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 (Mobitec) 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 manufacturerôs 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 manufacturerôs 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, 20mL imidazole pH 7.4, 10mM $MgCl_2$, 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-5ôTCGGATGTTGTGGGTCAG(**C/T**)GCATGATAGTGTA-3ǫ

5ô TACACTATCATGCGCTGACCCACAACATCCGA-3ô 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 Kaplanï 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.

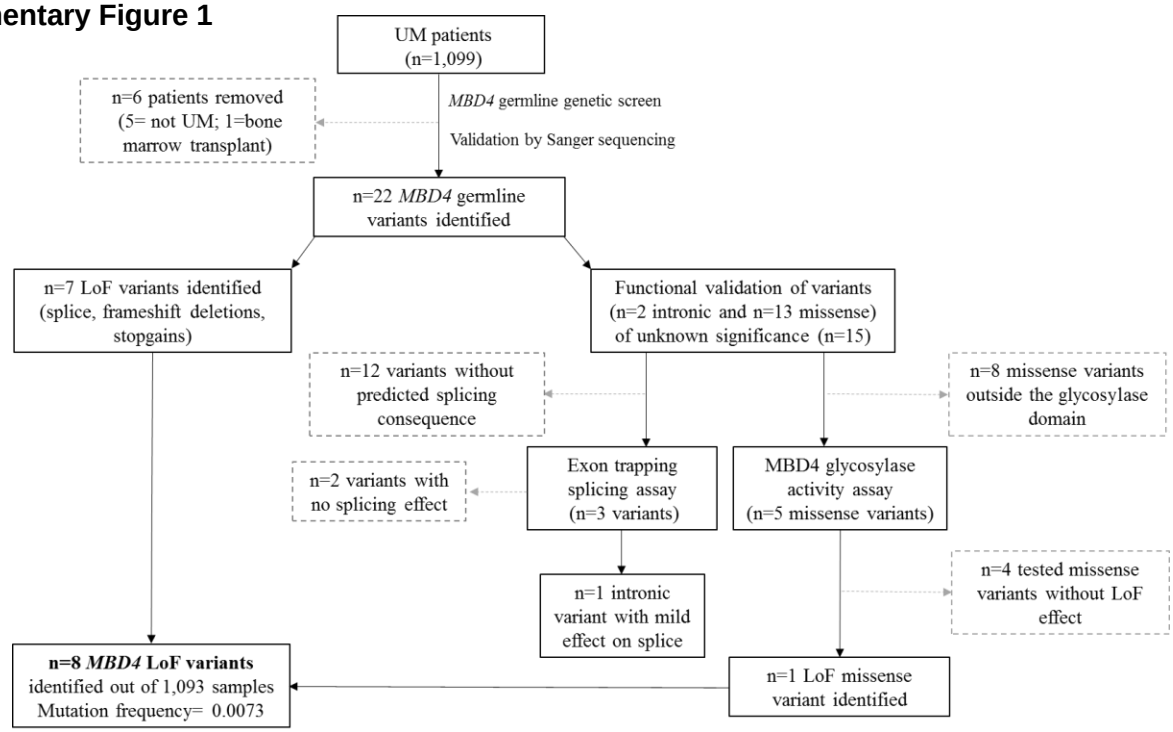
Supplementary References

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.

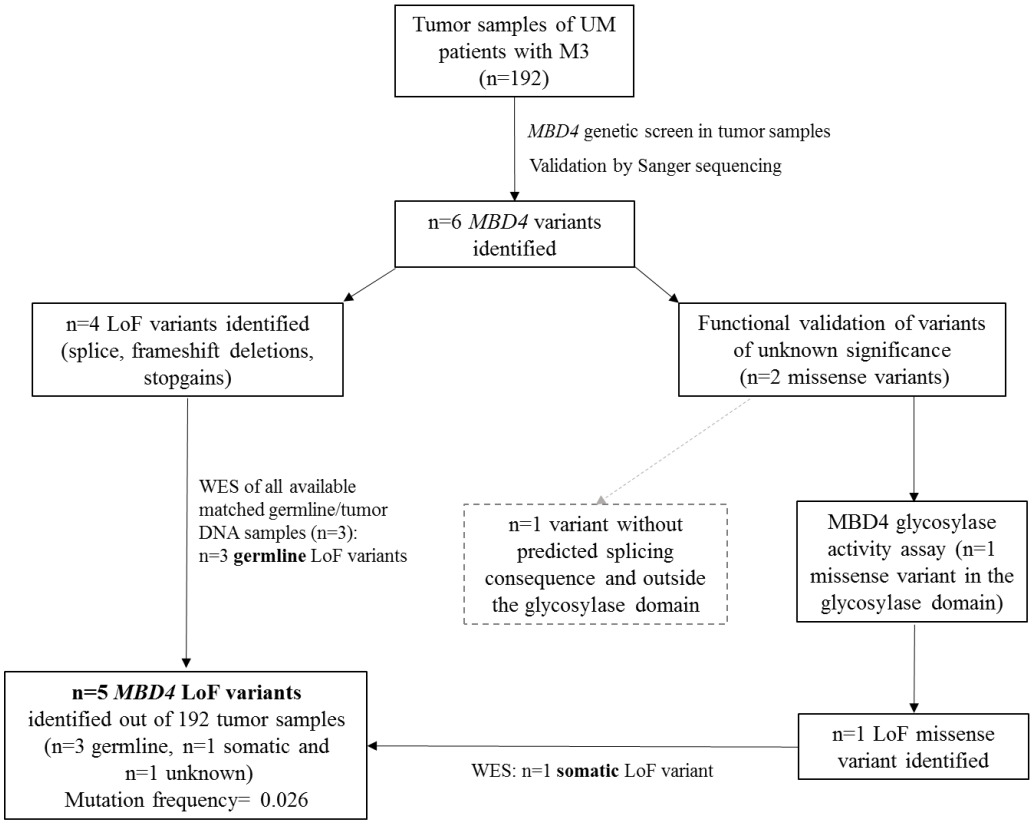
Supplementary Figures

Supplementary Figure 1

A.

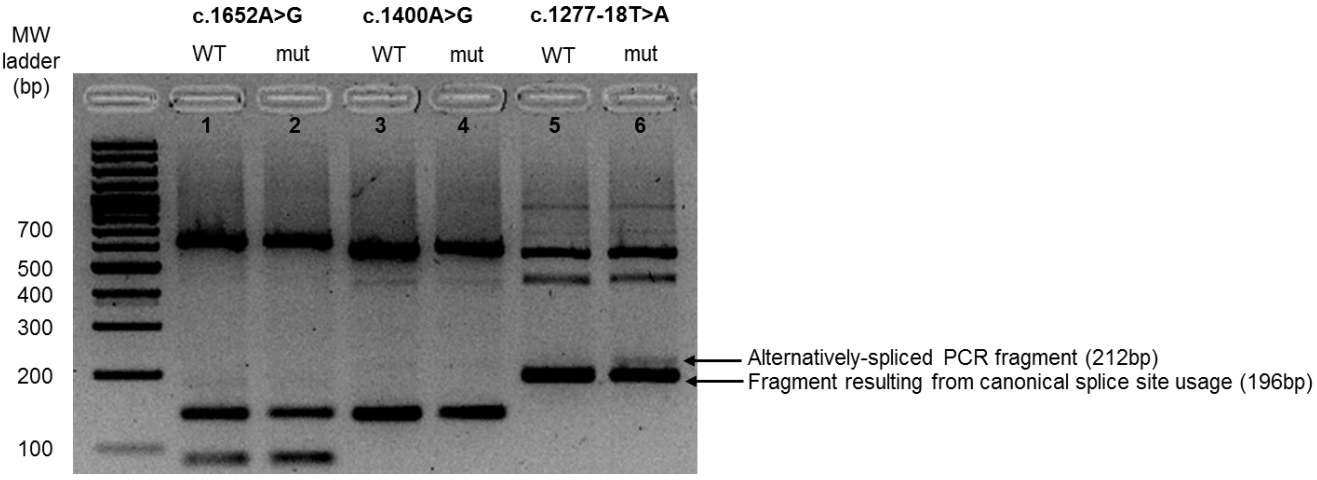


B.



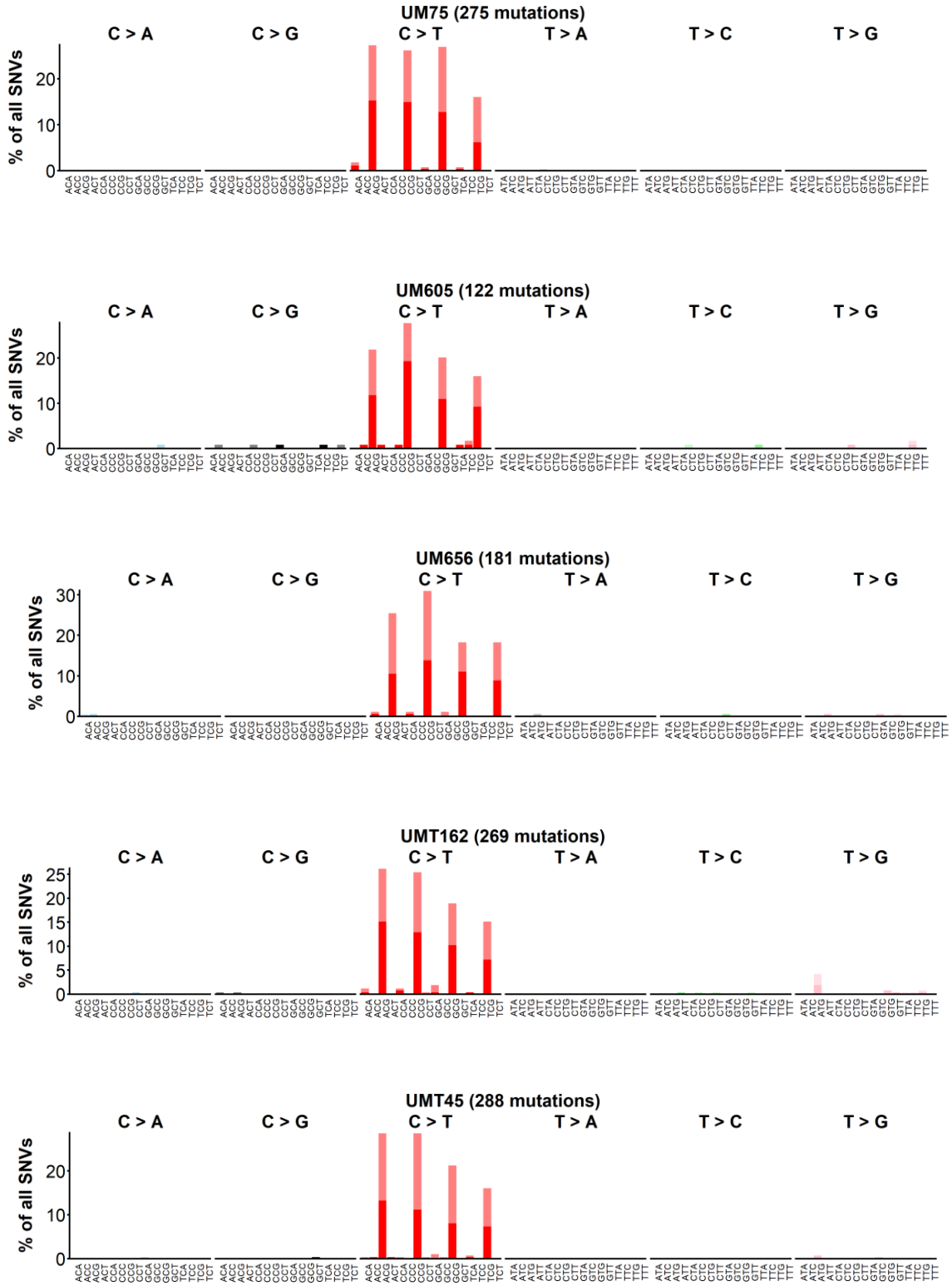
Supplementary Figure 1. Full workflow for the identification of loss-of-function *MBD4* variants among (A) our in-house cohort of 1,099 consecutive uveal melanoma patients and (B) our monosomy 3 tumor series of 192 uveal melanoma patients. Boxes in dotted lines indicate the patients that were excluded 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.

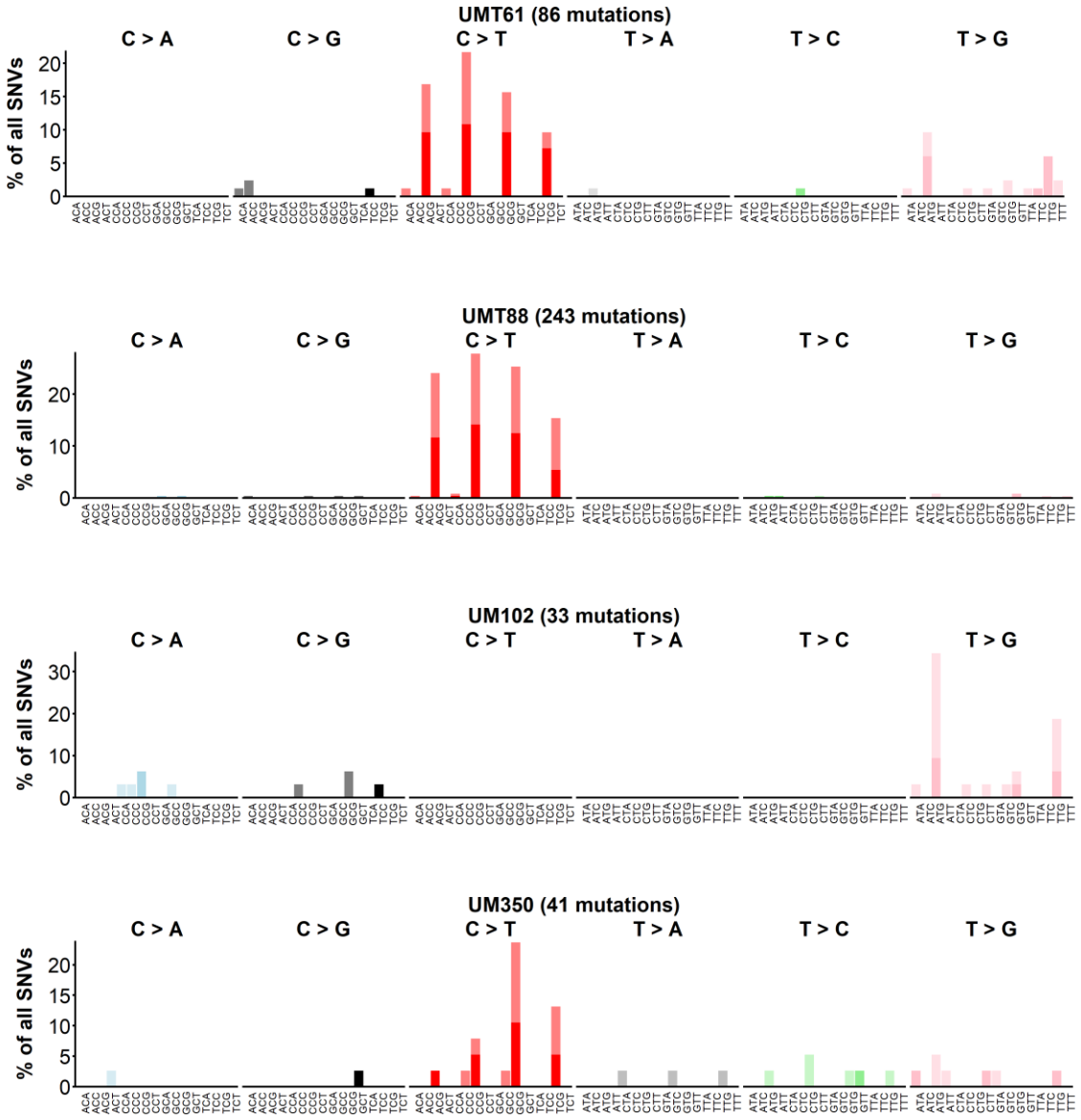
Supplementary Figure 2



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 Finder[18]. 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).

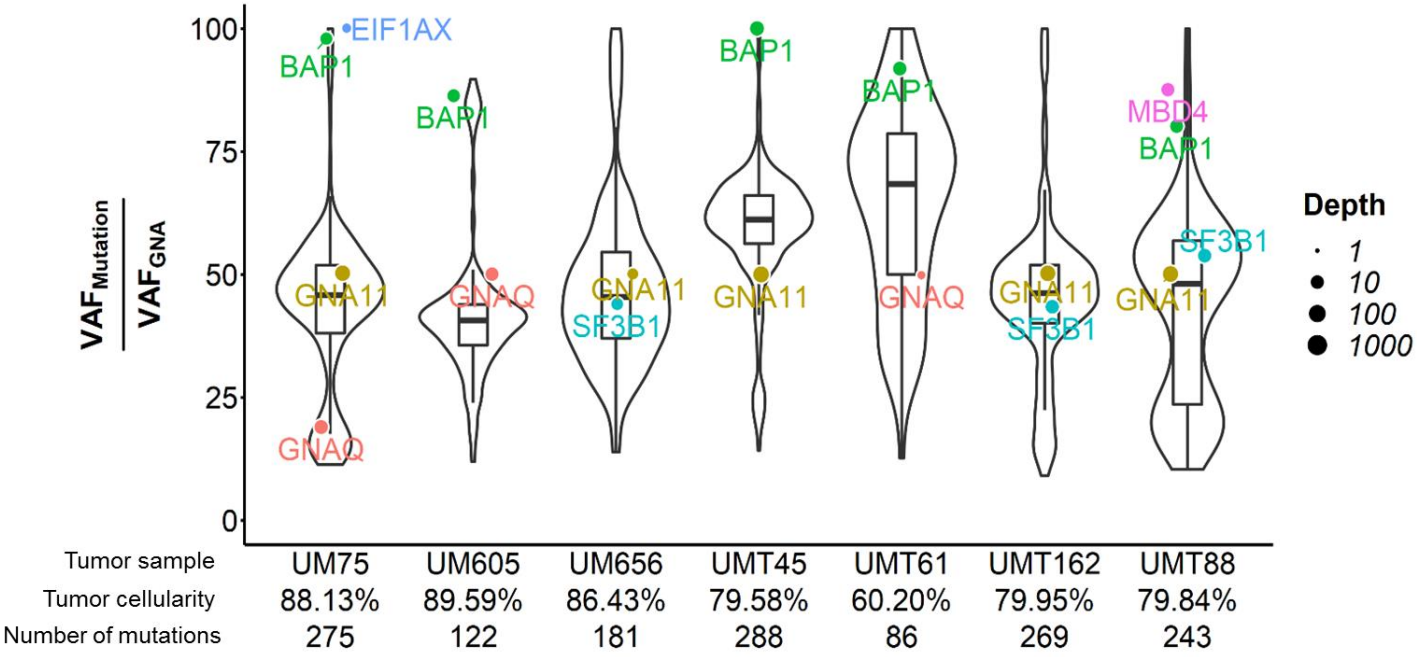
Supplementary Figure 3





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.

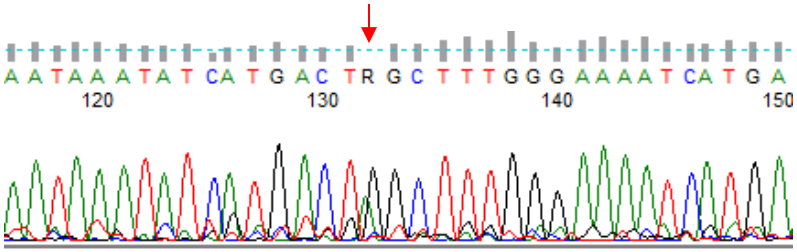
Supplementary Figure 4



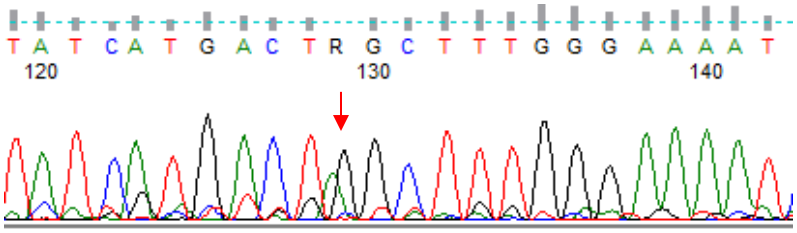
Supplementary Figure 4. **Variant Allele Frequency** (**VAF) distribution of all somatic variants in the tumors of the 7 *MBD4*-deficient UM patients with available exome data**. The box plot represents the median VAF and Q1 – Q3 quartile interval. The key driver events (mutations in *BAP1, GNA11, GNAQ, 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

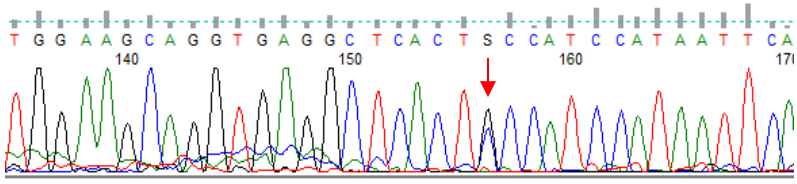
Consecutive germline UM series:



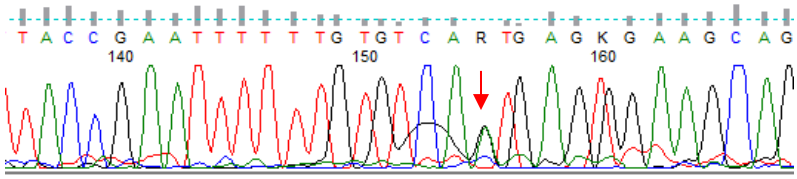
c.1706G>A, p.Trp569* (UM75)



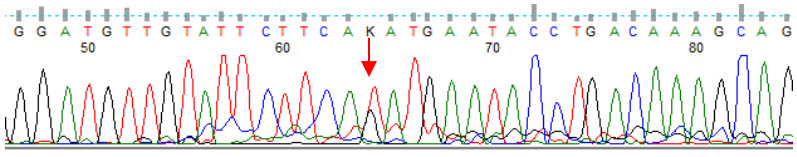
c.1706G>A, p.Trp569* (UM1033)



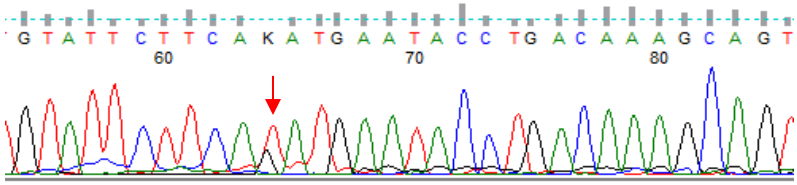
c.1665+13C>G (UM365)



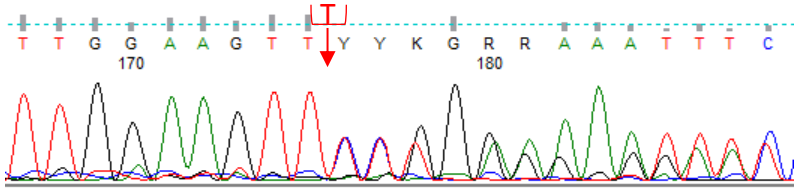
c.1652A>G, p.Asn551Ser (UM350)



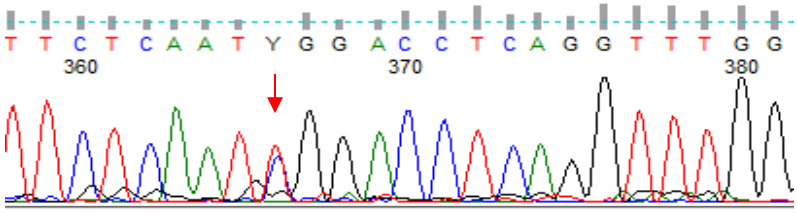
c.1562-1G>T, p.Asp521Profs*4 (UM49)



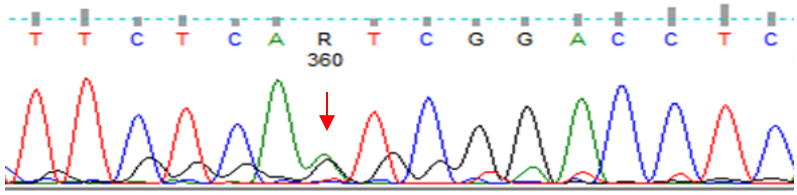
c.1562-1G>T, p.Asp521Profs*4 (UM1088)



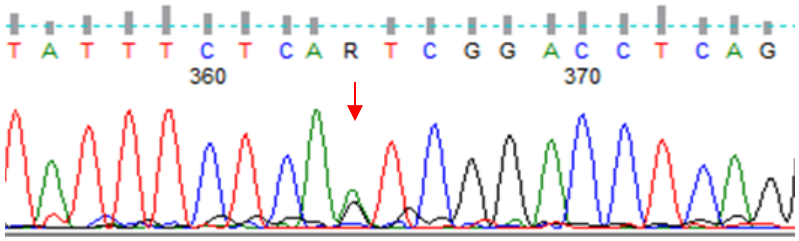
c.1443delT, p.Leu482Trpfs*9 (UM656)



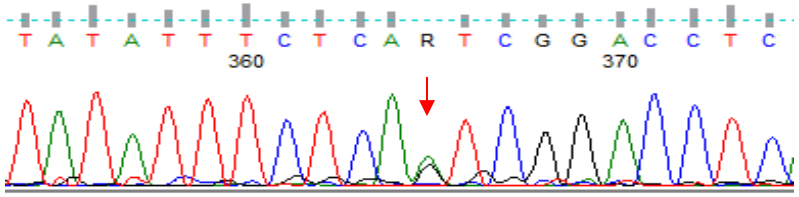
c.1402C>T, p.Arg468Trp (UM293)



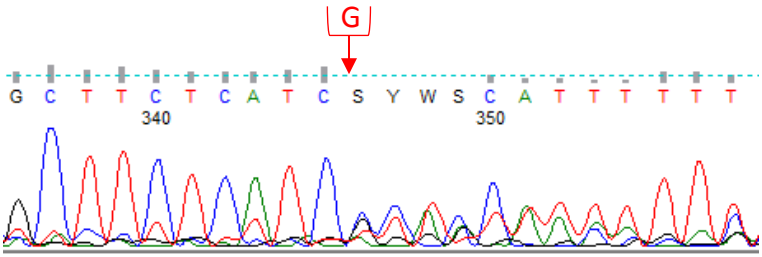
c.1400A>G, p.Asn467Ser (UM867)



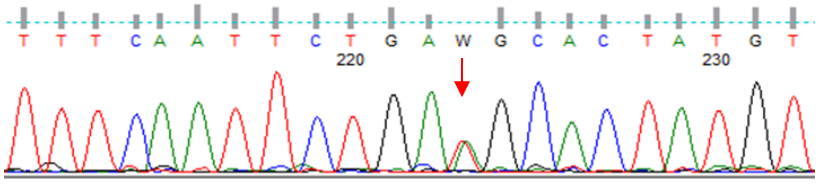
c.1400A>G, p.Asn467Ser (UM75)



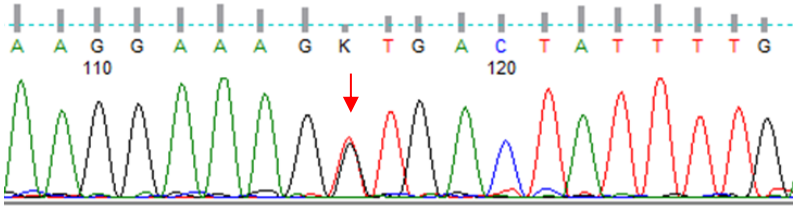
c.1400A>G, p.Asn467Ser (UM547)



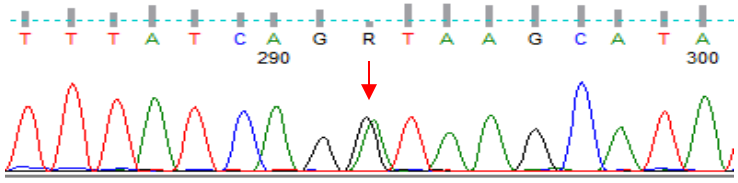
c.1384delG, p.Ala462Leufs*29 (UM605)



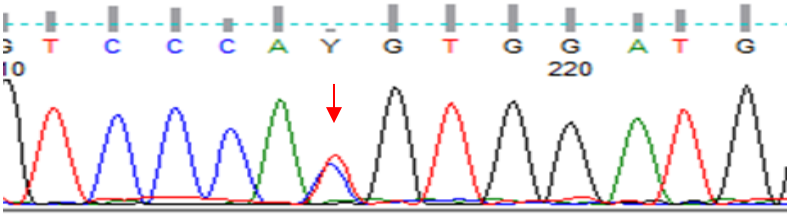
c.1277-18T>A (UM343)



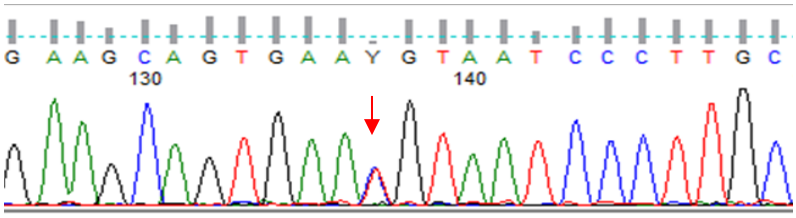
c.703G>T, p.Val235Leu (UM616)



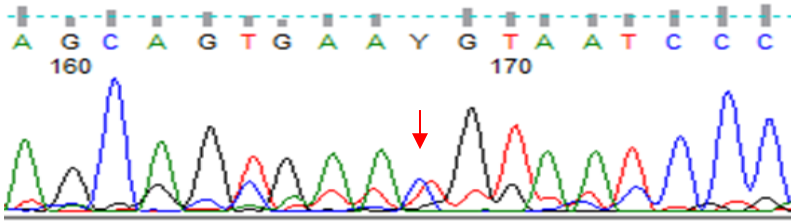
c.335+1G>A, p.Arg83Profs*5 (UM436)



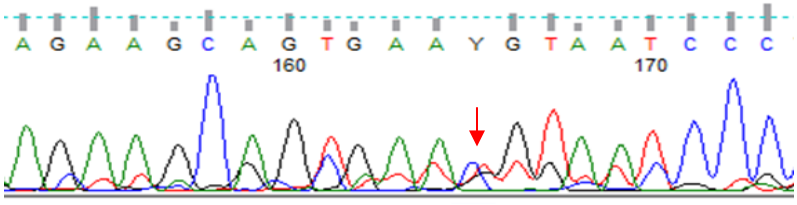
c.262T>C, p.Cys88Arg (UM1055)



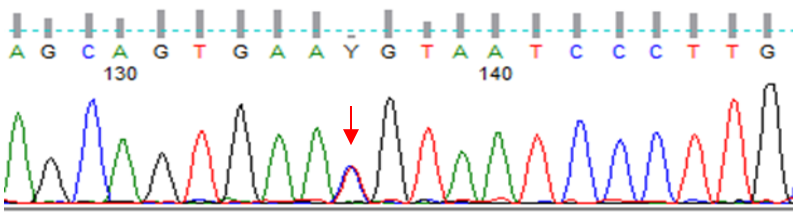
c.181T>C, p.Cys61Arg (UM402)



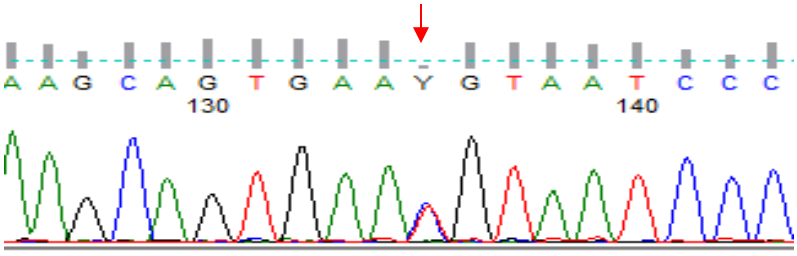
c.181T>C, p.Cys61Arg (UM835)



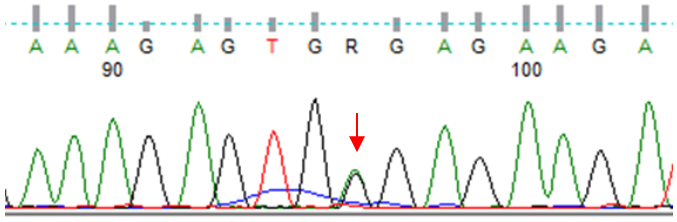
c.181T>C, p.Cys61Arg (UM872)



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

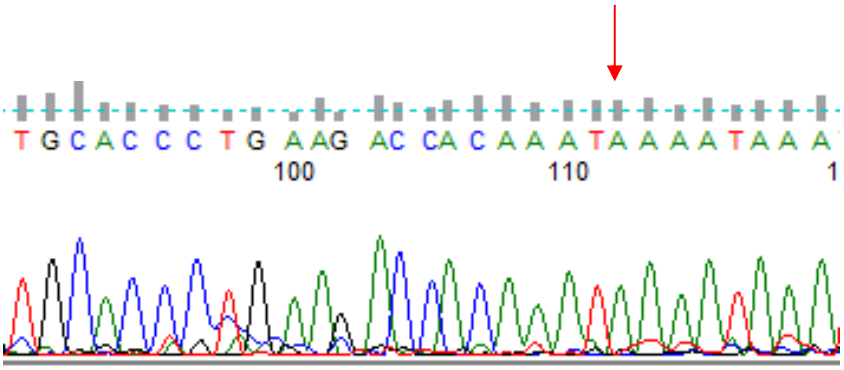


c.181T>C, p.Cys61Arg (UM42)

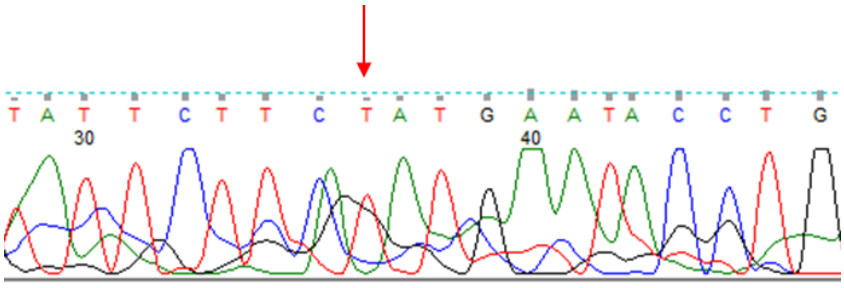


c.181T>C, p.Cys61Arg (UM102)

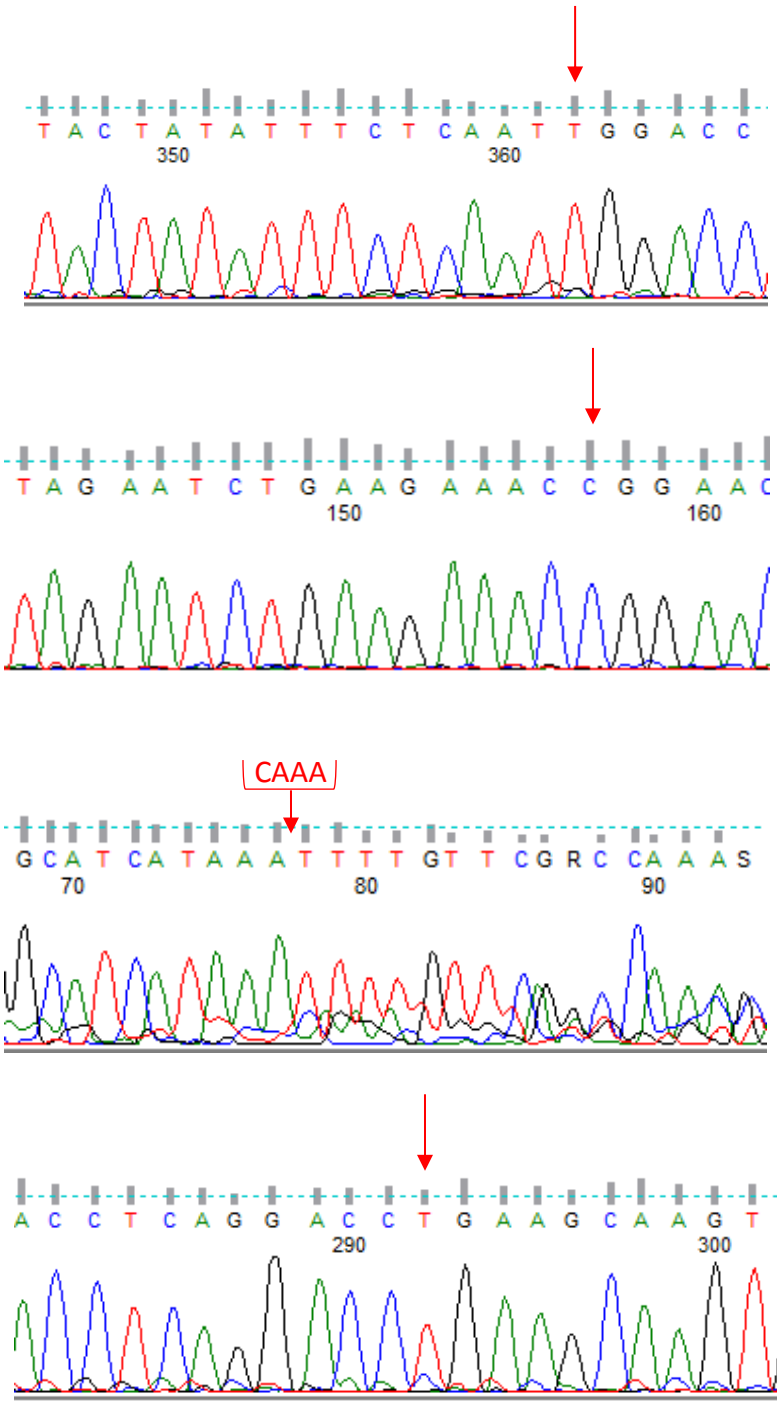
Monosomy 3 tumor UM series:



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

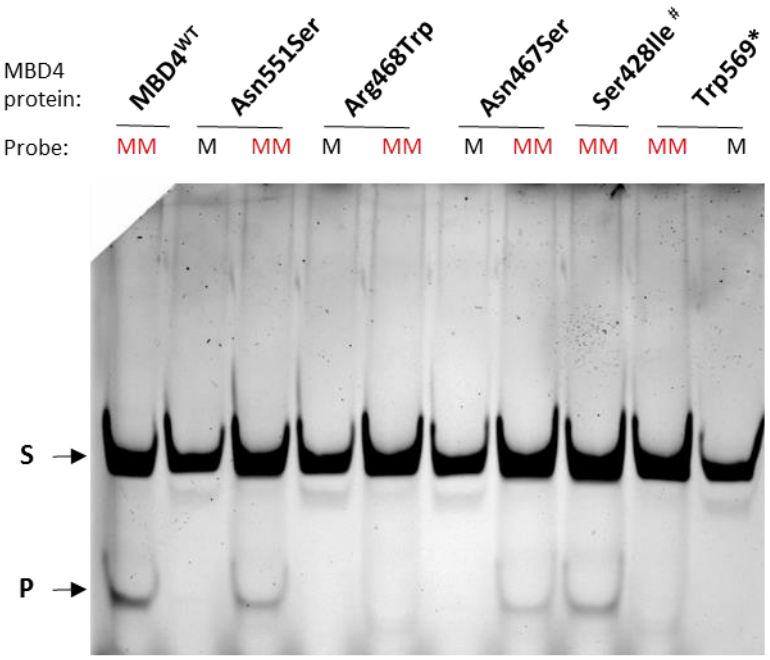


c.1562-1G>T, p.Asp521Profs*4 (UMT45)



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 Nurden (#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[À]	ï	ï	Inactive[§]	Yes (99.0%)	275	ï	ï	ND	2	282518	7,08E-06	2	129130	1,55E-05
UM1033											ND	ND									
UMT62	129150399	c.1688T>A	p.Leu563*	rs200758755	stop_gain	LoF	ND	ï	ï	ND[‖]	Yes (Sanger)	ND	ï	ï	ND	8	250988	3,19E-05	8	113702	7,04E-05
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	ï	ï	ND	ND	ND	ï	ï	ND	10	251472	3,98E-05	5	113766	7,03E-05
UM1088																					
UMT45											Yes (100%)	288									
UM656	129152059	c.1443delT	p.Leu482Trpfs*9	rs769076971	frameshift_deletion	LoF	GL	ï	ï	ND	Yes 97.7%	180	ï	ï	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							Som[ÿ]				Yes (100%)	243									
UM867	129152704	c.1400A>G	p.Asn467Ser	rs78782061	nonsynonymous_SNV	missense	GL	Possibly damaging	Tolerated	Benign	ND	ND	Crytpic acceptor site	79.9 vs 71.1	No	479	282746	1,69E-03	349	129084	2,64E-05
UM75											Yes (0.0%)**	275									
UM547											ND	ND									
UM605	129152720	c.1384delG	p.Ala462Leufs*29	ï	frameshift_deletion	LoF	GL	ï	ï	ND	Yes (89.8%)	122	ï	ï	ND	ï	ï	ï	ï	ï	ï
UM343	129152845	c.1277-18T>A	c.1277-18T>A	rs1434697310	intronic	intronic	GL	ï	ï	ND	ND	ND	Crytpic 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	ï	ï		Yes (97.2%)	85	ï	ï	ND	2	251306	7,96E-06	0	113650	0.00
UM616	129155784	c.703G>T	p.Val235Leu	ï	nonsynonymous_SNV	missense	GL	Benign	Deleterious	ND	ND	ND	None	None	ND	ï	ï	ï	ï	ï	ï
UMT162	129155946	c.541C>T	p.Arg181*	rs1270271346	stop_gain	LoF	GL	ï	ï	ND	Yes (99.9%)	269	ï	ï	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	ï	ï	ND	ND	ND	ï	ï	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																					
UM102	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

* loss of function (deleterious) mutation

À germline mutation

ÿ 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)

ï 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 singe-nucleotide variants and insertions-deletions. Percent of CpG>TpG transitions is calculated relative to the total number of SNVs.

Patient	UM Cohort	*MBD4* variant	*MBD4* mutation type	Number of mutations	Number of CpG>TpG SNVs*	Number of INDELs[À]	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[ÿ]	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

À insertions - deletions

ÿ 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	AGAAAGGCCCACACACTGTC
	Exon 3 (part 1)	AAAATTTGATCCTGAACTCAATG	GTTGCAGGAGAGCAGAGGAC
	Exon 6	TCTGAAAGTGGTTGCTGGTTC	AGTGGGAGACTGTGGTTTGG
	Exon 7	CACACATTTTGGGAGGGTG	GGTGGACTTATTTTGCCTCAG
MBD4 Multiplex PCR 2 (exons 2, 3, 4, 5 and 8)	Exon 2	GGTTCCTGCATTGTCATGG	GCTATGCTCCCACTACCTGC
	Exon 3 (part 2)	GGCACGAATACAAGATGCAG	GACCCTCAGTGTGACCAGTG
	Exon 3 (part 3)	CATCATCAACACCCTCATCTTC	CAGATACCTATGGCAACATTTGG
	Exons 4-5	ATAGTGCCTGGCATGCTTTG	ATGGACTTTGAACCCAGGC
	Exon 8	TGGTATCGTAATGTACTGTCCCC	CTCTATGGCTGGAAAGGTGG
Generation of mutant vectors by directed mutagenesis	p.Asn551Ser	CCGAATTTTTTGTGTCAGTGAGTGGAAGC AGGTGCACC	GGTGCACCTGCTTCCACTCACTGACACAAA AAATTCGG
	p.Arg468Trp	CATCGCTACTATATTTCTCAATTGGACCTC AGGCAAAATGGC	GCCATTTTGCCTGAGGTCCAATTGAGAAATA TAGTAGCGATG
	p.Asn467Ser	CATCGCTACTATATTTCTCAGTCGGACCT CAGGCAAAATGGC	GCCATTTTGCCTGAGGTCCGACTGAGAAAT ATAGTAGCGATG
	p.Trp569*	GACCACAAATTAAATAAATATCATGACTAG CTTTGGGAAAATCATG	CATGATTTTCCCAAAGCTAGTCATGATATTT ATTTAATTTGTGGTC
Production of recombinant hMBD4 (cloning of *MBD4* 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 CGGTCAC	TAGAACTAGTGGATCAGGGTGAAGGGGGAA 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 CGGTCAC	TAGAACTAGTGGATCAGGGTGAAGGGGGAA TGCC
	c.1277-18T>A	GCAGCCCGGGGGATCTTTCCCAATCAGA ACAGCAA	TAGAACTAGTGGATCAGGTCCGATTGAGAA ATATAGTAGC

Supplementary Table 4. Common variants detected with the targeted-sequencing pipeline for *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.

										GnomAD v2.1 Non-Finnish European population (NFE)			
Position on chromosome 3	Ref* allele	Alt* allele	dbSNP	Mutation type	Number of pools positive for variant allele	Cumulative VAF[Â] in pool population[ÿ]	Estimated variant allele count[§]	Total allele count	VAF in total pool population	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

Â VAF: variant allele frequency

ÿ 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.