 3 dans leur tumeur, indiquant que les mutations pathogènes de *MBD4* prédisposent spécifiquement au MU à haut risque métastatique, probablement en partie expliqué par la présence de *MBD4* sur le chromosome 3, fréquemment perdu lors de la progression tumorale du MU.

Ces résultats nous ont ainsi permis de démontrer que *MBD4* est un gène de prédisposition au MU, et que les mutations germinales délétères de ce gène prédisposent spécifiquement au MU à haut risque métastatique caractérisé par une monosomie 3, l'inactivation bi- allélique de ce gène dans la tumeur induisant un phénotype hypermuté et une prévalence du contexte mutationnel de type CpG>TpG. Suite à ces résultats, *MBD4* est maintenant inclus dans les panels oncogénétiques à l'Institut Curie, l'identification de patients porteurs de mutations *MBD4* étant particulièrement importante étant donné que les tumeurs *MBD4*-déficientes pourraient répondre à l'immunothérapie, contrairement au MU de manière générale.

Article 2: Different pigmentation risk loci for high-risk monosomy 3 and low-risk disomy 3 uveal melanomas

L'objectif premier de cette thèse était d'identifier et de caractériser les facteurs de prédisposition génétique au MU, à travers une étude de GWAS permettant d'identifier des polymorphismes (SNPs) communs de risque, ou allèles de susceptibilité, à faible pénétrance. Le MU étant associé avec les individus de souche européenne et avec ceux aux yeux et à la peau clairs, nous avons émis l'hypothèse de la présence d'allèles à risque dans ces populations. Comme mentionné précédemment, le premier GWAS mené par l'équipe a identifié un locus de risque au chromosome 5p15.33, *TERT/CLPTM1L*, dont le rôle dans la tumorigenèse du MU demeurait inconnu.⁵²⁶ Cependant, ces allèles de risque ne pouvaient expliquer la prévalence européenne du MU étant donné leur fréquence semblable dans les

populations européenne, asiatique et africaine. Nous avons ainsi initié une seconde phase de GWAS, combinant les données du 1^{er} GWAS avec deux séries supplémentaires d'ADNs constitutionnels de patients du MU traités à l'Institut Curie de 2013 à 2018, permettant d'augmenter la puissance statistique du GWAS (comprenant désormais les données de 1,142 patients de MU et 882 contrôles) et donc potentiellement d'identifier d'autres régions de risque du MU.

Dans un second article,⁶⁵⁰ publié dans le *Journal of the National Cancer Institute*, nous décrivons les résultats de ce GWAS, dans lequel les données imputées des trois jeux de données sont par la suite fusionnées. Ces résultats nous ont permis de confirmer l'association du locus *TERT/CLPTM1L* avec le MU, et d'identifier deux régions de susceptibilité supplémentaires: le locus *IRF4* (chromosome 6) et le locus *HERC2/OCA2* (chromosome 15), tous deux impliqués dans les mécanismes de pigmentation des yeux. Par ailleurs, en stratifiant les patients selon le statut génomique connu du chromosome 3 dans leurs tumeurs, pronostique du risque métastatique (disomie 3, D3, associé au faible risque métastatique, vs. monosomie 3, M3, risque élevé de métastases), nous avons pu montrer que le SNP rs12203592 d'*IRF4* est exclusivement associé aux MU avec D3, et inversement pour le SNP du locus *HERC2/OCA2* rs12913832, prédisposant spécifiquement au MU à haut risque métastatique (M3). Contrairement au locus *TERT/CLPTM1L*, associé tant aux MU-D3 qu'à ceux présentant une M3, les loci de pigmentation identifiés dans ces travaux sont différenciellement associés au MU avec D3 ou M3, représentant deux régions de risque distinctes qui influent la biologie des tumeurs de MU et leur potentiel métastatique, suggérant deux étiologies distinctes du risque du MU. Un impact clinique potentiel important de ces résultats est que les génotypes de ces loci pourraient indiquer le risque métastatique du MU.

Nos résultats confirment par ailleurs les travaux de Ferguson et collaborateurs,⁵⁵¹ qui identifient des régions de susceptibilité communes au MU et au mélanome cutané, dont les SNPs des loci *IRF4* et *HERC2* impliqués dans la pigmentation des yeux et de la peau. Si ces marqueurs de pigmentation sont concordants avec l'épidémiologie du MU, dans laquelle la peau et les yeux clairs sont des facteurs importants de risque (comme pour le mélanome cutané), le rôle de la synthèse de la mélanine dans la protection contre les UV n'est pas aussi évident dans le MU que dans le mélanome cutané pour lequel ce mécanisme protecteur contre les radiations liées aux UV est fortement documenté. De futures études seront

nécessaires pour élucider le mécanisme de ces allèles de susceptibilité au MU, qui pourraient impliquer les dérivés réactifs de l'oxygène (DRO, ou *reactive oxygen species*) de manière UV-indépendante. La caractérisation fonctionnelle du locus *TERT/CLPTM1L* dans le MU, quant à elle, est étudiée dans l'Article 3 ci-dessous.

Article 3: Functional characterization of CLPTM1L/TERT 5p15.33 risk locus in uveal melanoma identifies rs452384 as a functional variant regulating CLPTM1L and TERT expression through NKX2.4 allele-specific binding

Le premier GWAS mené par l'équipe en 2017 a identifié un locus de risque sur le chromosome 5p15.33, *TERT/CLPTM1L*, validé dans une cohorte indépendante, avec deux SNPs franchissant le seuil de signification statistique: rs421284 et rs452932 (odds ratio OR combiné dans la série découverte et validation: OR=1.71, $p=5 \times 10^{-9}$ et OR=1.72, $p=2 \times 10^{-9}$, respectivement). Les résultats décrits dans l'Article 2 confirment l'association de ce locus avec le MU (avec rs421284 et rs370348 à la tête de ce signal) caractérisé par de nombreux SNPs liés entre eux par déséquilibre de liaison (LD) situés dans les régions introniques de *CLPTM1L*. Une analyse de corrélation génotype-expression (eQTL) a par ailleurs démontré l'expression plus importante de *CLPTM1L* avec l'allèle à risque de rs421284.⁵²⁶

La région *TERT/CLPTM1L* est un locus de risque complexe impliqué dans de nombreux cancers, caractérisé par >10 signaux d'association indépendants les uns par rapport aux autres, et une direction d'effet dépendant du type tumoral, suggérant une régulation tissu- et cancer-spécifique.^{183,527-529} Un nombre limité d'études ont cherché à caractériser fonctionnellement les signaux de risque au 5p15.33, dans des cancers autres que le MU.^{183,537,541,636} Ces travaux, rédigés sous la forme d'un article qui sera prochainement soumis pour publication dans une revue scientifique, visent à la caractérisation par étude de génomique fonctionnelle (« post-GWAS ») des mécanismes biologiques sous-jacents au locus *TERT/CLPTM1L* dans le MU, afin de comprendre leur rôle dans le processus tumoral.

A travers une étude des marques chromatinienne et une analyse de données publiques (ENCODE), nous avons pu identifier une région intronique à *CLPTM1L*, comprenant des allèles à risque du MU, enrichie en marque H3K27ac (région à activité *enhancer*) et en sites hypersensibles à la DNase I (marqueur de chromatine ouverte, accessible à la fixation de facteurs et associé à des régions transcriptionnellement actives). Nous avons confirmé la

présence de la marque H3K27ac à cette même région dans des modèles cellulaires de MU. A travers un clonage allèle-spécifique de la région intronique associée au risque du MU dans un vecteur rapporteur de luciférase, testant l'induction l'activité transcriptionnelle, nous avons démontré que l'haplotype de risque du MU activait davantage la transcription génique par rapport à l'haplotype protecteur, et avons identifié rs452384 comme un SNP fonctionnel dans la région génomique de risque, induisant la plus forte allèle-spécificité dans la transcription génique. Par tests de retard sur gel (EMSA), nous avons ensuite montré la fixation différentielle de facteurs nucléaires à l'allèle à risque (C) ou protecteur (T) de rs452384, confirmée par une analyse de spectrométrie de masse quantitative qui a permis d'identifier NKX2.4 et GATA-4 comme facteurs de transcription se fixant préférentiellement sur un motif d'ADN comportant rs452384-T et -C, respectivement. Enfin, par une étude de sous-expression utilisant des siRNA, nous avons montré que la sous-expression de *NKX2.4* (mais pas *GATA-4*) résulte en une légère surexpression des gènes *CLPTM1L* et *TERT*. Enfin, nous émettons l'hypothèse que rs452384 a un rôle dans la tumorigenèse du MU à travers la régulation de la longueur des télomères, étant donné qu'une étude de la longueur des télomères dans l'ADN germinale (issu des lymphocytes du sang périphériques) d'une série de patients de MU nous a permis de montrer que le génotype à risque de rs452384 (CC) est associé avec des télomères plus longs que les génotypes CT et TT. Ces travaux identifient donc rs452384 comme un variant fonctionnel de la région de risque du chr5p15.33 dans le MU et représentent les premières étapes dans la caractérisation des conséquences biologiques du locus *TERT/CLPTM1L* dans la prédisposition au MU.

Conclusions générales et perspectives

Nos travaux ont contribué à l'identification et à la caractérisation des facteurs de risque du MU à pénétrance modérée (*MBD4*) et faible (allèles de susceptibilité aux loci *CLPTM1L/TERT*, *HERC2/OCA2* et *IRF4*). Nous avons démontré le rôle de *MBD4* dans la prédisposition au MU à travers un séquençage des exons dans une large cohorte de patients de MU, et avons réalisé une seconde étude de GWAS confirmant l'implication du locus *CLPTM1L/TERT* au chromosome 5p et identifiant deux loci de risque impliqués dans la pigmentation des yeux, *HERC2/OCA2* sur le chromosome 15 et *IRF4* sur le chromosome 6. En stratifiant les patients selon leur statut génomique au niveau du chromosome 3 (monosomie 3, M3, haut risque métastatique, vs. disomie 3, D3, faible risque métastatique), nous avons montré que les

allèles à risque situés sur le locus *HERC2/OCA2* prédisposent spécifiquement au MU à haut risque (M3), et inversement que le locus *IRF4* est exclusivement associé au risque du MU à faible risque (D3). Enfin, nous avons caractérisé par une étude de génomique fonctionnelle le locus *TERT/CLPTM1L*, identifiant rs452384 comme variant fonctionnel impliqué dans le risque du MU sur ce locus, à travers la régulation allèle-spécifique de l'expression génique, et la fixation préférentielle du facteur de transcription NKX2.4 sur l'allèle protecteur T.

Les mutations pathogènes germinales de *MBD4* prédisposent spécifiquement au MU à haut-risque métastatique, probablement en partie dû à la localisation de ce gène sur le chromosome 3, fréquemment perdu dans le MU métastatique. Nous avons montré que ce gène est associé à un risque relatif proche de 10 de développer un MU comparé aux patients de MU non-muté pour *MBD4*, et nous avons estimé la fréquence de mutations germinales de ce gène à 0.7% dans la population de MU, légèrement plus faible que celle des mutations prédisposantes de *BAP1*. Une application importante de ces résultats est l'intégration de *MBD4* dans les panels oncogénétiques, d'autant plus que le spectre tumoral affecté par ce syndrome de prédisposition risque de s'étendre. L'identification de mutations hotspots au sein de notre cohorte de MU et dans des patients affectés par d'autres cancers (notamment leucémie et glioblastome)^{553,623} suggère une origine fondatrice de ces mutations et des potentiels mécanismes oncogéniques communs. Nous avons initié une collaboration avec l'équipe du Dr Ahmed Ibdaih pour cribler les mutations germinales de *MBD4* dans une cohorte de patients de gliome, et avons prévu une analyse d'*Identity by state (IBS)* pour rechercher une origine fondatrice de certaines mutations de *MBD4*.

Un autre impact majeur est que les patients *MBD4*-déficients pourraient répondre à l'immunothérapie, du fait de leur taux mutationnel tumoral élevé et leur phénotype CpG>TpG, suggérant l'utilisation de la déficience en *MBD4* comme biomarqueur pour une potentielle réponse à ces traitements. Nous avons initié une collaboration avec Sandrine Moutel pour produire des *nanobodies* ciblant *MBD4*, qui pourraient être utilisés dans un contexte clinique en immunohistochimie pour détecter l'absence de *MBD4*. La réponse aux traitements anti-PD1 est effectivement très faible chez les patients de MU (de l'ordre de 5%), mais celle-ci s'élève jusqu'à 60% dans le cas d'une déficience *MBD4*.⁴⁶⁵ L'inactivation de *MBD4* pourrait par ailleurs sensibiliser les cellules tumorales à des drogues de chimiothérapie habituellement inefficaces dans le MU ; suivant cette hypothèse, nous entreprenons un large

criblage de l'efficacité de plusieurs drogues sur un modèle cellulaire isogénique de *MBD4* (CRISPR knockout inducible).

Si l'application clinique de la perte d'expression de *MBD4* est cruciale, les mécanismes biologiques sous-jacents demeurent majoritairement inconnus. Plusieurs mutations communes ont été identifiées dans des gènes impliqués dans la régulation épigénétique dans le contexte *MBD4*-déficient, tels que *BAP1* dans le MU, *DNMT3A* dans les leucémies myéloïdes aigues⁶²³ ou *ATRX* dans les glioblastomes,⁶²⁷ en accord avec le rôle de *MBD4* dans la maintenance de la méthylation de l'ADN. Nous sommes en cours d'investigation des conséquences transcriptionnelles et épigénétiques des patients de MU déficients en *MBD4*.

Le second et principal objectif de cette thèse était d'identifier et de caractériser les allèles communs de risque du MU, à la recherche des facteurs héréditaires de risque du MU autres que les rares gènes prédisposant connus (principalement *BAP1*). Le locus de risque *TERT/CLPTM1L* sur le chromosome 5 identifié dans notre étude contient des allèles de risque partagés avec le mélanome cutané (MC) et les cancers du pancréas et du poumon.⁵²⁹ De manière intéressante, rs452384 que nous identifions comme variant fonctionnel dans le MU est aussi impliqué dans le cancer du poumon, même si la direction d'effet est opposée à celle du MU. Notre second GWAS nous a également permis d'identifier d'autres SNPs sur ce même locus, dont certains sont partagés avec le mélanome cutané,^{171,535} tel que rs2447853 impliqué dans le développement des nevi.^{171,529}

Ces résultats indiquent des facteurs étiologiques communs au MU et au MC, au locus *CLPTM1L* du chromosome 5p15.33 mais aussi à travers les allèles à risque de ces deux cancers sur les loci marqueurs de pigmentation, *IRF4* et *HERC2*. Le rôle de ces deux gènes dans la pigmentation des yeux, dont les allèles à risque sont associés avec les yeux clairs, sont en accordance avec l'épidémiologie commune du MU et du MC, prévalents dans les populations à la peau et aux yeux clairs. Cependant, le rôle de l'exposition aux rayons ultra-violets dans le risque du MC et le rôle de la pigmentation dans la protection contre ceux-ci sont connus et bien documentés dans le MC, mais moins évidents dans le MU. Une explication potentielle est l'existence de rôles en dehors de la pigmentation conférant le risque du MU à ces deux loci de prédisposition. C'est ce que nous suggérons dans l'Article 2 en ajoutant à l'analyse GWAS la couleur des yeux en covariable, surtout pour *IRF4*. Cependant, une limitation évidente de notre étude est la circularité de l'analyse étant donné que la couleur des yeux est

ici prédite par certains SNPs connus pour le rôle dans la pigmentation, dont les allèles à risque rs12203592 (*IRF4*) et rs12913832 (*HERC2*). Nous planifions de valider nos résultats dans une cohorte indépendante comprenant l'information de la couleur des yeux, déjà disponible pour les patients de MU traités à l'Institut Curie, et disponible dans une cohorte contrôle (Rotterdam study) à laquelle nous avons demandé l'accès. En dehors du rôle connu de *IRF4* dans la pigmentation, à travers l'*upregulation* de son expression avec l'allèle protecteur C et à travers la synthèse de l'enzyme tyrosinase critique pour la mélanogénèse,^{631,643} le produit de ce gène est aussi connu pour ses fonctions immunes. Récemment, il a été montré que l'expression des gènes *PLA1A*, *NEO1*, *TMEM140* et *MIR3681HG* était associée au génotype de rs12203592 sur *IRF4*;¹³⁴ une étude fonctionnelle de ces gènes dans le MU pourrait ainsi être envisagée pour investiguer les mécanismes biologiques liés à *IRF4*.

Les régions de risque du *IRF4* et *HERC2/OCA2* prédisposent spécifiquement au MU à faible (disomie 3, D3) et haut (monosomie 3, M3) risque métastatique, respectivement. L'existence de facteurs de risque distincts pour ces deux sous-groupes vient s'ajouter à la liste de différences connues entre les MU M3 et D3, aux niveaux génétique, génomique, transcriptomique, épigénétique et clinique.^{260,261,297,410} Nous prévoyons de répliquer et d'étendre nos résultats dans une cohorte plus conséquente au statut génomique connu, et d'effectuer une analyse de GWAS M3 vs. D3, qui pourrait davantage caractériser les facteurs étiologiques distincts entre les deux sous-types de MU, associés à des risques différentiels d'évolution métastatique.

Contrairement aux loci de pigmentation, le locus *CLPTM1L/TERT* sur le chromosome 5p15.33 est autant associé avec le MU M3 que D3, suggérant un mécanisme oncogénique commun lié à ce locus. Le fait que ce locus de risque pan-cancer soit associé à de multiples types tumoraux et avec une direction d'effet parfois opposée rend l'analyse fonctionnelle de ce locus complexe, nécessitant des études cancer-spécifiques. A travers la caractérisation fonctionnelle d'une sous-région du 5p15.33, choisie en raison des marques chromatinienne actives visibles sur la base de données ENCODE (incluant notamment les marques H3K27ac, DNase I, fixation de facteurs de transcription), nous avons identifié un variant fonctionnel, rs452384, lié par fort déséquilibre de liaison au signal d'association marqué par les SNPs rs421284 et rs370348. Bien que nous ne puissions pas écarter l'hypothèse de l'existence d'autres variants fonctionnels à ce locus de risque dans le MU, un rôle important de rs452384 dans la régulation génique et dans le risque du MU est démontré à travers la forte régulation

allèle-spécifique conféré par ce SNP et par l'absence de signal de risque indépendant de cet haplotype. Cependant, notre second GWAS nous a permis d'identifier des allèles de risque additionnels au 5p15.33, sur une région génomique située 3-5kb en aval de la région étudiée, elle aussi enrichie en marques H3K27ac, qui n'a cependant pas pu être étudiée par approches classiques de tests rapporteurs de luciférase en raison de sa séquence mini-satellite hautement répétitive. Nous avons ainsi récemment séquencé ce fragment de 5kb dans l'ADN constitutionnel de 96 patients au génotypage connu pour les SNPs à risque du MU, à travers un séquençage *long-read* (PacBio), et sommes actuellement en cours d'analyse bio-informatique des résultats de séquençage, afin de déterminer si d'autres SNPs associés au risque du MU se trouve dans cette région, et si oui, si ceux-ci sont liés ou indépendants au signal d'association déjà décrit au 5p15.33 dans le MU.

Des travaux supplémentaires seront requis pour déterminer la fonction complète de rs452384 dans le risque du MU et la tumorigenèse. Nos données indiquent le rôle du facteur de transcription NKX2.4 dans la fixation allèle-spécifique au niveau de l'allèle protecteur rs452384-T. Nous nous attendons donc à ce que NKX2.4 induise l'expression de gènes suppresseurs de tumeurs, ou inversement réprime l'expression de gènes aux propriétés oncogéniques, telles que celles impliquées dans la prolifération ou survie cellulaire. Le fait que la répression de *NKX2.4* réprime à son tour l'expression des gènes *TERT* et *CLPTM1L* suggèrent que ces-derniers sont les cibles principales au 5p15.33, en accordance avec les résultats d'études dans d'autres cancers, tels que les cancers pancréatique, ovarien et du poumon.^{183,544,546,636} Nous planifions une étude fonctionnelle de *CLPTM1L* dans la survie cellulaire, la prolifération et la résistance à l'apoptose dans des modèles cellulaires de MU, en raison de la corrélation (légère mais significative) génotype-expression que nous observons entre rs452384 et *CLPTM1L* dans des échantillons tumoraux de MU et du rôle connu de *CLPTM1L* dans ces fonctions. *TERT* et sa régulation télomérique sont aussi des candidats majeurs dans le risque du MU, cependant la très faible expression de *TERT* dans les tumeurs de MU rend son analyse complexe, de manière concordante avec le fait que le MU est un *outlier* bas dans la taille des télomères comparé à de multiples autres types tumoraux.⁶³⁸ Nous sommes actuellement en train d'élaborer un protocole d'extraction des mélanocytes normaux de l'uvée en parallèle du tissu tumoral afin de comparer l'évolution de la longueur télomérique.

En plus de *CLPTM1L* et *TERT*, le locus 5p15.33 pourrait également réguler l'expression d'autres gènes cibles. Une étude de type *TWAS* (*transcriptome-wide association study*), combinant nos données de GWAS à celles d'expression génique dans les mélanocytes primaires en fonction du génotype (*eQTL*) pourrait potentiellement identifier de nouvelles cibles transcriptionnelles en aval de rs452384. Des études transcriptionnelles dans un contexte *NKX2.4* sauvage et *knockout* pourraient par ailleurs permettre l'identification de gènes dont l'expression est liée à *NKX2.4*. Un modèle isogénique de rs452384 (à travers un CRISPR sur lignée cellulaire de MU) est également prévu afin d'étudier le rôle fonctionnel de rs452384.

En conclusion, notre plan d'action suite à ces travaux est de continuer à caractériser les facteurs de risque génétique identifiés dans le MU, afin de raffiner les mécanismes de tumorigenèse et de pouvoir proposer des stratégies visant au diagnostic précoce de MU ou à l'identification de cibles thérapeutiques.

REFERENCES

1. Vogelstein, B. & Kinzler, K.W. Cancer genes and the pathways they control. *Nat Med* **10**, 789-99 (2004).
2. Hanahan, D. & Weinberg, R.A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-74 (2011).
3. Hanahan, D. & Weinberg, R.A. The hallmarks of cancer. *Cell* **100**, 57-70 (2000).
4. Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153-8 (2007).
5. Stratton, M.R., Campbell, P.J. & Futreal, P.A. The cancer genome. *Nature* **458**, 719-24 (2009).
6. Bodmer, W., Bielas, J.H. & Beckman, R.A. Genetic instability is not a requirement for tumor development. *Cancer Res* **68**, 3558-60; discussion 3560-1 (2008).
7. Vogelstein, B. *et al.* Cancer genome landscapes. *Science* **339**, 1546-58 (2013).
8. Stehelin, D., Varmus, H.E., Bishop, J.M. & Vogt, P.K. DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA. *Nature* **260**, 170-3 (1976).
9. Knudson, A.G. Two genetic hits (more or less) to cancer. *Nat Rev Cancer* **1**, 157-62 (2001).
10. Knudson, A.G. Cancer genetics. *Am J Med Genet* **111**, 96-102 (2002).
11. Johnson, A.F., Nguyen, H.T. & Veitia, R.A. Causes and effects of haploinsufficiency. *Biol Rev Camb Philos Soc* **94**, 1774-1785 (2019).
12. Inoue, K. & Fry, E.A. Haploinsufficient tumor suppressor genes. *Adv Med Biol* **118**, 83-122 (2017).
13. Venkatachalam, S. *et al.* Retention of wild-type p53 in tumors from p53 heterozygous mice: reduction of p53 dosage can promote cancer formation. *EMBO J* **17**, 4657-67 (1998).
14. Squatrito, M., Vanoli, F., Schultz, N., Jasin, M. & Holland, E.C. 53BP1 is a haploinsufficient tumor suppressor and protects cells from radiation response in glioma. *Cancer Res* **72**, 5250-60 (2012).
15. Friedberg, E.C. DNA damage and repair. *Nature* **421**, 436-40 (2003).
16. Takeshima, H. & Ushijima, T. Accumulation of genetic and epigenetic alterations in normal cells and cancer risk. *NPJ Precis Oncol* **3**, 7 (2019).
17. Jones, P.A. & Baylin, S.B. The fundamental role of epigenetic events in cancer. *Nat Rev Genet* **3**, 415-28 (2002).
18. Lengauer, C., Kinzler, K.W. & Vogelstein, B. Genetic instabilities in human cancers. *Nature* **396**, 643-9 (1998).
19. Alexandrov, L.B. & Stratton, M.R. Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. *Curr Opin Genet Dev* **24**, 52-60 (2014).
20. Helleday, T., Eshtad, S. & Nik-Zainal, S. Mechanisms underlying mutational signatures in human cancers. *Nat Rev Genet* **15**, 585-98 (2014).
21. Harris, R.S. Cancer mutation signatures, DNA damage mechanisms, and potential clinical implications. *Genome Med* **5**, 87 (2013).
22. Lawrence, M.S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214-218 (2013).
23. Alexandrov, L.B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415-21 (2013).
24. International Cancer Genome, C. *et al.* International network of cancer genome projects. *Nature* **464**, 993-8 (2010).
25. Alexandrov, L.B. *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578**, 94-101 (2020).
26. Consortium, I.T.P.-C.A.o.W.G. Pan-cancer analysis of whole genomes. *Nature* **578**, 82-93 (2020).
27. Pfeifer, G.P. Mutagenesis at methylated CpG sequences. *Curr Top Microbiol Immunol* **301**, 259-81 (2006).
28. Alexandrov, L.B. *et al.* Clock-like mutational processes in human somatic cells. *Nature Genetics* **47**, 1402-1407 (2015).
29. Chan, K. *et al.* An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. *Nat Genet* **47**, 1067-72 (2015).
30. Pfeifer, G.P., You, Y.H. & Besaratinia, A. Mutations induced by ultraviolet light. *Mutat Res* **571**, 19-31 (2005).

31. Maley, C.C. *et al.* Selectively advantageous mutations and hitchhikers in neoplasms: p16 lesions are selected in Barrett's esophagus. *Cancer Res* **64**, 3414-27 (2004).
32. Nowell, P.C. Tumor progression: a brief historical perspective. *Semin Cancer Biol* **12**, 261-6 (2002).
33. Chatsirisupachai, K., Lesluyes, T., Paraoan, L., Van Loo, P. & de Magalhaes, J.P. An integrative analysis of the age-associated multi-omic landscape across cancers. *Nat Commun* **12**, 2345 (2021).
34. de Magalhaes, J.P. How ageing processes influence cancer. *Nat Rev Cancer* **13**, 357-65 (2013).
35. Freed, D., Stevens, E.L. & Pevsner, J. Somatic mosaicism in the human genome. *Genes (Basel)* **5**, 1064-94 (2014).
36. Fernandez, L.C., Torres, M. & Real, F.X. Somatic mosaicism: on the road to cancer. *Nat Rev Cancer* **16**, 43-55 (2016).
37. Machiela, M.J. Mosaicism, aging and cancer. *Curr Opin Oncol* **31**, 108-113 (2019).
38. Ponder, B.A. Cancer genetics. *Nature* **411**, 336-41 (2001).
39. Hussussian, C.J. *et al.* Germline p16 mutations in familial melanoma. *Nat Genet* **8**, 15-21 (1994).
40. Zuo, L. *et al.* Germline mutations in the p16INK4a binding domain of CDK4 in familial melanoma. *Nat Genet* **12**, 97-9 (1996).
41. Puntervoll, H.E. *et al.* Melanoma prone families with CDK4 germline mutation: phenotypic profile and associations with MC1R variants. *J Med Genet* **50**, 264-70 (2013).
42. Welsh, P.L. & King, M.C. BRCA1 and BRCA2 and the genetics of breast and ovarian cancer. *Hum Mol Genet* **10**, 705-13 (2001).
43. Lynch, H.T. & de la Chapelle, A. Genetic susceptibility to non-polyposis colorectal cancer. *J Med Genet* **36**, 801-18 (1999).
44. Wendt, C. & Margolin, S. Identifying breast cancer susceptibility genes - a review of the genetic background in familial breast cancer. *Acta Oncol* **58**, 135-146 (2019).
45. Chen, J. *et al.* Penetrance of Breast and Ovarian Cancer in Women Who Carry a BRCA1/2 Mutation and Do Not Use Risk-Reducing Salpingo-Oophorectomy: An Updated Meta-Analysis. *JNCI Cancer Spectr* **4**, pkaa029 (2020).
46. Aloraifi, F., Boland, M.R., Green, A.J. & Geraghty, J.G. Gene analysis techniques and susceptibility gene discovery in non-BRCA1/BRCA2 familial breast cancer. *Surg Oncol* **24**, 100-9 (2015).
47. Baylin, S.B. & Jones, P.A. Epigenetic Determinants of Cancer. *Cold Spring Harb Perspect Biol* **8**(2016).
48. Baylin, S.B. & Jones, P.A. A decade of exploring the cancer epigenome - biological and translational implications. *Nat Rev Cancer* **11**, 726-34 (2011).
49. Bird, A. Perceptions of epigenetics. *Nature* **447**, 396-8 (2007).
50. Lesch, B.J. *et al.* Intergenerational epigenetic inheritance of cancer susceptibility in mammals. *Elife* **8**(2019).
51. Heard, E. & Martienssen, R.A. Transgenerational epigenetic inheritance: myths and mechanisms. *Cell* **157**, 95-109 (2014).
52. Arico, J.K., Katz, D.J., van der Vlag, J. & Kelly, W.G. Epigenetic patterns maintained in early *Caenorhabditis elegans* embryos can be established by gene activity in the parental germ cells. *PLoS Genet* **7**, e1001391 (2011).
53. Ciabrelli, F. *et al.* Stable Polycomb-dependent transgenerational inheritance of chromatin states in *Drosophila*. *Nat Genet* **49**, 876-886 (2017).
54. Lichtenstein, P. *et al.* Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* **343**, 78-85 (2000).
55. Mucci, L.A. *et al.* Familial Risk and Heritability of Cancer Among Twins in Nordic Countries. *JAMA* **315**, 68-76 (2016).
56. Polderman, T.J. *et al.* Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet* **47**, 702-9 (2015).
57. Visscher, P.M., Hill, W.G. & Wray, N.R. Heritability in the genomics era--concepts and misconceptions. *Nat Rev Genet* **9**, 255-66 (2008).
58. Harris, J.R. *et al.* The Nordic Twin Study on Cancer - NorTwinCan. *Twin Res Hum Genet* **22**, 817-823 (2019).
59. Hart, S.N. *et al.* Mutation prevalence tables for hereditary cancer derived from multigene panel testing. *Hum Mutat* **41**, e1-e6 (2020).
60. Mayer, D.K. *et al.* American Society of Clinical Oncology clinical expert statement on cancer survivorship care planning. *J Oncol Pract* **10**, 345-51 (2014).
61. Qing, T. *et al.* Germline variant burden in cancer genes correlates with age at diagnosis and somatic mutation burden. *Nat Commun* **11**, 2438 (2020).

62. Ramroop, J.R., Gerber, M.M. & Toland, A.E. Germline Variants Impact Somatic Events during Tumorigenesis. *Trends Genet* **35**, 515-526 (2019).
63. Frank, S.A. Genetic predisposition to cancer - insights from population genetics. *Nat Rev Genet* **5**, 764-72 (2004).
64. Clarke, G.M. *et al.* Basic statistical analysis in genetic case-control studies. *Nat Protoc* **6**, 121-33 (2011).
65. Sud, A., Kinnersley, B. & Houlston, R.S. Genome-wide association studies of cancer: current insights and future perspectives. *Nat Rev Cancer* **17**, 692-704 (2017).
66. Ballinger, M.L. *et al.* Monogenic and polygenic determinants of sarcoma risk: an international genetic study. *Lancet Oncol* **17**, 1261-71 (2016).
67. Rahman, N. Realizing the promise of cancer predisposition genes. *Nature* **505**, 302-8 (2014).
68. Glazier, A.M., Nadeau, J.H. & Aitman, T.J. Finding genes that underlie complex traits. *Science* **298**, 2345-9 (2002).
69. McCarthy, M.I., Smedley, D. & Hide, W. New methods for finding disease-susceptibility genes: impact and potential. *Genome Biol* **4**, 119 (2003).
70. Hall, J.M. *et al.* Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* **250**, 1684-9 (1990).
71. Peltomaki, P. *et al.* Genetic mapping of a locus predisposing to human colorectal cancer. *Science* **260**, 810-2 (1993).
72. Cannon-Albright, L.A. *et al.* Assignment of a locus for familial melanoma, MLM, to chromosome 9p13-p22. *Science* **258**, 1148-52 (1992).
73. Miki, Y. *et al.* A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* **266**, 66-71 (1994).
74. Smith, M.J. *et al.* Loss-of-function mutations in SMARCE1 cause an inherited disorder of multiple spinal meningiomas. *Nat Genet* **45**, 295-8 (2013).
75. Comino-Mendez, I. *et al.* Exome sequencing identifies MAX mutations as a cause of hereditary pheochromocytoma. *Nat Genet* **43**, 663-7 (2011).
76. Rotunno, M. *et al.* A Systematic Literature Review of Whole Exome and Genome Sequencing Population Studies of Genetic Susceptibility to Cancer. *Cancer Epidemiol Biomarkers Prev* **29**, 1519-1534 (2020).
77. Liu, P. *et al.* Reanalysis of Clinical Exome Sequencing Data. *N Engl J Med* **380**, 2478-2480 (2019).
78. Cirulli, E.T. *et al.* Genome-wide rare variant analysis for thousands of phenotypes in over 70,000 exomes from two cohorts. *Nat Commun* **11**, 542 (2020).
79. Wang, J. *et al.* Investigation of rare and low-frequency variants using high-throughput sequencing with pooled DNA samples. *Sci Rep* **6**, 33256 (2016).
80. Lee, S., Abecasis, G.R., Boehnke, M. & Lin, X. Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet* **95**, 5-23 (2014).
81. Lu, C. *et al.* Patterns and functional implications of rare germline variants across 12 cancer types. *Nat Commun* **6**, 10086 (2015).
82. Zhang, J. *et al.* Germline Mutations in Predisposition Genes in Pediatric Cancer. *N Engl J Med* **373**, 2336-2346 (2015).
83. Huang, K.L. *et al.* Pathogenic Germline Variants in 10,389 Adult Cancers. *Cell* **173**, 355-370 e14 (2018).
84. Rasnic, R., Linial, N. & Linial, M. Expanding cancer predisposition genes with ultra-rare cancer-exclusive human variations. *Sci Rep* **10**, 13462 (2020).
85. Qi, W., Allen, A.S. & Li, Y.J. Family-based association tests for rare variants with censored traits. *PLoS One* **14**, e0210870 (2019).
86. Li, B. & Leal, S.M. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet* **83**, 311-21 (2008).
87. Wu, M.C. *et al.* Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet* **89**, 82-93 (2011).
88. Ott, J., Wang, J. & Leal, S.M. Genetic linkage analysis in the age of whole-genome sequencing. *Nat Rev Genet* **16**, 275-84 (2015).
89. Wang, X. *et al.* Rare variant association test in family-based sequencing studies. *Brief Bioinform* **18**, 954-961 (2017).
90. Qiao, D. *et al.* Gene-based segregation method for identifying rare variants in family-based sequencing studies. *Genet Epidemiol* **41**, 309-319 (2017).

91. Artomov, M. *et al.* Rare Variant, Gene-Based Association Study of Hereditary Melanoma Using Whole-Exome Sequencing. *J Natl Cancer Inst* **109**(2017).
92. Chubb, D. *et al.* Rare disruptive mutations and their contribution to the heritable risk of colorectal cancer. *Nat Commun* **7**, 11883 (2016).
93. Palles, C. *et al.* Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nat Genet* **45**, 136-44 (2013).
94. Vahteristo, P. *et al.* A CHEK2 genetic variant contributing to a substantial fraction of familial breast cancer. *Am J Hum Genet* **71**, 432-8 (2002).
95. Lohmueller, K.E., Pearce, C.L., Pike, M., Lander, E.S. & Hirschhorn, J.N. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* **33**, 177-82 (2003).
96. Hanks, S. *et al.* Constitutional aneuploidy and cancer predisposition caused by biallelic mutations in BUB1B. *Nat Genet* **36**, 1159-61 (2004).
97. Miyaki, M. *et al.* Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer. *Nat Genet* **17**, 271-2 (1997).
98. Meijers-Heijboer, H. *et al.* Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. *Nat Genet* **31**, 55-9 (2002).
99. Seal, S. *et al.* Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. *Nat Genet* **38**, 1239-41 (2006).
100. Jarhelle, E. *et al.* Identifying sequence variants contributing to hereditary breast and ovarian cancer in BRCA1 and BRCA2 negative breast and ovarian cancer patients. *Sci Rep* **9**, 19986 (2019).
101. Cybulski, C. *et al.* Germline RECQL mutations are associated with breast cancer susceptibility. *Nat Genet* **47**, 643-6 (2015).
102. Renwick, A. *et al.* ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. *Nat Genet* **38**, 873-5 (2006).
103. Rahman, N. *et al.* PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nat Genet* **39**, 165-7 (2007).
104. Loveday, C. *et al.* Germline mutations in RAD51D confer susceptibility to ovarian cancer. *Nat Genet* **43**, 879-882 (2011).
105. Susswein, L.R. *et al.* Pathogenic and likely pathogenic variant prevalence among the first 10,000 patients referred for next-generation cancer panel testing. *Genet Med* **18**, 823-32 (2016).
106. Chakravarti, A. To a future of genetic medicine. *Nature* **409**, 822-3 (2001).
107. Forbes, S.A. *et al.* COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res* **39**, D945-50 (2011).
108. Weinstein, J.N. *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* **45**, 1113-20 (2013).
109. Venter, J.C. *et al.* The sequence of the human genome. *Science* **291**, 1304-51 (2001).
110. Lander, E.S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
111. International HapMap, C. The International HapMap Project. *Nature* **426**, 789-96 (2003).
112. Visscher, P.M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet* **101**, 5-22 (2017).
113. Cano-Gamez, E. & Trynka, G. From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. *Front Genet* **11**, 424 (2020).
114. Andrade, A.C.B., Viana, J.M.S., Pereira, H.D., Pinto, V.B. & Fonseca, E.S.F. Linkage disequilibrium and haplotype block patterns in popcorn populations. *PLoS One* **14**, e0219417 (2019).
115. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G.R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet* **44**, 955-9 (2012).
116. Genomes Project, C. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061-73 (2010).
117. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279-83 (2016).
118. Tam, V. *et al.* Benefits and limitations of genome-wide association studies. *Nat Rev Genet* **20**, 467-484 (2019).
119. Yang, J., Zaitlen, N.A., Goddard, M.E., Visscher, P.M. & Price, A.L. Advantages and pitfalls in the application of mixed-model association methods. *Nat Genet* **46**, 100-6 (2014).
120. Jiang, L. *et al.* A resource-efficient tool for mixed model association analysis of large-scale data. *Nat Genet* **51**, 1749-1755 (2019).

121. Yu, J. *et al.* A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet* **38**, 203-8 (2006).
122. Korte, A. *et al.* A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nat Genet* **44**, 1066-71 (2012).
123. Loh, P.R., Kichaev, G., Gazal, S., Schoech, A.P. & Price, A.L. Mixed-model association for biobank-scale datasets. *Nat Genet* **50**, 906-908 (2018).
124. Chung, J., Jun, G.R., Dupuis, J. & Farrer, L.A. Comparison of methods for multivariate gene-based association tests for complex diseases using common variants. *Eur J Hum Genet* **27**, 811-823 (2019).
125. Eichler, E.E. *et al.* Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet* **11**, 446-50 (2010).
126. Zhu, X. & Stephens, M. Large-scale genome-wide enrichment analyses identify new trait-associated genes and pathways across 31 human phenotypes. *Nat Commun* **9**, 4361 (2018).
127. White, M.J. *et al.* Strategies for Pathway Analysis Using GWAS and WGS Data. *Curr Protoc Hum Genet* **100**, e79 (2019).
128. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369-75, S1-3 (2012).
129. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* **48**, 245-52 (2016).
130. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-5 (2015).
131. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228-35 (2015).
132. Pickrell, J.K. *et al.* Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet* **48**, 709-17 (2016).
133. Zhong, J. *et al.* A Transcriptome-Wide Association Study Identifies Novel Candidate Susceptibility Genes for Pancreatic Cancer. *J Natl Cancer Inst* **112**, 1003-1012 (2020).
134. Zhang, T. *et al.* Cell-type-specific eQTL of primary melanocytes facilitates identification of melanoma susceptibility genes. *Genome Res* **28**, 1621-1635 (2018).
135. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* **44**, 512-25 (2015).
136. Burgess, S., Butterworth, A. & Thompson, S.G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* **37**, 658-65 (2013).
137. Porcu, E. *et al.* Mendelian randomization integrating GWAS and eQTL data reveals genetic determinants of complex and clinical traits. *Nat Commun* **10**, 3300 (2019).
138. Evans, D.M., Visscher, P.M. & Wray, N.R. Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk. *Hum Mol Genet* **18**, 3525-31 (2009).
139. Mavaddat, N. *et al.* Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *Am J Hum Genet* **104**, 21-34 (2019).
140. Wand, H. *et al.* Improving reporting standards for polygenic scores in risk prediction studies. *Nature* **591**, 211-219 (2021).
141. Barnes, D.R. *et al.* Polygenic risk scores and breast and epithelial ovarian cancer risks for carriers of BRCA1 and BRCA2 pathogenic variants. *Genet Med* **22**, 1653-1666 (2020).
142. Kachuri, L. *et al.* Pan-cancer analysis demonstrates that integrating polygenic risk scores with modifiable risk factors improves risk prediction. *Nat Commun* **11**, 6084 (2020).
143. Gallagher, M.D. & Chen-Plotkin, A.S. The Post-GWAS Era: From Association to Function. *Am J Hum Genet* **102**, 717-730 (2018).
144. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**, 1826 (2017).
145. Gabriel, S.B. *et al.* The structure of haplotype blocks in the human genome. *Science* **296**, 2225-9 (2002).
146. Consortium, E.P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
147. Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-30 (2015).
148. Schaid, D.J., Chen, W. & Larson, N.B. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet* **19**, 491-504 (2018).

149. Maurano, M.T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190-5 (2012).
150. Musunuru, K. *et al.* From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature* **466**, 714-9 (2010).
151. Fornes, O. *et al.* JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **48**, D87-D92 (2020).
152. Wingender, E., Dietze, P., Karas, H. & Knuppel, R. TRANSFAC: a database on transcription factors and their DNA binding sites. *Nucleic Acids Res* **24**, 238-41 (1996).
153. Ward, L.D. & Kellis, M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res* **44**, D877-81 (2016).
154. Boyle, A.P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* **22**, 1790-7 (2012).
155. Consortium, G.T. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-5 (2013).
156. Inoue, F. & Ahituv, N. Decoding enhancers using massively parallel reporter assays. *Genomics* **106**, 159-164 (2015).
157. Michailidou, K. *et al.* Association analysis identifies 65 new breast cancer risk loci. *Nature* **551**, 92-94 (2017).
158. Phelan, C.M. *et al.* Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. *Nat Genet* (2017).
159. Al Olama, A.A. *et al.* A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer. *Nat Genet* **46**, 1103-9 (2014).
160. McKay, J.D. *et al.* Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. *Nat Genet* **49**, 1126-1132 (2017).
161. Klein, A.P. *et al.* Genome-wide meta-analysis identifies five new susceptibility loci for pancreatic cancer. *Nat Commun* **9**, 556 (2018).
162. Amundadottir, L. *et al.* Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. *Nat Genet* **41**, 986-90 (2009).
163. Childs, E.J. *et al.* Common variation at 2p13.3, 3q29, 7p13 and 17q25.1 associated with susceptibility to pancreatic cancer. *Nat Genet* **47**, 911-6 (2015).
164. Huyghe, J.R. *et al.* Discovery of common and rare genetic risk variants for colorectal cancer. *Nat Genet* **51**, 76-87 (2019).
165. Schumacher, F.R. *et al.* Genome-wide association study of colorectal cancer identifies six new susceptibility loci. *Nat Commun* **6**, 7138 (2015).
166. Scelo, G. *et al.* Genome-wide association study identifies multiple risk loci for renal cell carcinoma. *Nat Commun* **8**, 15724 (2017).
167. Purdue, M.P. *et al.* Genome-wide association study of renal cell carcinoma identifies two susceptibility loci on 2p21 and 11q13.3. *Nat Genet* **43**, 60-5 (2011).
168. Chang, J. *et al.* Exome-wide analyses identify low-frequency variant in CYP26B1 and additional coding variants associated with esophageal squamous cell carcinoma. *Nat Genet* **50**, 338-343 (2018).
169. Gudmundsson, J. *et al.* A genome-wide association study yields five novel thyroid cancer risk loci. *Nat Commun* **8**, 14517 (2017).
170. Rogounovitch, T.I. *et al.* The common genetic variant rs944289 on chromosome 14q13.3 associates with risk of both malignant and benign thyroid tumors in the Japanese population. *Thyroid* **25**, 333-40 (2015).
171. Landi, M.T. *et al.* Genome-wide association meta-analyses combining multiple risk phenotypes provide insights into the genetic architecture of cutaneous melanoma susceptibility. *Nat Genet* **52**, 494-504 (2020).
172. Law, M.H. *et al.* Genome-wide meta-analysis identifies five new susceptibility loci for cutaneous malignant melanoma. *Nat Genet* **47**, 987-995 (2015).
173. Liang, B., Ding, H., Huang, L., Luo, H. & Zhu, X. GWAS in cancer: progress and challenges. *Mol Genet Genomics* **295**, 537-561 (2020).
174. Park, S.L., Cheng, I. & Haiman, C.A. Genome-Wide Association Studies of Cancer in Diverse Populations. *Cancer Epidemiol Biomarkers Prev* **27**, 405-417 (2018).
175. Purdue, M.P. *et al.* Pathway Analysis of Renal Cell Carcinoma Genome-Wide Association Studies Identifies Novel Associations. *Cancer Epidemiol Biomarkers Prev* **29**, 2065-2069 (2020).

176. Rashkin, S.R. *et al.* Pan-cancer study detects genetic risk variants and shared genetic basis in two large cohorts. *Nat Commun* **11**, 4423 (2020).
177. Sampson, J.N. *et al.* Analysis of Heritability and Shared Heritability Based on Genome-Wide Association Studies for Thirteen Cancer Types. *J Natl Cancer Inst* **107**, djv279 (2015).
178. Lindstrom, S. *et al.* Quantifying the Genetic Correlation between Multiple Cancer Types. *Cancer Epidemiol Biomarkers Prev* **26**, 1427-1435 (2017).
179. Jiang, X. *et al.* Shared heritability and functional enrichment across six solid cancers. *Nat Commun* **10**, 431 (2019).
180. Couch, F.J. *et al.* Genome-wide association study in BRCA1 mutation carriers identifies novel loci associated with breast and ovarian cancer risk. *PLoS Genet* **9**, e1003212 (2013).
181. Eeles, R.A. *et al.* Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. *Nat Genet* **45**, 385-91, 391e1-2 (2013).
182. Kote-Jarai, Z. *et al.* Seven prostate cancer susceptibility loci identified by a multi-stage genome-wide association study. *Nat Genet* **43**, 785-91 (2011).
183. Fang, J. *et al.* Functional characterization of a multi-cancer risk locus on chr5p15.33 reveals regulation of TERT by ZNF148. *Nat Commun* **8**, 15034 (2017).
184. Barrett, J.H. *et al.* Genome-wide association study identifies three new melanoma susceptibility loci. *Nat Genet* **43**, 1108-13 (2011).
185. Rothman, N. *et al.* A multi-stage genome-wide association study of bladder cancer identifies multiple susceptibility loci. *Nat Genet* **42**, 978-84 (2010).
186. McKay, J.D. *et al.* Lung cancer susceptibility locus at 5p15.33. *Nat Genet* **40**, 1404-6 (2008).
187. Bishop, D.T. *et al.* Genome-wide association study identifies three loci associated with melanoma risk. *Nat Genet* **41**, 920-5 (2009).
188. Timofeeva, M.N. *et al.* Influence of common genetic variation on lung cancer risk: meta-analysis of 14 900 cases and 29 485 controls. *Hum Mol Genet* **21**, 4980-95 (2012).
189. Codd, V. *et al.* Identification of seven loci affecting mean telomere length and their association with disease. *Nat Genet* **45**, 422-7, 427e1-2 (2013).
190. Takeda, K., Takahashi, N.H. & Shibahara, S. Neuroendocrine functions of melanocytes: beyond the skin-deep melanin maker. *Tohoku J Exp Med* **211**, 201-21 (2007).
191. Tachibana, M. Sound needs sound melanocytes to be heard. *Pigment Cell Res* **12**, 344-54 (1999).
192. Zecca, L. *et al.* Neuromelanin can protect against iron-mediated oxidative damage in system modeling iron overload of brain aging and Parkinson's disease. *J Neurochem* **106**, 1866-75 (2008).
193. Carneiro, F. *et al.* Relationships between melanocytes, mechanical properties and extracellular matrix composition in mouse heart valves. *J Long Term Eff Med Implants* **25**, 17-26 (2015).
194. Goldgeier, M.H., Klein, L.E., Klein-Angerer, S., Moellmann, G. & Nordlund, J.J. The distribution of melanocytes in the leptomeninges of the human brain. *J Invest Dermatol* **82**, 235-8 (1984).
195. Hu, D.N., Simon, J.D. & Sarna, T. Role of ocular melanin in ophthalmic physiology and pathology. *Photochem Photobiol* **84**, 639-44 (2008).
196. Randhawa, M. *et al.* Evidence for the ectopic synthesis of melanin in human adipose tissue. *FASEB J* **23**, 835-43 (2009).
197. Mort, R.L., Jackson, I.J. & Patton, E.E. The melanocyte lineage in development and disease. *Development* **142**, 1387 (2015).
198. Wilson, Y.M., Richards, K.L., Ford-Perriss, M.L., Panthier, J.J. & Murphy, M. Neural crest cell lineage segregation in the mouse neural tube. *Development* **131**, 6153-62 (2004).
199. Lamoreux, M.L., Wakamatsu, K. & Ito, S. Interaction of major coat color gene functions in mice as studied by chemical analysis of eumelanin and pheomelanin. *Pigment Cell Res* **14**, 23-31 (2001).
200. Baxter, L.L. & Pavan, W.J. Pmel17 expression is Mitf-dependent and reveals cranial melanoblast migration during murine development. *Gene Expr Patterns* **3**, 703-7 (2003).
201. Higdon, C.W., Mitra, R.D. & Johnson, S.L. Gene expression analysis of zebrafish melanocytes, iridophores, and retinal pigmented epithelium reveals indicators of biological function and developmental origin. *PLoS One* **8**, e67801 (2013).
202. Loring, J.F. & Erickson, C.A. Neural crest cell migratory pathways in the trunk of the chick embryo. *Dev Biol* **121**, 220-36 (1987).
203. Bonaventure, J., Domingues, M.J. & Larue, L. Cellular and molecular mechanisms controlling the migration of melanocytes and melanoma cells. *Pigment Cell Melanoma Res* **26**, 316-25 (2013).
204. Cichorek, M., Wachulska, M., Stasiewicz, A. & Tyminska, A. Skin melanocytes: biology and development. *Postepy Dermatol Alergol* **30**, 30-41 (2013).

205. Dupin, E., Calloni, G., Real, C., Goncalves-Trentin, A. & Le Douarin, N.M. Neural crest progenitors and stem cells. *C R Biol* **330**, 521-9 (2007).
206. Harris, M.L. & Erickson, C.A. Lineage specification in neural crest cell pathfinding. *Dev Dyn* **236**, 1-19 (2007).
207. Wiriyasermkul, P., Moriyama, S. & Nagamori, S. Membrane transport proteins in melanosomes: Regulation of ions for pigmentation. *Biochim Biophys Acta Biomembr* **1862**, 183318 (2020).
208. D'Mello, S.A., Finlay, G.J., Baguley, B.C. & Askarian-Amiri, M.E. Signaling Pathways in Melanogenesis. *Int J Mol Sci* **17**(2016).
209. Lin, J.Y. & Fisher, D.E. Melanocyte biology and skin pigmentation. *Nature* **445**, 843-50 (2007).
210. Pavan, W.J. & Sturm, R.A. The Genetics of Human Skin and Hair Pigmentation. *Annu Rev Genomics Hum Genet* **20**, 41-72 (2019).
211. Slominski, A., Zmijewski, M.A. & Pawelek, J. L-tyrosine and L-dihydroxyphenylalanine as hormone-like regulators of melanocyte functions. *Pigment Cell Melanoma Res* **25**, 14-27 (2012).
212. Simon, J.D., Peles, D., Wakamatsu, K. & Ito, S. Current challenges in understanding melanogenesis: bridging chemistry, biological control, morphology, and function. *Pigment Cell Melanoma Res* **22**, 563-79 (2009).
213. Nasti, T.H. & Timares, L. MC1R, eumelanin and pheomelanin: their role in determining the susceptibility to skin cancer. *Photochem Photobiol* **91**, 188-200 (2015).
214. Videira, I.F., Moura, D.F. & Magina, S. Mechanisms regulating melanogenesis. *An Bras Dermatol* **88**, 76-83 (2013).
215. Harris, M.L., Baxter, L.L., Loftus, S.K. & Pavan, W.J. Sox proteins in melanocyte development and melanoma. *Pigment Cell Melanoma Res* **23**, 496-513 (2010).
216. Brenner, M. & Hearing, V.J. The protective role of melanin against UV damage in human skin. *Photochem Photobiol* **84**, 539-49 (2008).
217. Houtzagers, L.E., Wierenga, A.P.A., Ruys, A.A.M., Luyten, G.P.M. & Jager, M.J. Iris Colour and the Risk of Developing Uveal Melanoma. *Int J Mol Sci* **21**(2020).
218. Fujiwara, S. *et al.* Gene Expression and Methylation Analysis in Melanomas and Melanocytes From the Same Patient: Loss of NPM2 Expression Is a Potential Immunohistochemical Marker for Melanoma. *Front Oncol* **8**, 675 (2018).
219. Schadendorf, D. *et al.* Melanoma. *Lancet* **392**, 971-984 (2018).
220. Arisi, M. *et al.* Sun Exposure and Melanoma, Certainties and Weaknesses of the Present Knowledge. *Front Med (Lausanne)* **5**, 235 (2018).
221. Sarna, T. *et al.* Loss of melanin from human RPE with aging: possible role of melanin photooxidation. *Exp Eye Res* **76**, 89-98 (2003).
222. Lopes, V.S., Wasmeier, C., Seabra, M.C. & Futter, C.E. Melanosome maturation defect in Rab38-deficient retinal pigment epithelium results in instability of immature melanosomes during transient melanogenesis. *Mol Biol Cell* **18**, 3914-27 (2007).
223. Jager, M.J. *et al.* Uveal melanoma. *Nat Rev Dis Primers* **6**, 24 (2020).
224. Rodrigues, M. *et al.* So Close, yet so Far: Discrepancies between Uveal and Other Melanomas. A Position Paper from UM Cure 2020. *Cancers (Basel)* **11**, 1032 (2019).
225. van der Kooij, M.K., Speetjens, F.M., van der Burg, S.H. & Kapiteijn, E. Uveal Versus Cutaneous Melanoma; Same Origin, Very Distinct Tumor Types. *Cancers (Basel)* **11**(2019).
226. Pandiani, C., Beranger, G.E., Leclerc, J., Ballotti, R. & Bertolotto, C. Focus on cutaneous and uveal melanoma specificities. *Genes Dev* **31**, 724-743 (2017).
227. Griewank, K.G. *et al.* Integrated Genomic Classification of Melanocytic Tumors of the Central Nervous System Using Mutation Analysis, Copy Number Alterations, and DNA Methylation Profiling. *Clin Cancer Res* **24**, 4494-4504 (2018).
228. Sitiwin, E. *et al.* Shedding light on melanins within in situ human eye melanocytes using 2-photon microscopy profiling techniques. *Sci Rep* **9**, 18585 (2019).
229. Hu, D.N. Regulation of growth and melanogenesis of uveal melanocytes. *Pigment Cell Res* **13 Suppl 8**, 81-6 (2000).
230. Wade, N.J. & Finger, S. The eye as an optical instrument: from camera obscura to Helmholtz's perspective. *Perception* **30**, 1157-77 (2001).
231. Sliney, D.H. How light reaches the eye and its components. *Int J Toxicol* **21**, 501-9 (2002).
232. Hong, L., Simon, J.D. & Sarna, T. Melanin structure and the potential functions of uveal melanosomes. *Pigment Cell Res* **19**, 465-6 (2006).

233. Mochizuki, M., Sugita, S. & Kamoi, K. Immunological homeostasis of the eye. *Prog Retin Eye Res* **33**, 10-27 (2013).
234. Wakamatsu, K., Hu, D.N., McCormick, S.A. & Ito, S. Characterization of melanin in human iridal and choroidal melanocytes from eyes with various colored irides. *Pigment Cell Melanoma Res* **21**, 97-105 (2008).
235. Wilkerson, C.L. *et al.* Melanocytes and iris color. Light microscopic findings. *Arch Ophthalmol* **114**, 437-42 (1996).
236. Li, L. *et al.* Uveal melanocytes do not respond to or express receptors for alpha-melanocyte-stimulating hormone. *Invest Ophthalmol Vis Sci* **47**, 4507-12 (2006).
237. Frudakis, T., Terravainen, T. & Thomas, M. Multilocus OCA2 genotypes specify human iris colors. *Hum Genet* **122**, 311-26 (2007).
238. White, D. & Rabago-Smith, M. Genotype-phenotype associations and human eye color. *J Hum Genet* **56**, 5-7 (2011).
239. Pospiech, E., Draus-Barini, J., Kupiec, T., Wojas-Pelc, A. & Branicki, W. Gene-gene interactions contribute to eye colour variation in humans. *J Hum Genet* **56**, 447-55 (2011).
240. Visser, M., Kayser, M. & Palstra, R.J. HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter. *Genome Res* **22**, 446-55 (2012).
241. Hirobe, T., Ito, S. & Wakamatsu, K. The mouse pink-eyed dilution allele of the P-gene greatly inhibits eumelanin but not pheomelanin synthesis. *Pigment Cell Melanoma Res* **24**, 241-6 (2011).
242. Walsh, S. *et al.* IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. *Forensic Sci Int Genet* **5**, 170-80 (2011).
243. Chaitanya, L. *et al.* The HirisPlex-S system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation. *Forensic Sci Int Genet* **35**, 123-135 (2018).
244. Walsh, S. *et al.* The HirisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Sci Int Genet* **7**, 98-115 (2013).
245. Sanchez-Mazas, A. & Meyer, D. The relevance of HLA sequencing in population genetics studies. *J Immunol Res* **2014**, 971818 (2014).
246. Sturm, R.A. & Duffy, D.L. Human pigmentation genes under environmental selection. *Genome Biol* **13**, 248 (2012).
247. Frost, P. European hair and eye color: A case of frequency-dependent sexual selection? *Evol. Hum. Behav.* **27:2**, 85-103 (2006).
248. Yuen, A.W. & Jablonski, N.G. Vitamin D: in the evolution of human skin colour. *Med Hypotheses* **74**, 39-44 (2010).
249. Singh, A.D., Turell, M.E. & Topham, A.K. Uveal melanoma: trends in incidence, treatment, and survival. *Ophthalmology* **118**, 1881-5 (2011).
250. Shields, C.L. *et al.* Iris melanoma: features and prognosis in 317 children and adults. *J AAPOS* **16**, 10-6 (2012).
251. Augsburger, J.J., Correa, Z.M. & Shaikh, A.H. Effectiveness of treatments for metastatic uveal melanoma. *Am J Ophthalmol* **148**, 119-27 (2009).
252. Carvajal, R.D. *et al.* Selumetinib in Combination With Dacarbazine in Patients With Metastatic Uveal Melanoma: A Phase III, Multicenter, Randomized Trial (SUMIT). *J Clin Oncol* **36**, 1232-1239 (2018).
253. Piperno-Neumann, S., Hassel, J.C., Rutkowski, P., *et al.* Phase 3 randomized trial compared tebentafusp with investigator's choice in first line metastatic uveal melanoma. . *Presented at: AACR Annual Meeting 2021; April 10-15, 2021; Virtual. CT002* (2021).
254. Sturm, V. & Richard, G. [The prevalence of bilateral malignant uveal melanoma]. *Klin Monbl Augenheilkd* **224**, 770-4 (2007).
255. Singh, A.D., Shields, C.L., Shields, J.A. & De Potter, P. Bilateral primary uveal melanoma. Bad luck or bad genes? *Ophthalmology* **103**, 256-62 (1996).
256. Echegaray, J.J., Medina, C.A., Biscotti, C.V., Plesec, T. & Singh, A.D. Multifocal Primary Uveal Melanoma: Clinical and Molecular Characteristics. *Ocul Oncol Pathol* **5**, 8-12 (2019).
257. Rodriguez, A., Duenas-Gonzalez, A. & Delgado-Pelayo, S. Clinical presentation and management of uveal melanoma. *Mol Clin Oncol* **5**, 675-677 (2016).
258. Johansson, P. *et al.* Deep sequencing of uveal melanoma identifies a recurrent mutation in PLCB4. *Oncotarget* **7**, 4624-31 (2016).
259. Furney, S.J. *et al.* SF3B1 mutations are associated with alternative splicing in uveal melanoma. *Cancer Discov* **3**, 1122-1129 (2013).

260. Robertson, A.G. *et al.* Integrative Analysis Identifies Four Molecular and Clinical Subsets in Uveal Melanoma. *Cancer Cell* **32**, 204-220 e15 (2017).
261. Royer-Bertrand, B. *et al.* Comprehensive Genetic Landscape of Uveal Melanoma by Whole-Genome Sequencing. *Am J Hum Genet* **99**, 1190-1198 (2016).
262. Ikehata, H. & Ono, T. The mechanisms of UV mutagenesis. *J Radiat Res* **52**, 115-25 (2011).
263. Berger, M.F. *et al.* Melanoma genome sequencing reveals frequent PREX2 mutations. *Nature* **485**, 502-6 (2012).
264. Cancer Genome Atlas, N. Genomic Classification of Cutaneous Melanoma. *Cell* **161**, 1681-96 (2015).
265. Johansson, P.A. *et al.* Whole genome landscapes of uveal melanoma show an ultraviolet radiation signature in iris tumours. *Nat Commun* **11**, 2408 (2020).
266. Field, M.G. *et al.* Punctuated evolution of canonical genomic aberrations in uveal melanoma. *Nat Commun* **9**, 116 (2018).
267. Onken, M.D. *et al.* Oncogenic mutations in GNAQ occur early in uveal melanoma. *Invest Ophthalmol Vis Sci* **49**, 5230-4 (2008).
268. Van Raamsdonk, C.D. *et al.* Mutations in GNA11 in uveal melanoma. *N Engl J Med* **363**, 2191-9 (2010).
269. Van Raamsdonk, C.D. *et al.* Frequent somatic mutations of GNAQ in uveal melanoma and blue naevi. *Nature* **457**, 599-602 (2009).
270. Shoushtari, A.N. & Carvajal, R.D. GNAQ and GNA11 mutations in uveal melanoma. *Melanoma Res* **24**, 525-34 (2014).
271. Amaro, A. *et al.* The biology of uveal melanoma. *Cancer Metastasis Rev* **36**, 109-140 (2017).
272. Hitchman, T.D. *et al.* Combined Inhibition of Gα(q) and MEK Enhances Therapeutic Efficacy in Uveal Melanoma. *Clin Cancer Res* **27**, 1476-1490 (2021).
273. Gerami, P., Pouryazdanparast, P., Vemula, S. & Bastian, B.C. Molecular analysis of a case of nevus of ota showing progressive evolution to melanoma with intermediate stages resembling cellular blue nevus. *Am J Dermatopathol* **32**, 301-305 (2010).
274. Shirley, M.D. *et al.* Sturge-Weber syndrome and port-wine stains caused by somatic mutation in GNAQ. *N Engl J Med* **368**, 1971-9 (2013).
275. Vader, M.J.C. *et al.* GNAQ and GNA11 mutations and downstream YAP activation in choroidal nevi. *Br J Cancer* **117**, 884-887 (2017).
276. Gresset, A., Sondek, J. & Harden, T.K. The phospholipase C isozymes and their regulation. *Subcell Biochem* **58**, 61-94 (2012).
277. Lyon, A.M. & Tesmer, J.J. Structural insights into phospholipase C-beta function. *Mol Pharmacol* **84**, 488-500 (2013).
278. Moore, A.R. *et al.* Recurrent activating mutations of G-protein-coupled receptor CYSLTR2 in uveal melanoma. *Nat Genet* **48**, 675-80 (2016).
279. Evans, J.F. Cysteinyl leukotriene receptors. *Prostaglandins Other Lipid Mediat* **68-69**, 587-97 (2002).
280. Vivet-Noguer, R., Tarin, M., Roman-Roman, S. & Alsafadi, S. Emerging Therapeutic Opportunities Based on Current Knowledge of Uveal Melanoma Biology. *Cancers (Basel)* **11**(2019).
281. Weber, A. *et al.* Absence of mutations of the BRAF gene and constitutive activation of extracellular-regulated kinase in malignant melanomas of the uvea. *Lab Invest* **83**, 1771-6 (2003).
282. Zuidervaat, W. *et al.* Activation of the MAPK pathway is a common event in uveal melanomas although it rarely occurs through mutation of BRAF or RAS. *Br J Cancer* **92**, 2032-8 (2005).
283. Rimoldi, D. *et al.* Lack of BRAF mutations in uveal melanoma. *Cancer Res* **63**, 5712-5 (2003).
284. Mouti, M.A., Dee, C., Coupland, S.E. & Hurlstone, A.F. Minimal contribution of ERK1/2-MAPK signalling towards the maintenance of oncogenic GNAQQ209P-driven uveal melanomas in zebrafish. *Oncotarget* **7**, 39654-39670 (2016).
285. Yoo, J.H. *et al.* ARF6 Is an Actionable Node that Orchestrates Oncogenic GNAQ Signaling in Uveal Melanoma. *Cancer Cell* **29**, 889-904 (2016).
286. Yu, F.X. *et al.* Mutant Gq/11 promote uveal melanoma tumorigenesis by activating YAP. *Cancer Cell* **25**, 822-30 (2014).
287. Feng, X. *et al.* Hippo-independent activation of YAP by the GNAQ uveal melanoma oncogene through a trio-regulated rho GTPase signaling circuitry. *Cancer Cell* **25**, 831-45 (2014).
288. Yu, F.X., Zhao, B. & Guan, K.L. Hippo Pathway in Organ Size Control, Tissue Homeostasis, and Cancer. *Cell* **163**, 811-28 (2015).
289. Zanconato, F., Cordenonsi, M. & Piccolo, S. YAP/TAZ at the Roots of Cancer. *Cancer Cell* **29**, 783-803 (2016).

290. Feng, X. *et al.* A Platform of Synthetic Lethal Gene Interaction Networks Reveals that the GNAQ Uveal Melanoma Oncogene Controls the Hippo Pathway through FAK. *Cancer Cell* **35**, 457-472 e5 (2019).
291. Paradis, J.S. *et al.* Synthetic Lethal Screens Reveal Cotargeting FAK and MEK as a Multimodal Precision Therapy for GNAQ-Driven Uveal Melanoma. *Clin Cancer Res* **27**, 3190-3200 (2021).
292. Amirouchene-Angelozzi, N. *et al.* The mTOR inhibitor Everolimus synergizes with the PI3K inhibitor GDC0941 to enhance anti-tumor efficacy in uveal melanoma. *Oncotarget* **7**, 23633-46 (2016).
293. Amirouchene-Angelozzi, N. *et al.* Establishment of novel cell lines recapitulating the genetic landscape of uveal melanoma and preclinical validation of mTOR as a therapeutic target. *Mol Oncol* **8**, 1508-20 (2014).
294. Babchia, N., Calipel, A., Mouriaux, F., Faussat, A.M. & Mascarelli, F. The PI3K/Akt and mTOR/P70S6K signaling pathways in human uveal melanoma cells: interaction with B-Raf/ERK. *Invest Ophthalmol Vis Sci* **51**, 421-9 (2010).
295. Shoushtari, A.N. *et al.* A phase 2 trial of everolimus and pasireotide long-acting release in patients with metastatic uveal melanoma. *Melanoma Res* **26**, 272-7 (2016).
296. Koopmans, A.E. *et al.* Patient survival in uveal melanoma is not affected by oncogenic mutations in GNAQ and GNA11. *Br J Cancer* **109**, 493-6 (2013).
297. Harbour, J.W. *et al.* Frequent mutation of BAP1 in metastasizing uveal melanomas. *Science* **330**, 1410-3 (2010).
298. Karlsson, J. *et al.* Molecular profiling of driver events in metastatic uveal melanoma. *Nat Commun* **11**, 1894 (2020).
299. Yavuziyigitoglu, S. *et al.* Metastatic Disease in Polyploid Uveal Melanoma Patients Is Associated With BAP1 Mutations. *Invest Ophthalmol Vis Sci* **57**, 2232-9 (2016).
300. Martin, M. *et al.* Exome sequencing identifies recurrent somatic mutations in EIF1AX and SF3B1 in uveal melanoma with disomy 3. *Nat Genet* **45**, 933-6 (2013).
301. Harbour, J.W. *et al.* Recurrent mutations at codon 625 of the splicing factor SF3B1 in uveal melanoma. *Nat Genet* **45**, 133-5 (2013).
302. Prescher, G. *et al.* Prognostic implications of monosomy 3 in uveal melanoma. *Lancet* **347**, 1222-5 (1996).
303. Murali, R., Wiesner, T. & Scolyer, R.A. Tumours associated with BAP1 mutations. *Pathology* **45**, 116-26 (2013).
304. Gupta, M.P. *et al.* Clinical Characteristics of Uveal Melanoma in Patients With Germline BAP1 Mutations. *JAMA Ophthalmol* **133**, 881-7 (2015).
305. Aoude, L.G., Vajdic, C.M., Krickler, A., Armstrong, B. & Hayward, N.K. Prevalence of germline BAP1 mutation in a population-based sample of uveal melanoma cases. *Pigment Cell Melanoma Res* **26**, 278-9 (2013).
306. Singh, N., Singh, R., Bowen, R.C., Abdel-Rahman, M.H. & Singh, A.D. Uveal Melanoma in BAP1 Tumor Predisposition Syndrome: Estimation of Risk. *Am J Ophthalmol* **224**, 172-177 (2021).
307. Abdel-Rahman, M.H. *et al.* Germline BAP1 mutation predisposes to uveal melanoma, lung adenocarcinoma, meningioma, and other cancers. *J Med Genet* **48**, 856-9 (2011).
308. Walpole, S. *et al.* Comprehensive Study of the Clinical Phenotype of Germline BAP1 Variant-Carrying Families Worldwide. *J Natl Cancer Inst* **110**, 1328-1341 (2018).
309. Popova, T. *et al.* Germline BAP1 mutations predispose to renal cell carcinomas. *Am J Hum Genet* **92**, 974-80 (2013).
310. Testa, J.R. *et al.* Germline BAP1 mutations predispose to malignant mesothelioma. *Nat Genet* **43**, 1022-5 (2011).
311. Wiesner, T. *et al.* Germline mutations in BAP1 predispose to melanocytic tumors. *Nat Genet* **43**, 1018-21 (2011).
312. Carbone, M. *et al.* BAP1 cancer syndrome: malignant mesothelioma, uveal and cutaneous melanoma, and MIBAITs. *J Transl Med* **10**, 179 (2012).
313. Pilarski, R. *et al.* Expanding the clinical phenotype of hereditary BAP1 cancer predisposition syndrome, reporting three new cases. *Genes Chromosomes Cancer* **53**, 177-82 (2014).
314. Carbone, M. *et al.* Biological Mechanisms and Clinical Significance of BAP1 Mutations in Human Cancer. *Cancer Discov* **10**, 1103-1120 (2020).
315. Louie, B.H. & Kurzrock, R. BAP1: Not just a BRCA1-associated protein. *Cancer Treat Rev* **90**, 102091 (2020).
316. Jensen, D.E. *et al.* BAP1: a novel ubiquitin hydrolase which binds to the BRCA1 RING finger and enhances BRCA1-mediated cell growth suppression. *Oncogene* **16**, 1097-112 (1998).

317. Hauri, S. *et al.* A High-Density Map for Navigating the Human Polycomb Complexome. *Cell Rep* **17**, 583-595 (2016).
318. Nishikawa, H. *et al.* BRCA1-associated protein 1 interferes with BRCA1/BARD1 RING heterodimer activity. *Cancer Res* **69**, 111-9 (2009).
319. Mallery, D.L., Vandenberg, C.J. & Hiom, K. Activation of the E3 ligase function of the BRCA1/BARD1 complex by polyubiquitin chains. *EMBO J* **21**, 6755-62 (2002).
320. Lee, H.S. *et al.* BAP1 promotes stalled fork restart and cell survival via INO80 in response to replication stress. *Biochem J* **476**, 3053-3066 (2019).
321. Machida, Y.J., Machida, Y., Vashisht, A.A., Wohlschlegel, J.A. & Dutta, A. The deubiquitinating enzyme BAP1 regulates cell growth via interaction with HCF-1. *J Biol Chem* **284**, 34179-88 (2009).
322. Okino, Y., Machida, Y., Frankland-Searby, S. & Machida, Y.J. BRCA1-associated protein 1 (BAP1) deubiquitinase antagonizes the ubiquitin-mediated activation of FoxK2 target genes. *J Biol Chem* **290**, 1580-91 (2015).
323. Daou, S. *et al.* The BAP1/ASXL2 Histone H2A Deubiquitinase Complex Regulates Cell Proliferation and Is Disrupted in Cancer. *J Biol Chem* **290**, 28643-63 (2015).
324. Daou, S. *et al.* Monoubiquitination of ASXLs controls the deubiquitinase activity of the tumor suppressor BAP1. *Nat Commun* **9**, 4385 (2018).
325. Zhang, Y. *et al.* BAP1 links metabolic regulation of ferroptosis to tumour suppression. *Nat Cell Biol* **20**, 1181-1192 (2018).
326. Yu, H. *et al.* Tumor suppressor and deubiquitinase BAP1 promotes DNA double-strand break repair. *Proc Natl Acad Sci U S A* **111**, 285-90 (2014).
327. Bononi, A. *et al.* BAP1 regulates IP3R3-mediated Ca(2+) flux to mitochondria suppressing cell transformation. *Nature* **546**, 549-553 (2017).
328. He, M. *et al.* Intrinsic apoptosis shapes the tumor spectrum linked to inactivation of the deubiquitinase BAP1. *Science* **364**, 283-285 (2019).
329. Dai, F. *et al.* BAP1 inhibits the ER stress gene regulatory network and modulates metabolic stress response. *Proc Natl Acad Sci U S A* **114**, 3192-3197 (2017).
330. Scheuermann, J.C. *et al.* Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. *Nature* **465**, 243-7 (2010).
331. Gaytan de Ayala Alonso, A. *et al.* A genetic screen identifies novel polycomb group genes in *Drosophila*. *Genetics* **176**, 2099-108 (2007).
332. Di Croce, L. & Helin, K. Transcriptional regulation by Polycomb group proteins. *Nat Struct Mol Biol* **20**, 1147-55 (2013).
333. Field, M.G. *et al.* BAP1 Loss Is Associated with DNA Methyloomic Repatterning in Highly Aggressive Class 2 Uveal Melanomas. *Clin Cancer Res* **25**, 5663-5673 (2019).
334. Lee, H.J. *et al.* The Tumor Suppressor BAP1 Regulates the Hippo Pathway in Pancreatic Ductal Adenocarcinoma. *Cancer Res* **80**, 1656-1668 (2020).
335. Cazzola, M., Rossi, M., Malcovati, L. & Associazione Italiana per la Ricerca sul Cancro Gruppo Italiano Malattie, M. Biologic and clinical significance of somatic mutations of SF3B1 in myeloid and lymphoid neoplasms. *Blood* **121**, 260-9 (2013).
336. Visconte, V. *et al.* SF3B1, a splicing factor is frequently mutated in refractory anemia with ring sideroblasts. *Leukemia* **26**, 542-5 (2012).
337. Yoshida, K. *et al.* Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478**, 64-9 (2011).
338. Yavuzigitoglu, S. *et al.* Uveal Melanomas with SF3B1 Mutations: A Distinct Subclass Associated with Late-Onset Metastases. *Ophthalmology* **123**, 1118-28 (2016).
339. Gozani, O., Potashkin, J. & Reed, R. A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. *Mol Cell Biol* **18**, 4752-60 (1998).
340. Alsafadi, S. *et al.* Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. *Nat Commun* **7**, 10615 (2016).
341. Darman, R.B. *et al.* Cancer-Associated SF3B1 Hotspot Mutations Induce Cryptic 3' Splice Site Selection through Use of a Different Branch Point. *Cell Rep* **13**, 1033-45 (2015).
342. DeBoever, C. *et al.* Transcriptome sequencing reveals potential mechanism of cryptic 3' splice site selection in SF3B1-mutated cancers. *PLoS Comput Biol* **11**, e1004105 (2015).
343. Jayasinghe, R.G. *et al.* Systematic Analysis of Splice-Site-Creating Mutations in Cancer. *Cell Rep* **23**, 270-281 e3 (2018).

344. Zhang, J. *et al.* Disease-Causing Mutations in SF3B1 Alter Splicing by Disrupting Interaction with SUGP1. *Mol Cell* (2019).
345. Alsafadi, S. *et al.* Genetic alterations of SUGP1 mimic mutant-SF3B1 splice pattern in lung adenocarcinoma and other cancers. *Oncogene* **40**, 85-96 (2021).
346. Gatchalian, J. *et al.* A non-canonical BRD9-containing BAF chromatin remodeling complex regulates naive pluripotency in mouse embryonic stem cells. *Nat Commun* **9**, 5139 (2018).
347. Inoue, D. *et al.* Spliceosomal disruption of the non-canonical BAF complex in cancer. *Nature* **574**, 432-436 (2019).
348. Zhou, Z. *et al.* The biological function and clinical significance of SF3B1 mutations in cancer. *Biomark Res* **8**, 38 (2020).
349. Liu, Z. *et al.* Mutations in the RNA Splicing Factor SF3B1 Promote Tumorigenesis through MYC Stabilization. *Cancer Discov* **10**, 806-821 (2020).
350. Bigot, J. *et al.* Splicing patterns in SF3B1 mutated uveal melanoma generate shared immunogenic tumor-specific neo-epitopes. *Cancer Discov* (2021).
351. Zhang, J. *et al.* Disease-associated mutation in SRSF2 misregulates splicing by altering RNA-binding affinities. *Proc Natl Acad Sci U S A* **112**, E4726-34 (2015).
352. Komeno, Y. *et al.* SRSF2 Is Essential for Hematopoiesis, and Its Myelodysplastic Syndrome-Related Mutations Dysregulate Alternative Pre-mRNA Splicing. *Mol Cell Biol* **35**, 3071-82 (2015).
353. Ewens, K.G. *et al.* Chromosome 3 status combined with BAP1 and EIF1AX mutation profiles are associated with metastasis in uveal melanoma. *Invest Ophthalmol Vis Sci* **55**, 5160-7 (2014).
354. Saini, A.K., Nanda, J.S., Lorsch, J.R. & Hinnebusch, A.G. Regulatory elements in eIF1A control the fidelity of start codon selection by modulating tRNA(i)(Met) binding to the ribosome. *Genes Dev* **24**, 97-110 (2010).
355. Johnson, C.P. *et al.* Systematic genomic and translational efficiency studies of uveal melanoma. *PLoS One* **12**, e0178189 (2017).
356. Kunstman, J.W. *et al.* Characterization of the mutational landscape of anaplastic thyroid cancer via whole-exome sequencing. *Hum Mol Genet* **24**, 2318-29 (2015).
357. Cancer Genome Atlas Research, N. Integrated genomic characterization of papillary thyroid carcinoma. *Cell* **159**, 676-90 (2014).
358. Krishnamoorthy, G.P. *et al.* EIF1AX and RAS Mutations Cooperate to Drive Thyroid Tumorigenesis through ATF4 and c-MYC. *Cancer Discov* **9**, 264-281 (2019).
359. Yoo, S.K. *et al.* Integrative analysis of genomic and transcriptomic characteristics associated with progression of aggressive thyroid cancer. *Nat Commun* **10**, 2764 (2019).
360. Robertson, A.G. *et al.* Integrative Analysis Identifies Four Molecular and Clinical Subsets in Uveal Melanoma. *Cancer Cell* **33**, 151 (2018).
361. Trolet, J. *et al.* Genomic profiling and identification of high-risk uveal melanoma by array CGH analysis of primary tumors and liver metastases. *Invest Ophthalmol Vis Sci* **50**, 2572-80 (2009).
362. Damato, B., Dopierala, J.A. & Coupland, S.E. Genotypic profiling of 452 choroidal melanomas with multiplex ligation-dependent probe amplification. *Clin Cancer Res* **16**, 6083-92 (2010).
363. Shields, C.L. *et al.* Cytogenetic Abnormalities in Uveal Melanoma Based on Tumor Features and Size in 1059 Patients: The 2016 W. Richard Green Lecture. *Ophthalmology* **124**, 609-618 (2017).
364. Horsman, D.E. & White, V.A. Cytogenetic analysis of uveal melanoma. Consistent occurrence of monosomy 3 and trisomy 8q. *Cancer* **71**, 811-9 (1993).
365. Sisley, K. *et al.* Abnormalities of chromosomes 3 and 8 in posterior uveal melanoma correlate with prognosis. *Genes Chromosomes Cancer* **19**, 22-8 (1997).
366. Cassoux, N. *et al.* Genome-wide profiling is a clinically relevant and affordable prognostic test in posterior uveal melanoma. *Br J Ophthalmol* **98**, 769-74 (2014).
367. Aalto, Y., Eriksson, L., Seregard, S., Larsson, O. & Knuutila, S. Concomitant loss of chromosome 3 and whole arm losses and gains of chromosome 1, 6, or 8 in metastasizing primary uveal melanoma. *Invest Ophthalmol Vis Sci* **42**, 313-7 (2001).
368. Shields, C.L. *et al.* Prognosis of uveal melanoma in 500 cases using genetic testing of fine-needle aspiration biopsy specimens. *Ophthalmology* **118**, 396-401 (2011).
369. Damato, B. & Coupland, S.E. Translating uveal melanoma cytogenetics into clinical care. *Arch Ophthalmol* **127**, 423-9 (2009).
370. Cross, N.A. *et al.* Multiple locations on chromosome 3 are the targets of specific deletions in uveal melanoma. *Eye (Lond)* **20**, 476-81 (2006).

371. Rodrigues, M. *et al.* Association of Partial Chromosome 3 Deletion in Uveal Melanomas With Metastasis-Free Survival. *JAMA Ophthalmol* (2020).
372. White, V.A., McNeil, B.K., Thiberville, L. & Horsman, D.E. Acquired homozygosity (isodisomy) of chromosome 3 during clonal evolution of a uveal melanoma: association with morphologic heterogeneity. *Genes Chromosomes Cancer* **15**, 138-43 (1996).
373. Chang, M.Y., Rao, N.P., Burgess, B.L., Johnson, L. & McCannel, T.A. Heterogeneity of monosomy 3 in fine needle aspiration biopsy of choroidal melanoma. *Mol Vis* **19**, 1892-900 (2013).
374. Mensink, H.W. *et al.* Chromosome 3 intratumor heterogeneity in uveal melanoma. *Invest Ophthalmol Vis Sci* **50**, 500-4 (2009).
375. de Lange, M.J. *et al.* Heterogeneity revealed by integrated genomic analysis uncovers a molecular switch in malignant uveal melanoma. *Oncotarget* **6**, 37824-35 (2015).
376. Damato, B. *et al.* Multiplex ligation-dependent probe amplification of uveal melanoma: correlation with metastatic death. *Invest Ophthalmol Vis Sci* **50**, 3048-55 (2009).
377. Laurent, C. *et al.* High PTP4A3 phosphatase expression correlates with metastatic risk in uveal melanoma patients. *Cancer Res* **71**, 666-74 (2011).
378. Onken, M.D., Worley, L.A., Ehlers, J.P. & Harbour, J.W. Gene expression profiling in uveal melanoma reveals two molecular classes and predicts metastatic death. *Cancer Res* **64**, 7205-9 (2004).
379. Chokhachi Baradaran, P., Kozovska, Z., Furdova, A. & Smolkova, B. Targeting Epigenetic Modifications in Uveal Melanoma. *Int J Mol Sci* **21**(2020).
380. Li, Y., Jia, R. & Ge, S. Role of Epigenetics in Uveal Melanoma. *Int J Biol Sci* **13**, 426-433 (2017).
381. Worley, L.A., Long, M.D., Onken, M.D. & Harbour, J.W. Micro-RNAs associated with metastasis in uveal melanoma identified by multiplexed microarray profiling. *Melanoma Res* **18**, 184-90 (2008).
382. Radhakrishnan, A., Badhrinarayanan, N., Biswas, J. & Krishnakumar, S. Analysis of chromosomal aberration (1, 3, and 8) and association of microRNAs in uveal melanoma. *Mol Vis* **15**, 2146-54 (2009).
383. van der Velden, P.A. *et al.* Promoter hypermethylation: a common cause of reduced p16(INK4a) expression in uveal melanoma. *Cancer Res* **61**, 5303-6 (2001).
384. Maat, W. *et al.* Epigenetic inactivation of RASSF1a in uveal melanoma. *Invest Ophthalmol Vis Sci* **48**, 486-90 (2007).
385. Merhavi, E. *et al.* Promoter methylation status of multiple genes in uveal melanoma. *Invest Ophthalmol Vis Sci* **48**, 4403-6 (2007).
386. Field, M.G. *et al.* PRAME as an Independent Biomarker for Metastasis in Uveal Melanoma. *Clin Cancer Res* **22**, 1234-42 (2016).
387. Landreville, S. *et al.* Histone deacetylase inhibitors induce growth arrest and differentiation in uveal melanoma. *Clin Cancer Res* **18**, 408-16 (2012).
388. Matatall, K.A. *et al.* BAP1 deficiency causes loss of melanocytic cell identity in uveal melanoma. *BMC Cancer* **13**, 371 (2013).
389. Ny, L. *et al.* The PEMDAC phase 2 study of pembrolizumab and entinostat in patients with metastatic uveal melanoma. *Nat Commun* **12**, 5155 (2021).
390. Diener-West, M. *et al.* Development of metastatic disease after enrollment in the COMS trials for treatment of choroidal melanoma: Collaborative Ocular Melanoma Study Group Report No. 26. *Arch Ophthalmol* **123**, 1639-43 (2005).
391. Shields, C.L. *et al.* Metastasis of uveal melanoma millimeter-by-millimeter in 8033 consecutive eyes. *Arch Ophthalmol* **127**, 989-98 (2009).
392. Singh, A.D., Shields, C.L. & Shields, J.A. Prognostic factors in uveal melanoma. *Melanoma Res* **11**, 255-63 (2001).
393. Shields, C.L., Kaliki, S., Furuta, M., Mashayekhi, A. & Shields, J.A. Clinical spectrum and prognosis of uveal melanoma based on age at presentation in 8,033 cases. *Retina* **32**, 1363-72 (2012).
394. Damato, B. *et al.* Cytogenetics of uveal melanoma: a 7-year clinical experience. *Ophthalmology* **114**, 1925-31 (2007).
395. Berus, T. *et al.* Clinical, Histopathological and Cytogenetic Prognosticators in Uveal Melanoma - A Comprehensive Review. *Anticancer Res* **37**, 6541-6549 (2017).
396. Skinner, C.C., Augsburger, J.J., Augsburger, B.D. & Correa, Z.M. Comparison of Alternative Tumor Size Classifications for Posterior Uveal Melanomas. *Invest Ophthalmol Vis Sci* **58**, 3335-3342 (2017).
397. Angi, M. *et al.* Immunohistochemical assessment of mitotic count in uveal melanoma. *Acta Ophthalmol* **89**, e155-60 (2011).
398. Scholes, A.G. *et al.* Monosomy 3 in uveal melanoma: correlation with clinical and histologic predictors of survival. *Invest Ophthalmol Vis Sci* **44**, 1008-11 (2003).

399. Ehlers, J.P., Worley, L., Onken, M.D. & Harbour, J.W. Integrative genomic analysis of aneuploidy in uveal melanoma. *Clin Cancer Res* **14**, 115-22 (2008).
400. White, V.A., Chambers, J.D., Courtright, P.D., Chang, W.Y. & Horsman, D.E. Correlation of cytogenetic abnormalities with the outcome of patients with uveal melanoma. *Cancer* **83**, 354-9 (1998).
401. Kilic, E. *et al.* Concurrent loss of chromosome arm 1p and chromosome 3 predicts a decreased disease-free survival in uveal melanoma patients. *Invest Ophthalmol Vis Sci* **46**, 2253-7 (2005).
402. Kilic, E. *et al.* Clinical and cytogenetic analyses in uveal melanoma. *Invest Ophthalmol Vis Sci* **47**, 3703-7 (2006).
403. Tschentscher, F., Prescher, G., Zeschnigk, M., Horsthemke, B. & Lohmann, D.R. Identification of chromosomes 3, 6, and 8 aberrations in uveal melanoma by microsatellite analysis in comparison to comparative genomic hybridization. *Cancer Genet Cytogenet* **122**, 13-7 (2000).
404. Shields, C.L. *et al.* Personalized Prognosis of Uveal Melanoma Based on Cytogenetic Profile in 1059 Patients over an 8-Year Period: The 2017 Harry S. Gradle Lecture. *Ophthalmology* **124**, 1523-1531 (2017).
405. Versluis, M. *et al.* Digital PCR validates 8q dosage as prognostic tool in uveal melanoma. *PLoS One* **10**, e0116371 (2015).
406. Onken, M.D., Worley, L.A., Tuscan, M.D. & Harbour, J.W. An accurate, clinically feasible multi-gene expression assay for predicting metastasis in uveal melanoma. *J Mol Diagn* **12**, 461-8 (2010).
407. Tschentscher, F. *et al.* Tumor classification based on gene expression profiling shows that uveal melanomas with and without monosomy 3 represent two distinct entities. *Cancer Res* **63**, 2578-84 (2003).
408. Onken, M.D. *et al.* Collaborative Ocular Oncology Group report number 1: prospective validation of a multi-gene prognostic assay in uveal melanoma. *Ophthalmology* **119**, 1596-603 (2012).
409. Harbour, J.W., Augsburger, J.J. & Char, D.H. Gene expression profiling versus TNM classification. *Ophthalmology* **120**, e52-3 (2013).
410. Harbour, J.W. & Chen, R. The DecisionDx-UM Gene Expression Profile Test Provides Risk Stratification and Individualized Patient Care in Uveal Melanoma. *PLoS Curr* **5**(2013).
411. Plasseraud, K.M. *et al.* Clinical Performance and Management Outcomes with the DecisionDx-UM Gene Expression Profile Test in a Prospective Multicenter Study. *J Oncol* **2016**, 5325762 (2016).
412. Onken, M.D., Worley, L.A. & Harbour, J.W. Association between gene expression profile, proliferation and metastasis in uveal melanoma. *Curr Eye Res* **35**, 857-63 (2010).
413. Onken, M.D. *et al.* Functional gene expression analysis uncovers phenotypic switch in aggressive uveal melanomas. *Cancer Res* **66**, 4602-9 (2006).
414. Chang, S.H., Worley, L.A., Onken, M.D. & Harbour, J.W. Prognostic biomarkers in uveal melanoma: evidence for a stem cell-like phenotype associated with metastasis. *Melanoma Res* **18**, 191-200 (2008).
415. Correa, Z.M. Assessing Prognosis in Uveal Melanoma. *Cancer Control* **23**, 93-8 (2016).
416. Shain, A.H. *et al.* The genetic evolution of metastatic uveal melanoma. *Nat Genet* **51**, 1123-1130 (2019).
417. Rodrigues, M. *et al.* Evolutionary Routes in Metastatic Uveal Melanomas Depend on MBD4 Alterations. *Clin Cancer Res* **25**, 5513-5524 (2019).
418. Uner, O.E., See, T.R.O., Szalai, E., Grossniklaus, H.E. & Stalhammar, G. Estimation of the timing of BAP1 mutation in uveal melanoma progression. *Sci Rep* **11**, 8923 (2021).
419. Piaggio, F. *et al.* Secondary Somatic Mutations in G-Protein-Related Pathways and Mutation Signatures in Uveal Melanoma. *Cancers (Basel)* **11**(2019).
420. Anbunathan, H., Verstraten, R., Singh, A.D., Harbour, J.W. & Bowcock, A.M. Integrative Copy Number Analysis of Uveal Melanoma Reveals Novel Candidate Genes Involved in Tumorigenesis Including a Tumor Suppressor Role for PHF10/BAF45a. *Clin Cancer Res* **25**, 5156-5166 (2019).
421. Durante, M.A. *et al.* Single-cell analysis reveals new evolutionary complexity in uveal melanoma. *Nat Commun* **11**, 496 (2020).
422. Pandiani, C. *et al.* Single-cell RNA sequencing reveals intratumoral heterogeneity in primary uveal melanomas and identifies HES6 as a driver of the metastatic disease. *Cell Death Differ* **28**, 1990-2000 (2021).
423. Makitie, T., Summanen, P., Tarkkanen, A. & Kivela, T. Tumor-infiltrating macrophages (CD68(+) cells) and prognosis in malignant uveal melanoma. *Invest Ophthalmol Vis Sci* **42**, 1414-21 (2001).
424. Bronkhorst, I.H. *et al.* Detection of M2-macrophages in uveal melanoma and relation with survival. *Invest Ophthalmol Vis Sci* **52**, 643-50 (2011).

425. Maat, W. *et al.* Monosomy of chromosome 3 and an inflammatory phenotype occur together in uveal melanoma. *Invest Ophthalmol Vis Sci* **49**, 505-10 (2008).
426. Krishna, Y., McCarthy, C., Kalirai, H. & Coupland, S.E. Inflammatory cell infiltrates in advanced metastatic uveal melanoma. *Hum Pathol* **66**, 159-166 (2017).
427. de Waard-Siebinga, I., Creyghton, W.M., Kool, J. & Jager, M.J. Effects of interferon alfa and gamma on human uveal melanoma cells in vitro. *Br J Ophthalmol* **79**, 847-55 (1995).
428. Bronkhorst, I.H. & Jager, M.J. Uveal melanoma: the inflammatory microenvironment. *J Innate Immun* **4**, 454-62 (2012).
429. Brouwer, N.J. *et al.* Tumour Angiogenesis in Uveal Melanoma Is Related to Genetic Evolution. *Cancers (Basel)* **11**(2019).
430. Jager, M.J., Ly, L.V., El Filali, M. & Madigan, M.C. Macrophages in uveal melanoma and in experimental ocular tumor models: Friends or foes? *Prog Retin Eye Res* **30**, 129-46 (2011).
431. Gezgin, G. *et al.* Genetic evolution of uveal melanoma guides the development of an inflammatory microenvironment. *Cancer Immunol Immunother* **66**, 903-912 (2017).
432. Figueiredo, C.R. *et al.* Loss of BAP1 expression is associated with an immunosuppressive microenvironment in uveal melanoma, with implications for immunotherapy development. *J Pathol* **250**, 420-439 (2020).
433. Wierenga, A.P.A., Cao, J., Luyten, G.P.M. & Jager, M.J. Immune Checkpoint Inhibitors in Uveal and Conjunctival Melanoma. *Int Ophthalmol Clin* **59**, 53-63 (2019).
434. Verbik, D.J., Murray, T.G., Tran, J.M. & Ksander, B.R. Melanomas that develop within the eye inhibit lymphocyte proliferation. *Int J Cancer* **73**, 470-8 (1997).
435. Shields, J.A. & Shields, C.L. Management of posterior uveal melanoma: past, present, and future: the 2014 Charles L. Schepens lecture. *Ophthalmology* **122**, 414-28 (2015).
436. Dogrusoz, M., Jager, M.J. & Damato, B. Uveal Melanoma Treatment and Prognostication. *Asia Pac J Ophthalmol (Phila)* **6**, 186-196 (2017).
437. Shields, C.L. *et al.* Iris melanoma management with iodine-125 plaque radiotherapy in 144 patients: impact of melanoma-related glaucoma on outcomes. *Ophthalmology* **120**, 55-61 (2013).
438. Gragoudas, E.S., Lane, A.M., Munzenrider, J., Egan, K.M. & Li, W. Long-term risk of local failure after proton therapy for choroidal/ciliary body melanoma. *Trans Am Ophthalmol Soc* **100**, 43-8; discussion 48-9 (2002).
439. Damato, B. The role of eyewall resection in uveal melanoma management. *Int Ophthalmol Clin* **46**, 81-93 (2006).
440. Damato, B. & Heimann, H. Personalized treatment of uveal melanoma. *Eye (Lond)* **27**, 172-9 (2013).
441. Piperno-Neumann, S. *et al.* A randomized multicenter phase 3 trial of adjuvant fotemustine versus surveillance in high risk uveal melanoma (UM) patients (FOTEADJ). *Journal of Clinical Oncology* **35**, 9502-9502 (2017).
442. Rodriguez-Vidal, C. *et al.* Treatment of Metastatic Uveal Melanoma: Systematic Review. *Cancers (Basel)* **12**(2020).
443. Carvajal, R.D. *et al.* Metastatic disease from uveal melanoma: treatment options and future prospects. *Br J Ophthalmol* **101**, 38-44 (2017).
444. Bhatia, S. *et al.* Phase II trial of sorafenib in combination with carboplatin and paclitaxel in patients with metastatic uveal melanoma: SWOG S0512. *PLoS One* **7**, e48787 (2012).
445. Homsí, J. *et al.* Phase 2 open-label study of weekly docosahexaenoic acid-paclitaxel in patients with metastatic uveal melanoma. *Melanoma Res* **20**, 507-10 (2010).
446. Schmitt, A. *et al.* A two-cohort phase II clinical trial of gemcitabine plus treosulfan in patients with metastatic uveal melanoma. *Melanoma Res* **15**, 447-51 (2005).
447. Kivela, T. *et al.* Bleomycin, vincristine, lomustine and dacarbazine (BOLD) in combination with recombinant interferon alpha-2b for metastatic uveal melanoma. *Eur J Cancer* **39**, 1115-20 (2003).
448. Pons, F. *et al.* Metastatic uveal melanoma: is there a role for conventional chemotherapy? - A single center study based on 58 patients. *Melanoma Res* **21**, 217-22 (2011).
449. Pfohler, C. *et al.* Treosulfan and gemcitabine in metastatic uveal melanoma patients: results of a multicenter feasibility study. *Anticancer Drugs* **14**, 337-40 (2003).
450. Schmitt, A. *et al.* Phase II trial of cisplatin, gemcitabine and treosulfan in patients with metastatic uveal melanoma. *Melanoma Res* **15**, 205-7 (2005).
451. Leyvraz, S. *et al.* Hepatic intra-arterial versus intravenous fotemustine in patients with liver metastases from uveal melanoma (EORTC 18021): a multicentric randomized trial. *Ann Oncol* **25**, 742-6 (2014).

452. Hughes, M.S. *et al.* Results of a Randomized Controlled Multicenter Phase III Trial of Percutaneous Hepatic Perfusion Compared with Best Available Care for Patients with Melanoma Liver Metastases. *Ann Surg Oncol* **23**, 1309-19 (2016).
453. Huppert, P.E. *et al.* Transarterial chemoembolization of liver metastases in patients with uveal melanoma. *Eur J Radiol* **74**, e38-44 (2010).
454. Fiorentini, G. *et al.* Intra-arterial hepatic chemoembolization (TACE) of liver metastases from ocular melanoma with slow-release irinotecan-eluting beads. Early results of a phase II clinical study. *In Vivo* **23**, 131-7 (2009).
455. Frenkel, S. *et al.* Long-term survival of uveal melanoma patients after surgery for liver metastases. *Br J Ophthalmol* **93**, 1042-6 (2009).
456. Mariani, P. *et al.* Surgical management of liver metastases from uveal melanoma: 16 years' experience at the Institut Curie. *Eur J Surg Oncol* **35**, 1192-7 (2009).
457. Wu, X., Zhu, M., Fletcher, J.A., Giobbie-Hurder, A. & Hodi, F.S. The protein kinase C inhibitor enzastaurin exhibits antitumor activity against uveal melanoma. *PLoS One* **7**, e29622 (2012).
458. Heijkants, R.C., Nieveen, M., Hart, K.C., Teunisse, A. & Jochemsen, A.G. Targeting MDMX and PKCdelta to improve current uveal melanoma therapeutic strategies. *Oncogenesis* **7**, 33 (2018).
459. Nathan, P., Needham, A., Corrie, P.G. *et al.* SELPAC: A 3 arm randomised phase II study of the MEK inhibitor selumetinib alone or in combination with paclitaxel (PT) in metastatic uveal melanoma (UM). *Annals of oncology* **30**(2019).
460. Piperno-Neumann, S. *et al.* Phase I dose-escalation study of the protein kinase C (PKC) inhibitor AEB071 in patients with metastatic uveal melanoma. *Journal of Clinical Oncology* **32**, 9030-9030 (2014).
461. Onken, M.D. *et al.* Targeting nucleotide exchange to inhibit constitutively active G protein alpha subunits in cancer cells. *Sci Signal* **11**(2018).
462. Chua, V. *et al.* Stromal fibroblast growth factor 2 reduces the efficacy of bromodomain inhibitors in uveal melanoma. *EMBO Mol Med* **11**(2019).
463. Khattak, M.A., Fisher, R., Hughes, P., Gore, M. & Larkin, J. Ipilimumab activity in advanced uveal melanoma. *Melanoma Res* **23**, 79-81 (2013).
464. Wolchok, J.D. *et al.* Four-year survival rates for patients with metastatic melanoma who received ipilimumab in phase II clinical trials. *Ann Oncol* **24**, 2174-80 (2013).
465. Saint-Ghislain, M., Geoffrois L., Gastaud L., Lesimple L., *et al.* Immune checkpoint inhibitors in a cohort of 206 metastatic uveal melanomas patients. Poster presentation. *Annals of Oncology* (2019).
466. Chandran, S.S. *et al.* Treatment of metastatic uveal melanoma with adoptive transfer of tumour-infiltrating lymphocytes: a single-centre, two-stage, single-arm, phase 2 study. *Lancet Oncol* (2017).
467. Smith, J.R. *et al.* Proceedings of the Association for Research in Vision and Ophthalmology and Champalimaud Foundation Ocular Oncogenesis and Oncology Conference. *Transl Vis Sci Technol* **8**, 9 (2019).
468. TCR Bispecific Boosts Survival in Uveal Melanoma. *Cancer Discov* (2021).
469. Middleton, M.R. *et al.* Tebentafusp, A TCR/Anti-CD3 Bispecific Fusion Protein Targeting gp100, Potently Activated Antitumor Immune Responses in Patients with Metastatic Melanoma. *Clin Cancer Res* **26**, 5869-5878 (2020).
470. Damato, B.E., Dukes, J., Goodall, H. & Carvajal, R.D. Tebentafusp: T Cell Redirection for the Treatment of Metastatic Uveal Melanoma. *Cancers (Basel)* **11**(2019).
471. Aronow, M.E., Topham, A.K. & Singh, A.D. Uveal Melanoma: 5-Year Update on Incidence, Treatment, and Survival (SEER 1973-2013). *Ocul Oncol Pathol* **4**, 145-151 (2018).
472. Xu, Y. *et al.* Epidemiological Study of Uveal Melanoma from US Surveillance, Epidemiology, and End Results Program (2010-2015). *J Ophthalmol* **2020**, 3614039 (2020).
473. Krantz, B.A., Dave, N., Komatsubara, K.M., Marr, B.P. & Carvajal, R.D. Uveal melanoma: epidemiology, etiology, and treatment of primary disease. *Clin Ophthalmol* **11**, 279-289 (2017).
474. Al-Jamal, R.T. *et al.* The Pediatric Choroidal and Ciliary Body Melanoma Study: A Survey by the European Ophthalmic Oncology Group. *Ophthalmology* **123**, 898-907 (2016).
475. Kujala, E., Makitie, T. & Kivela, T. Very long-term prognosis of patients with malignant uveal melanoma. *Invest Ophthalmol Vis Sci* **44**, 4651-9 (2003).
476. Singh, M., Durairaj, P. & Yeung, J. Uveal Melanoma: A Review of the Literature. *Oncol Ther* **6**, 87-104 (2018).
477. Park, S.J. *et al.* Nationwide Incidence of Ocular Melanoma in South Korea by Using the National Cancer Registry Database (1999-2011). *Invest Ophthalmol Vis Sci* **56**, 4719-24 (2015).

478. Tomizuka, T., Namikawa, K. & Higashi, T. Characteristics of melanoma in Japan: a nationwide registry analysis 2011-2013. *Melanoma Res* **27**, 492-497 (2017).
479. Virgili, G. *et al.* Incidence of uveal melanoma in Europe. *Ophthalmology* **114**, 2309-15 (2007).
480. Baily, C. *et al.* Uveal Melanoma in Ireland. *Ocul Oncol Pathol* **5**, 195-204 (2019).
481. Yonekawa, Y. & Kim, I.K. Epidemiology and management of uveal melanoma. *Hematol Oncol Clin North Am* **26**, 1169-84 (2012).
482. Shields, C.L. *et al.* Prognosis of uveal melanoma based on race in 8100 patients: The 2015 Doyne Lecture. *Eye (Lond)* **29**, 1027-35 (2015).
483. Weis, E., Shah, C.P., Lajous, M., Shields, J.A. & Shields, C.L. The association between host susceptibility factors and uveal melanoma: a meta-analysis. *Arch Ophthalmol* **124**, 54-60 (2006).
484. Chien, J.L., Sioufi, K., Surakiatchanukul, T., Shields, J.A. & Shields, C.L. Choroidal nevus: a review of prevalence, features, genetics, risks, and outcomes. *Curr Opin Ophthalmol* **28**, 228-237 (2017).
485. Singh, A.D. & Borden, E.C. Metastatic uveal melanoma. *Ophthalmol Clin North Am* **18**, 143-50, ix (2005).
486. Kivela, T. & Eskelin, S. Transformation of nevus to melanoma. *Ophthalmology* **113**, 887-8 e1 (2006).
487. Shields, C.L. *et al.* Association of ocular and oculodermal melanocytosis with the rate of uveal melanoma metastasis: analysis of 7872 consecutive eyes. *JAMA Ophthalmol* **131**, 993-1003 (2013).
488. Konstantinov, N.K., Berry, T.M., Elwood, H.R. & Zlotoff, B.J. Nevus of Ota associated with a primary uveal melanoma and intracranial melanoma metastasis. *Cutis* **102**, E2-E4 (2018).
489. Singh, A.D. *et al.* Lifetime prevalence of uveal melanoma in white patients with oculo(dermal) melanocytosis. *Ophthalmology* **105**, 195-8 (1998).
490. Vivancos, A. *et al.* Genetic evolution of nevus of Ota reveals clonal heterogeneity acquiring BAP1 and TP53 mutations. *Pigment Cell Melanoma Res* **29**, 247-53 (2016).
491. Thomas, A.C. *et al.* Mosaic Activating Mutations in GNA11 and GNAQ Are Associated with Phakomatosis Pigmentovascularis and Extensive Dermal Melanocytosis. *J Invest Dermatol* **136**, 770-778 (2016).
492. Nayman, T., Bostan, C., Logan, P. & Burnier, M.N., Jr. Uveal Melanoma Risk Factors: A Systematic Review of Meta-Analyses. *Curr Eye Res* **42**, 1085-1093 (2017).
493. Holly, E.A., Aston, D.A., Ahn, D.K. & Smith, A.H. Intraocular melanoma linked to occupations and chemical exposures. *Epidemiology* **7**, 55-61 (1996).
494. Weis, E., Shah, C.P., Lajous, M., Shields, J.A. & Shields, C.L. The association of cutaneous and iris nevi with uveal melanoma: a meta-analysis. *Ophthalmology* **116**, 536-543 e2 (2009).
495. Schwab, C. *et al.* New insights into oculodermal nevogenesis and proposal for a new iris nevus classification. *Br J Ophthalmol* **99**, 644-9 (2015).
496. Singh, A.D. *et al.* Iris melanoma in a ten-year-old boy with familial atypical mole-melanoma (FAM-M) syndrome. *Ophthalmic Genet* **15**, 145-9 (1994).
497. Shields, C.L. *et al.* Uveal melanoma in teenagers and children. A report of 40 cases. *Ophthalmology* **98**, 1662-6 (1991).
498. Jay, M. & McCartney, A.C. Familial malignant melanoma of the uvea and p53: a Victorian detective story. *Surv Ophthalmol* **37**, 457-62 (1993).
499. Singh, A.D. *et al.* Familial uveal melanoma. Clinical observations on 56 patients. *Arch Ophthalmol* **114**, 392-9 (1996).
500. Van Ruymbeke, K. & Leys, A. Malignant melanoma of skin and eye in the same family: more than a coincidence? *Bull Soc Belge Ophtalmol* **231**, 149-55 (1989).
501. van Hees, C.L., Jager, M.J., Bleeker, J.C., Kemme, H. & Bergman, W. Occurrence of cutaneous and uveal melanoma in patients with uveal melanoma and their first degree relatives. *Melanoma Res* **8**, 175-80 (1998).
502. Singh, A.D. *et al.* Familial uveal melanoma, III. Is the occurrence of familial uveal melanoma coincidental? *Arch Ophthalmol* **114**, 1101-4 (1996).
503. Abdel-Rahman, M.H., Pilarski, R., Ezzat, S., Sexton, J. & Davidorf, F.H. Cancer family history characterization in an unselected cohort of 121 patients with uveal melanoma. *Fam Cancer* **9**, 431-8 (2010).
504. Rednam, K.R., Jampol, L.M., Levine, R.A. & Goldberg, M.F. Uveal melanoma in association with multiple malignancies. A case report and review. *Retina* **1**, 100-6 (1981).
505. Diener-West, M. *et al.* Second primary cancers after enrollment in the COMS trials for treatment of choroidal melanoma: COMS Report No. 25. *Arch Ophthalmol* **123**, 601-4 (2005).

506. Rai, K. *et al.* Germline BAP1 alterations in familial uveal melanoma. *Genes Chromosomes Cancer* **56**, 168-174 (2017).
507. Bergman, L., Nilsson, B., Ragnarsson-Olding, B. & Seregard, S. Uveal melanoma: a study on incidence of additional cancers in the Swedish population. *Invest Ophthalmol Vis Sci* **47**, 72-7 (2006).
508. Hemminki, K. & Jiang, Y. Association of ocular melanoma with breast cancer but not with cutaneous melanoma: results from the Swedish Family-Cancer Database. *Int J Cancer* **94**, 907-9 (2001).
509. Scelo, G. *et al.* Associations between ocular melanoma and other primary cancers: an international population-based study. *Int J Cancer* **120**, 152-9 (2007).
510. Singh, A.D., Shields, C.L., Shields, J.A. & Sato, T. Uveal melanoma in young patients. *Arch Ophthalmol* **118**, 918-23 (2000).
511. van Hees, C.L. *et al.* Are atypical nevi a risk factor for uveal melanoma? A case-control study. *J Invest Dermatol* **103**, 202-5 (1994).
512. Helgadottir, H. & Hoium, V. The genetics of uveal melanoma: current insights. *Appl Clin Genet* **9**, 147-55 (2016).
513. Rai, K., Pilarski, R., Cebulla, C.M. & Abdel-Rahman, M.H. Comprehensive review of BAP1 tumor predisposition syndrome with report of two new cases. *Clin Genet* **89**, 285-94 (2016).
514. Yu, M.D., Masoomian, B., Shields, J.A. & Shields, C.L. BAP1 Germline Mutation Associated with Bilateral Primary Uveal Melanoma. *Ocul Oncol Pathol* **6**, 10-14 (2020).
515. Strande, N.T. *et al.* Evaluating the Clinical Validity of Gene-Disease Associations: An Evidence-Based Framework Developed by the Clinical Genome Resource. *Am J Hum Genet* **100**, 895-906 (2017).
516. Aoude, L.G. *et al.* Prevalence of Germline BAP1, CDKN2A, and CDK4 Mutations in an Australian Population-Based Sample of Cutaneous Melanoma Cases. *Twin Res Hum Genet* **18**, 126-33 (2015).
517. Nathan, V. *et al.* Loss-of-function variants in POT1 predispose to uveal melanoma. *J Med Genet* **58**, 234-236 (2021).
518. Abdel-Rahman, M.H. *et al.* Whole exome sequencing identifies candidate genes associated with hereditary predisposition to uveal melanoma. *Ophthalmology* (2019).
519. Lobo, J. *et al.* Ovarian metastasis from uveal melanoma with MLH1/PMS2 protein loss in a patient with germline MLH1 mutated Lynch syndrome: consequence or coincidence? *Virchows Arch* **470**, 347-352 (2017).
520. Kannengiesser, C. *et al.* CDKN2A as a uveal and cutaneous melanoma susceptibility gene. *Genes Chromosomes Cancer* **38**, 265-8 (2003).
521. Baglietto, L. *et al.* Risks of Lynch syndrome cancers for MSH6 mutation carriers. *J Natl Cancer Inst* **102**, 193-201 (2010).
522. Cruz, C., Teule, A., Caminal, J.M., Blanco, I. & Piulats, J.M. Uveal melanoma and BRCA1/BRCA2 genes: a relationship that needs further investigation. *J Clin Oncol* **29**, e827-9 (2011).
523. Toomey, C.B. *et al.* Prevalence of Mismatch Repair Gene Mutations in Uveal Melanoma. *Retina* **40**, 2216-2220 (2020).
524. Iscovich, J. *et al.* Prevalence of the BRCA2 6174 del T mutation in Israeli uveal melanoma patients. *Int J Cancer* **98**, 42-4 (2002).
525. Sinilnikova, O.M. *et al.* Germline brca2 sequence variants in patients with ocular melanoma. *Int J Cancer* **82**, 325-8 (1999).
526. Mobuchon, L. *et al.* A GWAS in uveal melanoma identifies risk polymorphisms in the CLPTM1L locus. *NPJ Genom Med* **2**(2017).
527. Wang, Z. *et al.* Imputation and subset-based association analysis across different cancer types identifies multiple independent risk loci in the TERT-CLPTM1L region on chromosome 5p15.33. *Hum Mol Genet* **23**, 6616-33 (2014).
528. Rafnar, T. *et al.* Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. *Nat Genet* **41**, 221-7 (2009).
529. Chen, H.M., A.; *et al.* Large-scale cross-cancer fine-mapping of the 5p15.33 region reveals multiple independent signals. *Human Genetics and Genomics Advances* **2**(2021).
530. Schumacher, F.R. *et al.* Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat Genet* **50**, 928-936 (2018).
531. Lesueur, C. *et al.* Genome-wide association analyses identify new susceptibility loci for oral cavity and pharyngeal cancer. *Nat Genet* **48**, 1544-1550 (2016).
532. Gharahkhani, P. *et al.* Genome-wide association studies in oesophageal adenocarcinoma and Barrett's oesophagus: a large-scale meta-analysis. *Lancet Oncol* **17**, 1363-1373 (2016).

533. O'Mara, T.A. *et al.* Identification of nine new susceptibility loci for endometrial cancer. *Nat Commun* **9**, 3166 (2018).
534. Chung, C.C. *et al.* Meta-analysis identifies four new loci associated with testicular germ cell tumor. *Nat Genet* **45**, 680-5 (2013).
535. Law, M.H. *et al.* Genome-wide meta-analysis identifies five new susceptibility loci for cutaneous malignant melanoma. *Nat Genet* **47**, 987-95 (2015).
536. Melin, B.S. *et al.* Genome-wide association study of glioma subtypes identifies specific differences in genetic susceptibility to glioblastoma and non-glioblastoma tumors. *Nat Genet* **49**, 789-794 (2017).
537. Yamamoto, K., Okamoto, A., Isonishi, S., Ochiai, K. & Ohtake, Y. A novel gene, CRR9, which was up-regulated in CDDP-resistant ovarian tumor cell line, was associated with apoptosis. *Biochem Biophys Res Commun* **280**, 1148-54 (2001).
538. Parashar, D. *et al.* Targeted biologic inhibition of both tumor cell-intrinsic and intercellular CLPTM1L/CRR9-mediated chemotherapeutic drug resistance. *NPJ Precis Oncol* **5**, 16 (2021).
539. Kang, J.U., Koo, S.H., Kwon, K.C., Park, J.W. & Kim, J.M. Gain at chromosomal region 5p15.33, containing TERT, is the most frequent genetic event in early stages of non-small cell lung cancer. *Cancer Genet Cytogenet* **182**, 1-11 (2008).
540. Vazquez-Mena, O. *et al.* Amplified genes may be overexpressed, unchanged, or downregulated in cervical cancer cell lines. *PLoS One* **7**, e32667 (2012).
541. James, M.A., Vikis, H.G., Tate, E., Rymaszewski, A.L. & You, M. CRR9/CLPTM1L regulates cell survival signaling and is required for Ras transformation and lung tumorigenesis. *Cancer Res* **74**, 1116-27 (2014).
542. Li, H. *et al.* CLPTM1L induces estrogen receptor beta signaling-mediated radioresistance in non-small cell lung cancer cells. *Cell Commun Signal* **18**, 152 (2020).
543. Jia, J. *et al.* CLPTM1L promotes growth and enhances aneuploidy in pancreatic cancer cells. *Cancer Res* **74**, 2785-95 (2014).
544. Clarke, W.R., Amundadottir, L. & James, M.A. CLPTM1L/CRR9 ectodomain interaction with GRP78 at the cell surface signals for survival and chemoresistance upon ER stress in pancreatic adenocarcinoma cells. *Int J Cancer* **144**, 1367-1378 (2019).
545. Maciejowski, J. & de Lange, T. Telomeres in cancer: tumour suppression and genome instability. *Nat Rev Mol Cell Biol* **18**, 175-186 (2017).
546. Huang, F.W. *et al.* Highly recurrent TERT promoter mutations in human melanoma. *Science* **339**, 957-9 (2013).
547. Griewank, K.G. *et al.* TERT promoter mutations in ocular melanoma distinguish between conjunctival and uveal tumours. *Br J Cancer* **109**, 497-501 (2013).
548. Dono, M. *et al.* Mutation frequencies of GNAQ, GNA11, BAP1, SF3B1, EIF1AX and TERT in uveal melanoma: detection of an activating mutation in the TERT gene promoter in a single case of uveal melanoma. *Br J Cancer* **110**, 1058-65 (2014).
549. Koopmans, A.E. *et al.* Prevalence and implications of TERT promoter mutation in uveal and conjunctival melanoma and in benign and premalignant conjunctival melanocytic lesions. *Invest Ophthalmol Vis Sci* **55**, 6024-30 (2014).
550. Amos, C.I. *et al.* Genome-wide association study identifies novel loci predisposing to cutaneous melanoma. *Hum Mol Genet* **20**, 5012-23 (2011).
551. Ferguson, R. *et al.* Genetic markers of pigmentation are novel risk loci for uveal melanoma. *Sci Rep* **6**, 31191 (2016).
552. Wei, L. *et al.* Variants at the OCA2/HERC2 locus affect time to first cutaneous squamous cell carcinoma in solid organ transplant recipients collected using two different study designs. *Br J Dermatol* **177**, 1066-1073 (2017).
553. Rodrigues, M. *et al.* Outlier response to anti-PD1 in uveal melanoma reveals germline MBD4 mutations in hypermutated tumors. *Nat Commun* **9**, 1866 (2018).
554. Beard, W.A., Horton, J.K., Prasad, R. & Wilson, S.H. Eukaryotic Base Excision Repair: New Approaches Shine Light on Mechanism. *Annu Rev Biochem* **88**, 137-162 (2019).
555. Dizdaroglu, M. Oxidatively induced DNA damage and its repair in cancer. *Mutat Res Rev Mutat Res* **763**, 212-45 (2015).
556. Grundy, G.J. & Parsons, J.L. Base excision repair and its implications to cancer therapy. *Essays Biochem* **64**, 831-843 (2020).
557. Fromme, J.C., Banerjee, A. & Verdine, G.L. DNA glycosylase recognition and catalysis. *Curr Opin Struct Biol* **14**, 43-9 (2004).

558. Chastain, P.D., 2nd *et al.* Abasic sites preferentially form at regions undergoing DNA replication. *FASEB J* **24**, 3674-80 (2010).
559. Robson, C.N. & Hickson, I.D. Isolation of cDNA clones encoding a human apurinic/aprimidinic endonuclease that corrects DNA repair and mutagenesis defects in *E. coli* xth (exonuclease III) mutants. *Nucleic Acids Res* **19**, 5519-23 (1991).
560. Braithwaite, E.K. *et al.* DNA polymerase lambda mediates a back-up base excision repair activity in extracts of mouse embryonic fibroblasts. *J Biol Chem* **280**, 18469-75 (2005).
561. Zharkov, D.O. Base excision DNA repair. *Cell Mol Life Sci* **65**, 1544-65 (2008).
562. Fortini, P. & Dogliotti, E. Base damage and single-strand break repair: mechanisms and functional significance of short- and long-patch repair subpathways. *DNA Repair (Amst)* **6**, 398-409 (2007).
563. Chen, X. *et al.* Distinct kinetics of human DNA ligases I, IIIalpha, IIIbeta, and IV reveal direct DNA sensing ability and differential physiological functions in DNA repair. *DNA Repair (Amst)* **8**, 961-8 (2009).
564. Thayer, M.M., Ahern, H., Xing, D., Cunningham, R.P. & Tainer, J.A. Novel DNA binding motifs in the DNA repair enzyme endonuclease III crystal structure. *EMBO J* **14**, 4108-20 (1995).
565. Doublet, S., Bandaru, V., Bond, J.P. & Wallace, S.S. The crystal structure of human endonuclease VIII-like 1 (NEIL1) reveals a zincless finger motif required for glycosylase activity. *Proc Natl Acad Sci U S A* **101**, 10284-9 (2004).
566. Bandaru, V., Sunkara, S., Wallace, S.S. & Bond, J.P. A novel human DNA glycosylase that removes oxidative DNA damage and is homologous to *Escherichia coli* endonuclease VIII. *DNA Repair (Amst)* **1**, 517-29 (2002).
567. Hans, F., Senariso, M., Bhaskar Naidu, C. & Timmins, J. Focus on DNA Glycosylases-A Set of Tightly Regulated Enzymes with a High Potential as Anticancer Drug Targets. *Int J Mol Sci* **21**(2020).
568. Hardeland, U., Steinacher, R., Jiricny, J. & Schar, P. Modification of the human thymine-DNA glycosylase by ubiquitin-like proteins facilitates enzymatic turnover. *EMBO J* **21**, 1456-64 (2002).
569. Sjolund, A.B., Senejani, A.G. & Sweasy, J.B. MBD4 and TDG: Multifaceted DNA glycosylases with ever expanding biological roles. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **743-744**, 12-25 (2013).
570. Wu, P. *et al.* Mismatch Repair in Methylated DNA. *Journal of Biological Chemistry* **278**, 5285-5291 (2003).
571. Petronzelli, F. *et al.* Biphasic kinetics of the human DNA repair protein MED1 (MBD4), a mismatch-specific DNA N-glycosylase. *J Biol Chem* **275**, 32422-9 (2000).
572. Petronzelli, F. *et al.* Investigation of the substrate spectrum of the human mismatch-specific DNA N-glycosylase MED1 (MBD4): fundamental role of the catalytic domain. *J Cell Physiol* **185**, 473-80 (2000).
573. Hendrich, B., Hardeland, U., Ng, H.H., Jiricny, J. & Bird, A. The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. *Nature* **401**, 301-4 (1999).
574. Bellacosa, A. *et al.* MED1, a novel human methyl-CpG-binding endonuclease, interacts with DNA mismatch repair protein MLH1. *Proc Natl Acad Sci U S A* **96**, 3969-74 (1999).
575. Bellacosa, A. & Drohat, A.C. Role of base excision repair in maintaining the genetic and epigenetic integrity of CpG sites. *DNA Repair (Amst)* **32**, 33-42 (2015).
576. Du, Q., Luu, P.L., Stirzaker, C. & Clark, S.J. Methyl-CpG-binding domain proteins: readers of the epigenome. *Epigenomics* **7**, 1051-73 (2015).
577. Allis, C.D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat Rev Genet* **17**, 487-500 (2016).
578. Ehrlich, M. *et al.* Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *Nucleic Acids Res* **10**, 2709-21 (1982).
579. Bird, A.P. CpG-rich islands and the function of DNA methylation. *Nature* **321**, 209-13 (1986).
580. Cooper, D.N. & Youssoufian, H. The CpG dinucleotide and human genetic disease. *Hum Genet* **78**, 151-5 (1988).
581. Schmutte, C. & Jones, P.A. Involvement of DNA methylation in human carcinogenesis. *Biol Chem* **379**, 377-88 (1998).
582. Ouzon-Shubeita, H., Jung, H., Lee, M.H., Koag, M.C. & Lee, S. Catalytic mechanism of the mismatch-specific DNA glycosylase methyl-CpG-binding domain 4. *Biochem J* **477**, 1601-1612 (2020).
583. Zhang, W. *et al.* Crystal structure of the mismatch-specific thymine glycosylase domain of human methyl-CpG-binding protein MBD4. *Biochem Biophys Res Commun* **412**, 425-8 (2011).
584. Manvilla, B.A., Maiti, A., Begley, M.C., Toth, E.A. & Drohat, A.C. Crystal structure of human methyl-binding domain IV glycosylase bound to abasic DNA. *J Mol Biol* **420**, 164-75 (2012).

585. Hendrich, B. & Bird, A. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol Cell Biol* **18**, 6538-47 (1998).
586. Otani, J. *et al.* Structural basis of the versatile DNA recognition ability of the methyl-CpG binding domain of methyl-CpG binding domain protein 4. *J Biol Chem* **288**, 6351-62 (2013).
587. Wakefield, R.I. *et al.* The solution structure of the domain from MeCP2 that binds to methylated DNA. *J Mol Biol* **291**, 1055-65 (1999).
588. Meng, H., Harrison, D.J. & Meehan, R.R. MBD4 interacts with and recruits USP7 to heterochromatic foci. *J Cell Biochem* **116**, 476-85 (2015).
589. Millar, C.B. *et al.* Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. *Science* **297**, 403-5 (2002).
590. Wong, E. *et al.* Mbd4 inactivation increases Cright-arrowT transition mutations and promotes gastrointestinal tumor formation. *Proc Natl Acad Sci U S A* **99**, 14937-42 (2002).
591. Prolla, T.A. *et al.* Tumour susceptibility and spontaneous mutation in mice deficient in Mlh1, Pms1 and Pms2 DNA mismatch repair. *Nat Genet* **18**, 276-9 (1998).
592. Sansom, O.J., Bishop, S.M., Bird, A. & Clarke, A.R. MBD4 deficiency does not increase mutation or accelerate tumorigenesis in mice lacking MMR. *Oncogene* **23**, 5693-6 (2004).
593. Tricarico, R. *et al.* Involvement of MBD4 inactivation in mismatch repair-deficient tumorigenesis. *Oncotarget* **6**, 42892-904 (2015).
594. Bader, S. *et al.* Somatic frameshift mutations in the MBD4 gene of sporadic colon cancers with mismatch repair deficiency. *Oncogene* **18**, 8044-7 (1999).
595. Riccio, A. *et al.* The DNA repair gene MBD4 (MED1) is mutated in human carcinomas with microsatellite instability. *Nat Genet* **23**, 266-8 (1999).
596. Yamada, T. *et al.* Frameshift mutations in the MBD4/MED1 gene in primary gastric cancer with high-frequency microsatellite instability. *Cancer Lett* **181**, 115-20 (2002).
597. Evertson, S. *et al.* Microsatellite instability and MBD4 mutation in unselected colorectal cancer. *Anticancer Res* **23**, 3569-74 (2003).
598. Pinto, M. *et al.* MBD4 mutations are rare in gastric carcinomas with microsatellite instability. *Cancer Genet Cytogenet* **145**, 103-7 (2003).
599. Boland, M.J. & Christman, J.K. Characterization of Dnmt3b:thymine-DNA glycosylase interaction and stimulation of thymine glycosylase-mediated repair by DNA methyltransferase(s) and RNA. *J Mol Biol* **379**, 492-504 (2008).
600. Ruzov, A. *et al.* MBD4 and MLH1 are required for apoptotic induction in xDNMT1-depleted embryos. *Development* **136**, 2277-86 (2009).
601. Laget, S. *et al.* MBD4 cooperates with DNMT1 to mediate methyl-DNA repression and protects mammalian cells from oxidative stress. *Epigenetics* **9**, 546-56 (2014).
602. Jin, B. & Robertson, K.D. DNA methyltransferases, DNA damage repair, and cancer. *Adv Exp Med Biol* **754**, 3-29 (2013).
603. Cortellino, S. *et al.* The base excision repair enzyme MED1 mediates DNA damage response to antitumor drugs and is associated with mismatch repair system integrity. *Proc Natl Acad Sci U S A* **100**, 15071-6 (2003).
604. Aebi, S. *et al.* Resistance to cytotoxic drugs in DNA mismatch repair-deficient cells. *Clin Cancer Res* **3**, 1763-7 (1997).
605. Hardman, R.A., Afshari, C.A. & Barrett, J.C. Involvement of mammalian MLH1 in the apoptotic response to peroxide-induced oxidative stress. *Cancer Res* **61**, 1392-7 (2001).
606. Sansom, O.J. *et al.* MBD4 deficiency reduces the apoptotic response to DNA-damaging agents in the murine small intestine. *Oncogene* **22**, 7130-6 (2003).
607. Screatton, R.A. *et al.* Fas-associated death domain protein interacts with methyl-CpG binding domain protein 4: a potential link between genome surveillance and apoptosis. *Proc Natl Acad Sci U S A* **100**, 5211-6 (2003).
608. Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* **33 Suppl**, 245-54 (2003).
609. Kondo, E., Gu, Z., Horii, A. & Fukushima, S. The thymine DNA glycosylase MBD4 represses transcription and is associated with methylated p16(INK4a) and hMLH1 genes. *Mol Cell Biol* **25**, 4388-96 (2005).
610. Papait, R. *et al.* Np95 is implicated in pericentromeric heterochromatin replication and in major satellite silencing. *Mol Biol Cell* **18**, 1098-106 (2007).
611. Papait, R. *et al.* The PHD domain of Np95 (mUHRF1) is involved in large-scale reorganization of pericentromeric heterochromatin. *Mol Biol Cell* **19**, 3554-63 (2008).

612. Rai, K. *et al.* DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell* **135**, 1201-12 (2008).
613. Shimoda, N. *et al.* No evidence for AID/MBD4-coupled DNA demethylation in zebrafish embryos. *PLoS One* **9**, e114816 (2014).
614. Kim, M.S. *et al.* DNA demethylation in hormone-induced transcriptional derepression. *Nature* **461**, 1007-12 (2009).
615. Kim, M.S. *et al.* Retraction: DNA demethylation in hormone-induced transcriptional derepression. *Nature* **486**, 280 (2012).
616. Sabag, O. *et al.* Establishment of methylation patterns in ES cells. *Nat Struct Mol Biol* **21**, 110-2 (2014).
617. Nabel, C.S. *et al.* AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat Chem Biol* **8**, 751-8 (2012).
618. Rangam, G., Schmitz, K.M., Cobb, A.J. & Petersen-Mahrt, S.K. AID enzymatic activity is inversely proportional to the size of cytosine C5 orbital cloud. *PLoS One* **7**, e43279 (2012).
619. Kohli, R.M. & Zhang, Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* **502**, 472-9 (2013).
620. Song, J.H. *et al.* The Glu346Lys polymorphism and frameshift mutations of the Methyl-CpG Binding Domain 4 gene in gastrointestinal cancer. *Neoplasma* **56**, 343-7 (2009).
621. Menoyo, A. *et al.* Somatic mutations in the DNA damage-response genes ATR and CHK1 in sporadic stomach tumors with microsatellite instability. *Cancer Res* **61**, 7727-30 (2001).
622. Bader, S.A., Walker, M. & Harrison, D.J. A human cancer-associated truncation of MBD4 causes dominant negative impairment of DNA repair in colon cancer cells. *Br J Cancer* **96**, 660-6 (2007).
623. Sanders, M.A. *et al.* MBD4 guards against methylation damage and germ line deficiency predisposes to clonal hematopoiesis and early-onset AML. *Blood* **132**, 1526-1534 (2018).
624. Johansson, P.A. *et al.* Prolonged stable disease in a uveal melanoma patient with germline MBD4 nonsense mutation treated with pembrolizumab and ipilimumab. *Immunogenetics* **71**, 433-436 (2019).
625. Shen, S.Y. *et al.* Sensitive tumour detection and classification using plasma cell-free DNA methylomes. *Nature* **563**, 579-583 (2018).
626. Tanakaya, K. *et al.* A germline MBD4 mutation was identified in a patient with colorectal oligopolyposis and early-onset cancer: A case report. *Oncology Reports* **42**, 1133-1140 (2019).
627. Waszak, S.M. *et al.* Germline determinants of the somatic mutation landscape in 2,642 cancer genomes. *bioRxiv*, 208330 (2017).
628. Davies, H.R. *et al.* Epigenetic dysregulation underpins tumorigenesis in a cutaneous tumor syndrome. *bioRxiv*, 687459 (2019).
629. Wang, L., Birch, N.W., Zhao Z. *et al.* Epigenetic targeted therapy of stabilized BAP1 in ASXL1 gain-of-function mutated leukemia. *Nature Cancer* **2**, 515-526 (2021).
630. Hagman, J. Critical Functions of IRF4 in B and T Lymphocytes. *J Immunol* **199**, 3715-3716 (2017).
631. Praetorius, C. *et al.* A polymorphism in IRF4 affects human pigmentation through a tyrosinase-dependent MITF/TFAP2A pathway. *Cell* **155**, 1022-33 (2013).
632. Pelloso, M.C. *et al.* Effects of the melanin precursor 5,6-dihydroxy-indole-2-carboxylic acid (DHICA) on DNA damage and repair in the presence of reactive oxygen species. *Arch Biochem Biophys* **557**, 55-64 (2014).
633. Mitra, D. *et al.* An ultraviolet-radiation-independent pathway to melanoma carcinogenesis in the red hair/fair skin background. *Nature* **491**, 449-53 (2012).
634. Ge, M. *et al.* Functional evaluation of TERT-CLPTM1L genetic variants associated with susceptibility of papillary thyroid carcinoma. *Sci Rep* **6**, 26037 (2016).
635. Zhang, Y. *et al.* Common variations in TERT-CLPTM1L locus are reproducibly associated with the risk of nasopharyngeal carcinoma in Chinese populations. *Oncotarget* **7**, 759-70 (2016).
636. James, M.A. *et al.* Functional characterization of CLPTM1L as a lung cancer risk candidate gene in the 5p15.33 locus. *PLoS One* **7**, e36116 (2012).
637. Yang, C. *et al.* Positional integration of lung adenocarcinoma susceptibility loci with primary human alveolar epithelial cell epigenomes. *Epigenomics* **10**, 1167-1187 (2018).
638. Barthel, F.P. *et al.* Systematic analysis of telomere length and somatic alterations in 31 cancer types. *Nat Genet* **49**, 349-357 (2017).
639. Puskás, L.G. *et al.* Novel Anti-CRR9/CLPTM1L Antibodies with Antitumorigenic Activity Inhibit Cell Surface Accumulation, PI3K Interaction, and Survival Signaling. *Mol Cancer Ther* **15**, 985-97 (2016).
640. Busque, L. & Godley, L.A. MBD4: guardian of the epigenetic galaxy. *Blood* **132**, 1468-1469 (2018).

641. Francis, J.H. *et al.* Investigation of Somatic GNAQ, GNA11, BAP1 and SF3B1 Mutations in Ophthalmic Melanocytomas. *Ocul Oncol Pathol* **2**, 171-7 (2016).
642. Bronner-Fraser, M. Neural crest cell formation and migration in the developing embryo. *FASEB J* **8**, 699-706 (1994).
643. Visser, M., Palstra, R.J. & Kayser, M. Allele-specific transcriptional regulation of IRF4 in melanocytes is mediated by chromatin looping of the intronic rs12203592 enhancer to the IRF4 promoter. *Hum Mol Genet* **24**, 2649-61 (2015).
644. Huber, M. & Lohoff, M. IRF4 at the crossroads of effector T-cell fate decision. *Eur J Immunol* **44**, 1886-95 (2014).
645. Grusdat, M. *et al.* IRF4 and BATF are critical for CD8(+) T-cell function following infection with LCMV. *Cell Death Differ* **21**, 1050-60 (2014).
646. Qu, H., Sun, H. & Wang, X. Neogenin-1 Promotes Cell Proliferation, Motility, and Adhesion by Up-Regulation of Zinc Finger E-Box Binding Homeobox 1 Via Activating the Rac1/PI3K/AKT Pathway in Gastric Cancer Cells. *Cell Physiol Biochem* **48**, 1457-1467 (2018).
647. Kane, M. *et al.* Identification of Interferon-Stimulated Genes with Antiretroviral Activity. *Cell Host Microbe* **20**, 392-405 (2016).
648. Thomsen, H. *et al.* Genome-wide study on uveal melanoma patients finds association to DNA repair gene TDP1. *Melanoma Res* (2019).
649. Derrien, A.C. *et al.* Germline MBD4 Mutations and Predisposition to Uveal Melanoma. *J Natl Cancer Inst* **113**, 80-87 (2021).
650. Mobuchon, L. *et al.* Different Pigmentation Risk Loci for High-Risk Monosomy 3 and Low-Risk Disomy 3 Uveal Melanomas. *J Natl Cancer Inst* (2021).

RÉSUMÉ

Le mélanome uvéal (MU) est un cancer rare, mais reste la tumeur intraoculaire primaire la plus fréquente chez l'adulte avec 500 nouveaux cas par an en France. Le MU est lié à la transformation maligne des mélanocytes de l'uvée, composée de la choroïde (90% des cas), du corps ciliaire (6%) et de l'iris (4%). Le MU a une faible charge mutationnelle et comporte deux événements oncogéniques majeurs : l'activation constitutive de la voie de signalisation $G\alpha q$ par mutations activatrices mutuellement exclusives de *GNAQ*, *GNA11*, *PLCB4* ou *CYSLTR2*, et une seconde étape dans la transformation maligne, caractérisée elle aussi par des mutations mutuellement exclusives impactant le risque métastatique : les mutations de *BAP1* (associées à une inactivation bi-allélique du gène par perte de l'allèle sauvage par monosomie 3, M3) constituent le MU à haut risque métastatique et mauvais pronostic, tandis que les mutations de *SF3B1* ou *EIF1AX* (associées à la disomie 3, D3) confèrent un risque respectif intermédiaire et faible de métastases. Le très mauvais pronostic du MU métastatique, hépatique dans la majorité des cas, se caractérise par une survie médiane de 12 mois. Le MU est associé aux populations européennes, et aux individus à la peau et aux yeux clairs, avec un risque relatif (RR) 10 à 20 fois plus élevé chez les populations de souche européenne qu'africaine ou asiatique. L'absence de signature UV exclue le rôle des différences de pigmentation dans la protection des UV pour expliquer cette épidémiologie remarquable. De même, les mutations germinales de *BAP1*, seul facteur de risque génétique à forte pénétrance connu, n'expliquent qu'une partie des cas familiaux du MU, suggérant l'existence d'autres facteurs de prédisposition génétique.

Nous avons émis l'hypothèse d'allèles à risque prédisposant les populations européennes au MU. Nous avons utilisé deux approches : (i) l'évaluation du gène candidat *MBD4* par criblage de mutations dans une cohorte consécutive de 1,093 patients de MU, nous permettant de démontrer que *MBD4* prédispose au MU à haut risque métastatique avec un RR de 9 comparé à la population générale ; et (ii) l'identification de polymorphismes fréquents de risque par étude d'association pan-génomique et la caractérisation de leurs conséquences moléculaires par une étude de génomique fonctionnelle. Nous avons ainsi pu démontrer l'implication de régions de risque sur les chromosomes 5, 6 et 15 (loci respectifs *TERT/CLPTM1L*, *IRF4* et *HERC2/OCA2*). Les deux derniers loci ont un rôle majeur dans la pigmentation ; nous avons pu montrer leur association différentielle avec le MU à haut (M3) et faible (D3) risque métastatique (*HERC2/OCA2* et *IRF4* respectivement), indiquant deux facteurs de risque distincts ayant un impact sur le potentiel métastatique. Enfin, la caractérisation fonctionnelle du locus *TERT/CLPTM1L* nous a permis de démontrer la spécificité allélique de la régulation génique à cette région et d'identifier un variant fonctionnel du MU : rs452384. Celui-ci permet la fixation différentielle de facteurs nucléaires, parmi lesquels nous avons identifié par spectrométrie de masse quantitative NKX2.4, facteur de transcription spécifiquement associé à l'allèle T de rs452384. La sous-expression de NKX2.4 dans des modèles cellulaires du MU résulte en l'augmentation légère mais significative de l'expression des gènes *TERT* et *CLPTM1L*, constituant la première étape dans l'élucidation des mécanismes biologiques de cette région de risque. Enfin, nous émettons l'hypothèse que rs452384 régule de manière différentielle la longueur des télomères, avant l'apparition de la tumeur. Ces résultats mettent en avant le rôle de facteurs de risque à pénétrance faible ou modérée dans le MU et l'implication potentielle de mécanismes biologiques sous-jacents. Notamment, l'impact clinique du MU dans un contexte d'inactivation de *MBD4* et l'inclusion de *MBD4* dans les panels oncogénétiques pourraient permettre l'identification de patients répondant à l'immunothérapie.

MOTS CLÉS

Mélanome uvéal, génomique fonctionnelle, génétique, cancer, prédisposition génétique, oncogénèse, *MBD4*

ABSTRACT

Uveal melanoma (UM) is a rare tumor, yet it is the most common primary intraocular malignancy in adults with 500 new cases per year in France. It arises from the malignant transformation of melanocytes of the uvea, composed of the choroid (90% of cases), ciliary body (6%) and iris (4%). UM is a genetically simple tumor characterized by two major somatic driver events: mutually exclusive mutations of *GNAQ*, *GNA11*, *PLCB4* or *CYSLTR2*, initiating tumorigenesis by constitutive activation of the Gαq signaling pathway, and a second event through mutually exclusive mutations that define metastatic risk and patient outcome: loss-of-function mutations of *BAP1* associated with its bi-allelic inactivation by monosomy 3 (M3) leads to UM with high metastatic risk and poor prognosis, while disomy 3 (D3) is associated with either *SF3B1* or *EIF1AX* mutations and respectively confer intermediate and low metastatic risk. Prognosis is dismal after occurrence of metastasis in 30-50% of cases, almost invariably to the liver, with a median survival of 12 months. Strikingly, individuals of European ancestry with light eyes and fair skin are particularly at risk of developing UM, with a relative risk (RR) of up to 20 in populations of European ancestry compared to those of African or Asian ancestries. However, the absence of ultra-violet (UV) mutational signature and stable UM incidence over the past decades rule out a role for pigmentation protecting against UV light to explain this peculiar epidemiology. Besides, familial cases of UM are seen in 1% of UMs, yet germline mutations of *BAP1* are the only strongly predisposing genetic risk factor known so far and only explain a fraction of familial cases, suggesting that genetic risk of UM mostly remains unaccounted for.

We hypothesized that some susceptibility alleles could predispose populations of European ancestry to UM. We sought to investigate the genetic predisposing factors in UM following two approaches: (i) a candidate-gene approach by targeted-sequencing of *MBD4* in a large consecutive cohort of 1,093 UM patients, allowing us to demonstrate that *MBD4* is a UM predisposing gene associated with high-risk M3 UM, conferring a RR = 9.2 of developing UM compared to general population; and (ii) a genome-wide association study in UM followed by a functional genomics study to identify common, low-penetrant UM risk alleles. Through this second approach, we confirmed the UM risk region at chromosome (chr) 5 on the *TERT/CLPTM1L* locus and further identified two pigmentation susceptibility loci, *HERC2/OCA2* (chr15) and *IRF4* (chr6). Importantly, these two loci were differentially associated with high-risk, M3 and low-risk, D3 UMs respectively, distinguishing two UM subgroups marked by specific genetic risk factors influencing tumor biology and metastatic potential. Lastly, we functionally characterized the *TERT/CLPTM1L* UM risk locus. After demonstrating an allele-specific regulation of gene expression at this locus, we narrowed down the genomic region to one functional variant, rs452384, which exhibited allele-specific nuclear factor binding properties. Using a quantitative proteomic approach, we evidenced binding of NKX2.4 transcription factor specifically to the rs452384-T allele. We further show that knockdown of NKX2.4 in UM cell lines results in a subtle yet significant increase in *TERT* and *CLPTM1L* expression, representing the first steps towards deciphering the tumorigenic mechanism of this risk locus in UM. Finally, we suggest that rs452384 may mediate its activity through differential telomere length regulation, potentially arising prior to UM tumorigenesis. Taken together, these results shed light on multiple moderate and low-risk genetic predisposition factors in UM, which bring new insights on the underlying biological mechanisms in UM predisposition and has clinical relevance as patients with *MBD4* deficiencies may respond to immunotherapy, resulting in the addition of *MBD4* to clinical oncogenetic panels.

KEYWORDS

Uveal melanoma, functional genomics, genetics, cancer, predisposition factors, oncogenesis, *MBD4*