haematologica
Journal of the Ferrata Storti Foundation

A founder heterozygous mutation in Methyl-CpG binding domain protein 4 (MBD4) prevalent among Israeli Christian Arabs predisposes to increased mutagenesis

by Gal Dadi, Sara Rosen, Hadas Naor, Elias Hellou, Noa Chapal-Ilani, Tal Bacharach, Nathali Kaushansky and Liran I. Shlush

Received: February 26, 2025.
Accepted: August 22, 2025.

Citation: Gal Dadi, Sara Rosen, Hadas Naor, Elias Hellou, Noa Chapal-Ilani, Tal Bacharach, Nathali Kaushansky and Liran I. Shlush. A founder heterozygous mutation in Methyl-CpG binding domain protein 4 (MBD4) prevalent among Israeli Christian Arabs predisposes to increased mutagenesis. Haematologica. 2025 Sept 4. doi: 10.3324/haematol.2025.287690 [Epub ahead of print]

A founder heterozygous mutation in Methyl-CpG binding domain protein 4 (MBD4) prevalent among Israeli Christian Arabs predisposes to increased mutagenesis

Gal Dadi[1], Sara Rosen[1], Hadas Naor[1], Elias Hellou[2,3], Noa Chapal-Ilani[1], Tal Bacharach[1], Nathali Kaushansky[1], Liran I Shlush[1]

[1] Weizmann Institute of Science, Rehovot, Israel

[2] Mayo Clinic Rochester, MN, USA

[3] Nazareth Hospital EMMS, Nazareth, Israel

Running heads: MBD4 founder mutation

Corresponding author: Gal Dadi, Weizmann Institute of Science, molecular cell biology, Shlush Lab, Ulmann building. Email: galdadi@gmail.com, Phone: +972-507224992

Authorship Contributions: G.D. and L.I.S. designed research, interpreted sequencing results, and revised the manuscript. N.K. supervised the protocol and reviewed the

manuscript. G.D. H.N, E.H. T.B performed the research. G.D. ,S.R. and N.C.I analyzed and interpreted data.

Competing Interests Disclosures: There was no conflict of interest

Data Availability Statement:

Code available at: https://github.com/ShlushLab/Somatic_mutation_detection

Sequencing data is available at Annotare under accession number E-MTAB-14787

Acknowledgments

- Grant support: Israel Science Foundation (ISF))1123/21(Israel Cancer Research Fund (ICRF) (22-107-PG)
- The results here are part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

We identified a rare case of familial germline loss-of-function (LOF) mutation in Methyl-CpG Binding Domain 4, DNA Glycosylase (*MBD4*) associated with early-onset acute myeloid leukemia (AML) in an Israeli Christian-Arab family from a highly endogamous community that reported no known consanguinity. This discovery prompted broader screening, revealing additional unrelated carriers of the same *MBD4* frameshift mutation in the population. To investigate the consequences of heterozygous *MBD4* loss, we analyzed clonal hematopoiesis (CH) and mutational patterns by whole exome sequencing (WES) and found that heterozygous *MBD4* deficiency is associated with increased mutagenesis—particularly characterized by CG>TG transitions.

The base excision repair pathway maintains genomic stability by correcting DNA mismatches. *MBD4* and Thymine-DNA glycosylase (*TDG*) specifically repair T:G mismatches caused by 5-methylcytosine (5mC) deamination, with some substrate specificity [1–3]. While biallelic *TDG* loss is embryonically lethal[4], *MBD4* deficiency increases mutagenesis and tumor risk[5]

Germline biallelic *MBD4* LOF has been linked to cancer including AML, myelodysplastic syndrome (MDS), CH[6], and colonic polyposis, defining the *MBD4*-associated neoplasia syndrome (MANS)[7]. Heterozygous germline mutations with somatic loss of the wild-type allele have also been reported in uveal melanoma[8]. Affected tumors typically exhibit excess CG>TG mutations, with recent reports suggesting a potential SBS96 signature specifically associated with biallelic *MBD4* deficiency[9].

While the phenotype of the homozygous trait of the *MBD4* mutation appears clearer, particularly concerning the hematopoietic and gastrointestinal tract, the phenotype of the heterozygous state remains less defined.

A case involving a 42-year-old male diagnosed with AML harboring eight CH mutations—predominantly CG>TG—detected by a myeloid panel sequencing was presented to our laboratory (figure 1A). The patient's sister had previously succumbed to AML at the age of 30. WES of leukemic blasts revealed a germline biallelic 4-bp deletion in *MBD4* (c.612_615del; p.Ser205ThrfsTer9), confirmed by amplicon sequencing of peripheral blood and shown to segregate disease in the family (figure 1B).

This variant is rare in Genome Aggregation Database (gnomAD v4.1.0) at allele frequency 0.0000399, observed only in the heterozygous state (0.00000248), with no individuals carrying biallelic mutations. A similar familial case of the same biallelic c.612_615del frameshift deletion was reported in a patient with colorectal adenomas and MDS that evolved to AML[10].

Given the patient's origin in an Israeli Christian Arab community with notable endogamy and a high inbreeding coefficient[11], and the family's denial of consanguinity, we hypothesized that this *MBD4* mutation may represent a founder variant predisposing to MANS among Israeli Christian Arabs.

In the current study we aimed to assess the prevalence of the *MBD4* c.612_615del mutation in the Christian-Arab population in Israel, and to study the consequences of the *MBD4* c.612_615del heterozygous carriers.

We conducted a pilot screening of healthy unrelated volunteers from the Israeli Christian-Arab community, unrelated to index family, in collaboration with the Orthodox Church and EMMS Nazareth Hospital using amplicon sequencing. Volunteers with active oncological conditions were excluded.

To assess the phenotype of *MBD4* c.612_615del heterozygous carriers, we performed deep targeted sequencing for CH mutations and WES of peripheral blood to evaluate the impact of *MBD4* c.612_615del on mutagenesis.

The study received approval from the Weizmann institute of science local ethics committee (Institutional Review Board approval 1773-2).

Genomic DNA was extracted from peripheral blood using Qiagen DNA purification kit. For whole-exome sequencing (WES), 1 µg of DNA was fragmented (~220 bp) and libraries were prepared using the xGen Exome v1.0 Panel (IDT). The sequencing was performed on an Illumina NovaSeq X Plus using 100 bp paired-end reads, with an average depth of ~50× and minimum target coverage of 20×.

Public datasets were harmonized for comparison. Beat AML[12] samples were sequenced using the Illumina Nextera RapidCapture Exome kit. The Cancer Genome Atlas (TCGA) blood-derived normal samples—used as germline references for the TCGA project—were selected from chemotherapy-naïve, aged-matched individuals across BRCA (breast invasive carcinoma), LGG (brain lower grade glioma), LIHC (liver hepatocellular carcinoma), and TGCT (testicular germ cell tumor) cohorts. Capture kits included Agilent v3/v5 and Roche VCRome. All raw FASTQ or aligned BAM files from public datasets were reprocessed through the same alignment and variant calling pipeline to minimize bias.

Reads were aligned to GRCh38/hg38 using BWA. Variant calling was performed with Mutect2 (GATK v4.1.7.0) in tumor-only mode with default parameters[13]. Germline filtering used gnomAD v4.1.0 and the 1000 Genomes Project Phase 3 v5 as references. Orientation artifacts were modeled with LearnReadOrientationModel, and calls were filtered with FilterMutectCalls. We retained only PASS variants with variant allele frequency (VAF) between 0.11 and 0.8, TLOD - Log 10 likelihood ratio score of variant existing versus non-existing, with a threshold of 40.

Additional filters excluded variants with: (1) strand bias ($P < 0.01$, chi-square), (2) alternate/reference base quality difference >5, (3) extreme depth (>150×), (4) <3 alternate-supporting reads, (5) <3.35 gnomAD MAF (−log10), (6) proximity <50 bp to another variant, and (7) recurrence across individuals. Annotation was performed using ANNOVAR.

Mutational signature analysis was performed with the MutationalPatterns R package for context-specific C>T substitutions, and with the signatureFit_pipeline() function from the signature.tools.lib[10] package for SBS fitting, focusing on SBS1, SBS5, (signatures associated with spontaneous and clock-like C>T transitions) and SBS96. Signature fitting included 100 bootstraps.

Deep targeted sequencing for clonal hematopoiesis (CH) was done with a 47-gene Molecular Inversion Probe panel[14], requiring depth >100×. Variant significance was assessed using a Poisson exact test with Benjamini–Hochberg multiple testing correction.

Amplicon-based sequencing of *MBD4* c.612_615del was performed using primers with 5′ Illumina adaptors (Fwd: CTACACGACGCTCTTCCGATCTttctgaagttaacatcatcaaca,

Rev: CAGACGTGTGCTCTTCCGATCTaaccaaagtaacaattcaaactg) and sequenced on an Illumina MiniSeq (2×151 bp).

For the index case, somatic myeloid mutations were profiled using the Oncomine™ Myeloid Assay (Ion Torrent S5, GRCh37 reference), using a 5% threshold, excluding synonymous and common SNPs.

Sequencing the peripheral blood from 312 healthy unrelated Christian-Arab individuals (18–93 years) using amplicon-based sequencing of *MBD4* c.612_615del identified 3 *MBD4* c.612_615del heterozygous carriers (~1%) and no biallelic cases. None had a personal or family history of cancer.

To assess CH, we analyzed 11 *MBD4* c.612_615del heterozygous carriers: 3 from our population screen and 8 relatives of the index patient. CH mutations were found in 3, including the 14-year-old daughter (Figure 2A).

To assess the mutational impact of the *MBD4* c.612_615del variant, we compared WES of leukemic blasts from the biallelic case to the Beat AML samples, and WES from *MBD4* c.612_615del heterozygous carriers to WES from peripheral blood of non-leukemic donors (in-house $MBD4^{+/+}$ donors), and blood-derived normal from the TCGA database.

WES of leukemic blasts from the biallelic c.612_615del AML patient confirmed previous observations, were 72% of C>T substitutions occurred at CpG sites (CG>TG), which

was significantly higher compared to AML samples from the Beat AML cohort (n = 11, p = 1.72 × 10^{-1}□, Fisher's exact test; Figure 1C, D). The biallelic c.612_615del patient exhibited an average higher overall mutation burden (552 mutations, n=1) compared to the combined all Beat AML samples (746 mutations, n=11).

To assess mutational impact, we performed WES on peripheral blood from 9 healthy *MBD4* c.612_615del heterozygotes and compared profiles to two non-leukemic control cohorts: (1) blood-derived normal samples from age-matched individuals in the TCGA database who had not received chemotherapy, and (2) 19 non-leukemic *MBD4*$^{+}$/$^{+}$ donors sampled in-house to minimize potential sample handling bias.

MBD4 c.612_615del heterozygous carriers exhibited a significantly higher proportion of C>T transitions at CpG sites compared to both control groups (41% vs. 31% in *MBD4*$^{+}$/$^{+}$ in-house controls, p = 0.00178; 41% vs. 28% in TCGA normal, p = 0.0000245; Wilcoxon rank-sum test; Figure 2B).

We observed 3,274 mutations in *MBD4* c.612_615del heterozygous carriers (n = 9), compared to 5,530 in TCGA controls (n = 36) and 4,174 in *MBD4*$^{+}$/$^{+}$ donors (n = 19), supporting modest CpG□biased mutagenesis in carriers. No significant correlation between age and mutation count was observed (*MBD4*$^{+}$/$^{-}$: r = –0.5, p = 0.17; *MBD4*$^{+}$/$^{+}$: r = 0, p = 0.99; TCGA: r = –0.06, p = 0.75), suggesting age alone does not explain the increased burden (Figure S1A). Our cohort small size and drift might reduce power to detect age related changes.

We used the signatureFit_pipeline function from the signature.tools.lib R package to assess mutational signatures, focusing on SBS96.

The *MBD4* biallelic c.612_615del patient showed a stronger SBS96 fit than Beat AML samples (mean 0.453 vs. 0.318, $p = 9.42 \times 10^{-}$□, Figure S1B).
MBD4 c.612_615del heterozygous carriers had higher SBS96 contributions than both $MBD4^{+}/^{+}$ donors (mean 0.403 vs. 0.335, $p < 2.6 \times 10^{-3}$□) and TCGA controls (0.403 vs. 0.252, $p < 2.6 \times 10^{-3}$□; Figure S1C).

Altogether, this study demonstrates that the *MBD4* c.612_615del variant contributes to increased somatic mutagenesis, with a more pronounced effect in the homozygous state. While *MBD4* c.612_615del heterozygous carriers exhibit a milder phenotype, the consistent rise in C>T transitions at CpG sites suggests a progressive mutagenic process. The ability to detect somatic mutations was probably feasible due to drift[15]. Given its ~1% carrier frequency in Israeli Christian Arabs and potential long□term genomic instability, genetic counseling may be advisable. This may guide reproductive choices and clinical follow□up as penetrance and risks become clearer.

References

1. Hendrich B, Hardeland U, Ng HH, Jiricny J, et al. The Thymine Glycosylase MBD4 Can Bind to the Product of Deamination at Methylated CpG Sites. Nature. 1999;401(6750):301-304.

2. Wiebauer K, Jiricny J. In vitro correction of G o T mispairs to G o C pairs in nuclear extracts from human cells. Nature. 1989;339(6221):234-236.

3. Moréra S, Grin I, Vigouroux A, et al. Biochemical and structural characterization of the glycosylase domain of MBD4 bound to thymine and 5-hydroxymethyuracil-containing DNA. Nucleic Acids Res. 2012;40(19):9917-9926.

4. Cortázar D, Kunz C, Selfridge J, et al. Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. Nature. 2011;470(7334):419-423.

5. Millar CB, Guy J, Sansom OJ, et al. Enhanced CpG Mutability and Tumorigenesis inMBD4-Deficient Mice. Science. 2002;297(5580):403-405.

6. 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.

7. Blombery P, Ryland GL, Fox LC, et al. Methyl-CpG binding domain 4, DNA glycosylase (MBD4)-associated neoplasia syndrome associated with a homozygous missense variant in MBD4: Expansion of an emerging phenotype. Br J Haematol. 2022;198(1):196-199.

8. Villy MC, Le Ven A, Le Mentec M, et al. Familial uveal melanoma and other tumors in 25 families with monoallelic germline MBD4 variants. J Natl Cancer Inst. 2024;116(4):580-587.

9. Degasperi A, Zou X, Amarante TD, et al. Substitution mutational signatures in whole-genome–sequenced cancers in the UK population. Science. 2022;376(6591):science.abl9283.

10. Palles C, West HD, Chew E, et al. Germline MBD4 deficiency causes a multi-tumor predisposition syndrome. Am J Hum Genet. 2022;109(5):953-960.

11. Haber M, Gauguier D, Youhanna S, et al. Genome-Wide Diversity in the Levant Reveals Recent Structuring by Culture. PLoS Genet. 2013;9(2):e1003316.

12. Beat Acute Myeloid Leukemia (AML) 1.0. Available from https://registry.opendata.aws/beataml. (Accessed on 2024, November 19)

13. Bernstein N, Spencer Chapman M, Nyamondo K, et al. Analysis of somatic mutations in whole blood from 200,618 individuals identifies pervasive positive selection and novel drivers of clonal hematopoiesis. Nat Genet. 2024;56(6):1147-1155.

14. Biezuner T, Brilon Y, Arye A Ben, et al. An improved molecular inversion probe based targeted sequencing approach for low variant allele frequency. NAR Genom Bioinform. 2022;4(1):lqab125.

15. Zink F, Stacey SN, Norddahl GL, et al. Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. Blood. 2017;130(6):742-752.

Figure legends:

Figure 1. **Somatic Mutations, Pedigree, and Unique Mutational Signature of the *MBD4* biallelic c.612_615del Patient extracted from leukemic blasts** (A) Table of somatic mutations in CH-related genes identified using a myeloid next generation sequencing panel on leukemic blasts from a 42-year-old AML patient. Several mutations, including those in *STAG2, TET2*, *DNMT3A*, and *IDH1*, occur at CpG sites and represent C>T transitions characteristic of CG>TG mutagenesis. (B) Pedigree of the biallelic c.612_615del patient (marked with X) from an Israeli Christian-Arab family with the *MBD4* mutation. Black represents the mutated allele; white indicates wild type; squares denote males, circles denote females. The year of birth is indicated. The patient's parents declined genetic testing.

(C-D) Comparison of mutational signatures between the *MBD4* biallelic c.612_615del patient and the Beat AML leukemic samples cohort. The total number of mutations was 552 in the biallelic *MBD4* c.612_615del patient and 746 across 11 Beat AML samples. The biallelic patient exhibited a significantly higher proportion of C>T transitions at CpG sites (light red bar), with error bars representing 95% confidence intervals (Fisher's exact test, $p = 1.72 \times 10^{-1}$□). Abbreviations: AML – Acute myeloid leukemia; HGVS – Human Genome Variation Society; VAF – Variant allele frequency

Figure 2. Prevalence and Mutation Patterns in *MBD4* c.612_615del Heterozygous Carriers in the Christian-Arab Population

(A) Clonal hematopoiesis (CH) in heterozygous carriers: CH mutations were detected from peripheral blood DNA in 3 out of 11 *MBD4* c.612_615del heterozygous carriers, including one 14-year-old individual. HGVS annotations are based on canonical transcripts: *TP53* (NM_000546.6), *DNMT3A* (NM_022552.5), and *EZH2* (NM_004456.4).

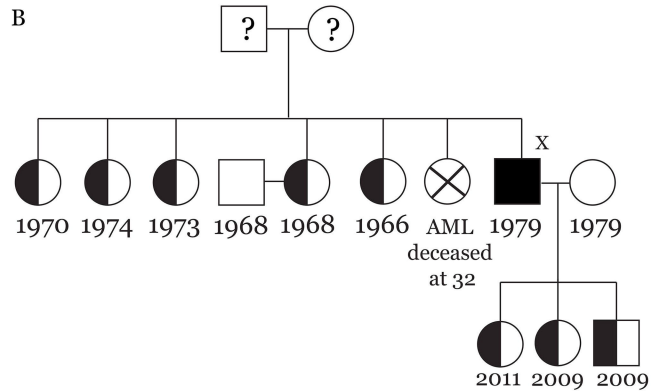
(B) Comparison of the percentage of C>T transitions at CpG sites among *MBD4* c.612_615del heterozygous carriers, in-house *MBD4*$^{+}$/$^{+}$ healthy donors, and WES data from the TCGA project extracted from peripheral blood. *MBD4* c.612_615del heterozygous carriers exhibited significantly higher C>T transitions at CpG sites (41% vs. 31% in *MBD4*$^{+}$/$^{+}$ controls, p = 0.00178; 41% vs. 28% in TCGA normal, p = 0.0000245; Wilcoxon rank-sum test).

(C) Mutational spectra from peripheral blood WES data across cohorts. Total mutations: *MBD4* c.612_615del heterozygous carriers (n = 9) = 3,274; TCGA (n = 36) = 5,530; *MBD4*$^{+}$/$^{+}$ (n = 19) = 4,174. Abbreviations: VAF – Variant allele frequency; TCGA- The Cancer Genome Atlas

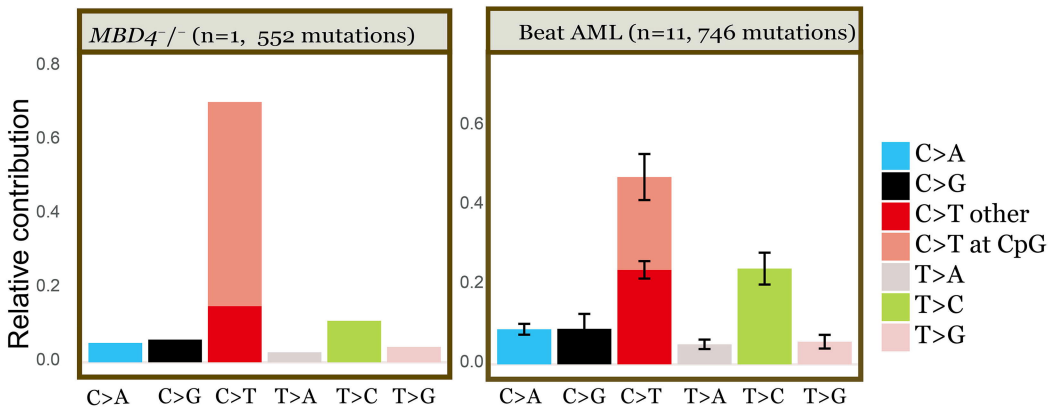
A

Gene	Transcript	HGVS_c	HGVS_p	VAF(%)
DNMT3A	NM_022552.4	c.2644C>T	p.Arg882Cys	31
IDH1	NM_005896.3	c.394C>T	p.Arg132Cys	42
TET2	NM_001127208.2	c.4393C>T	p.Arg1465Ter	51
ETV6	NM_001987.5	c.832_833insC	p.Ile278ThrfsTer22	8
STAG2	NM_001042749.2	c.3133C>T	p.Arg1045Ter	84
ASXL1	NM_015338.6	c.1934_1935insG	p.Gly646TrpfsTer12	45
DNMT3A	NM_022552.4	c.2053G>A	p.Gly685Arg	45
TET2	NM_001127208.2	c.4075C>T	p.Arg1359Cys	40

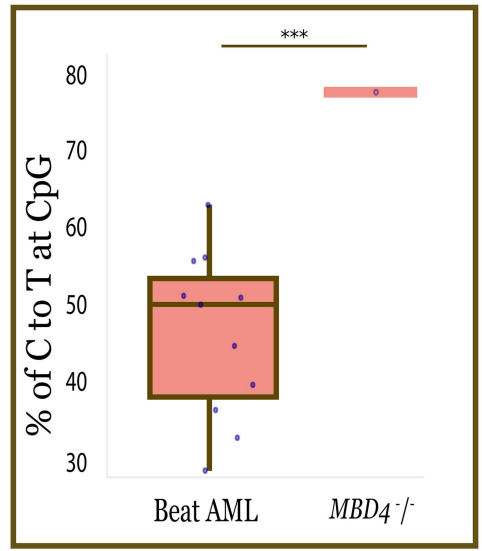
B



C



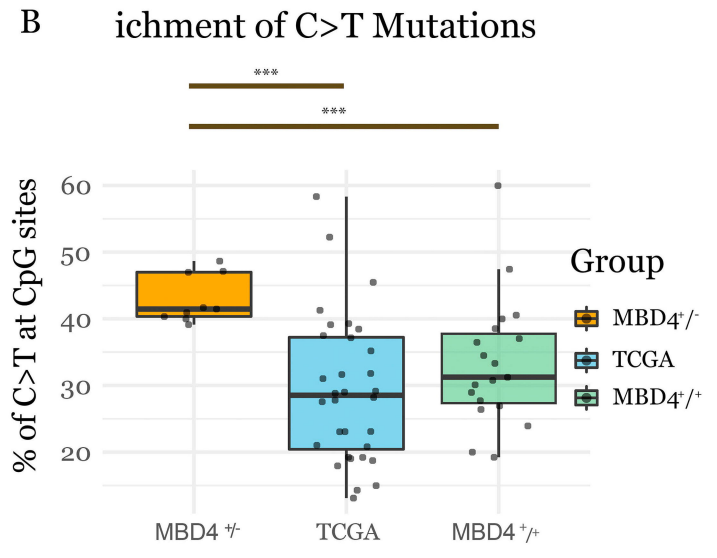
D



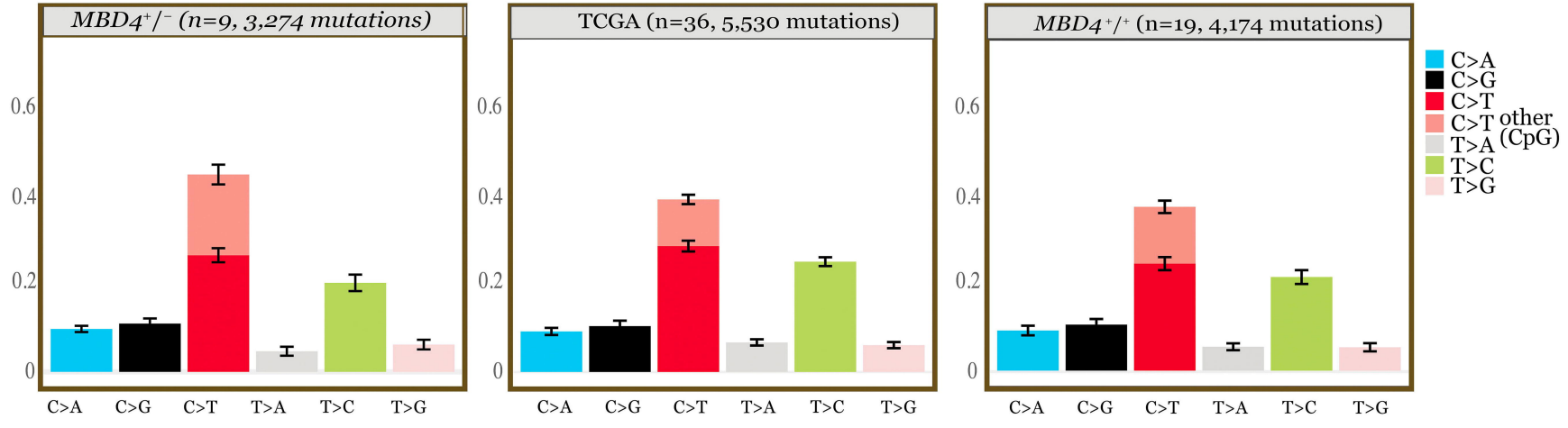
A

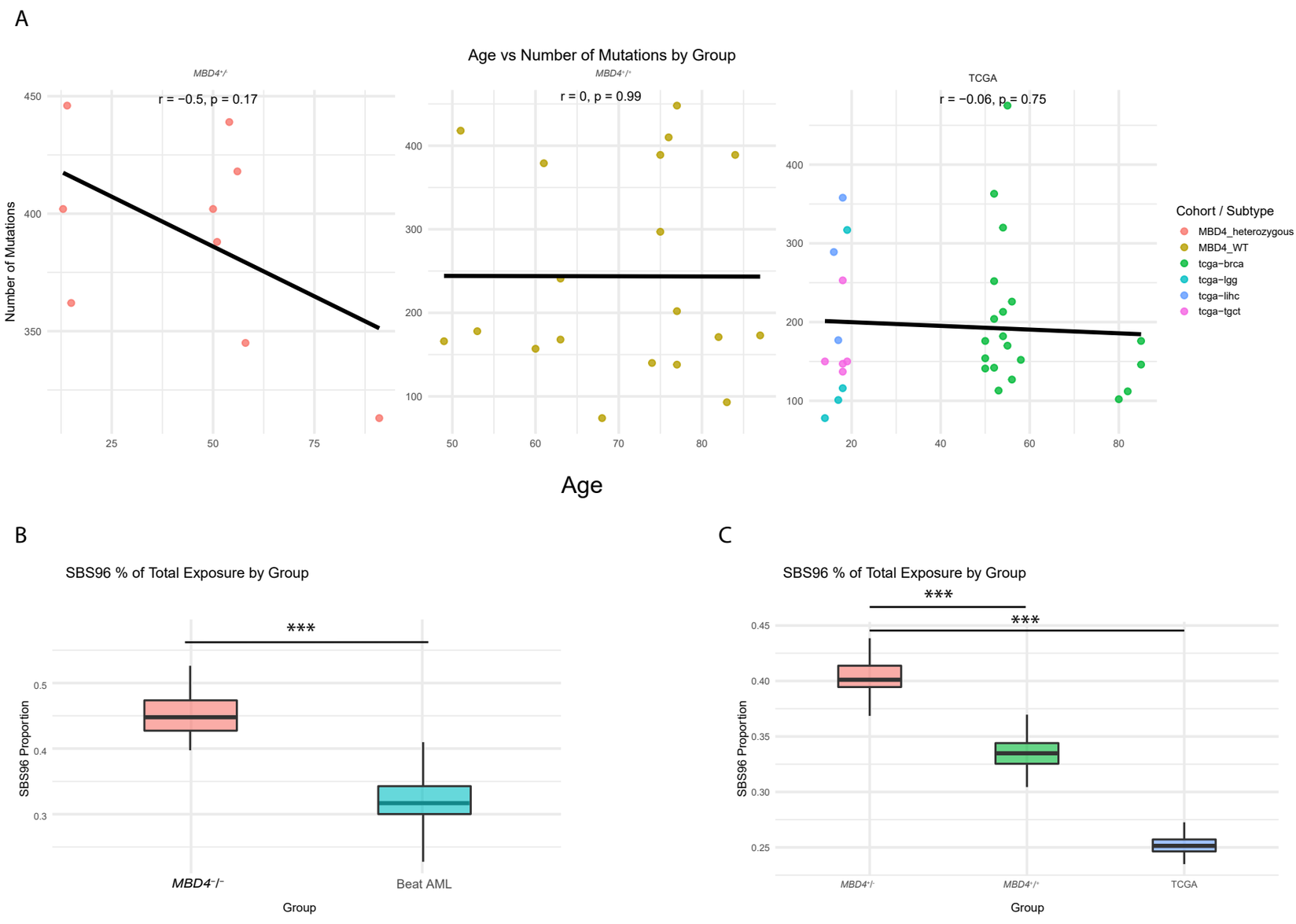
Donor	age(y)	Gene	Position	Nucleotide	Protein	VAF(%)
N6	14	EZH2	chr7:148811708	c.1864G>A	p.Ala622Thr	4
N9	55	DNMT3A	chr2:25248158	c.734C>A	p.Pro245His	6
V1	93	DNMT3A	chr2:25243931	c.1903C>T	p.Arg635Trp	9
		TP53	chr7:7674221	c.742C>T	p.Arg248Trp	1.7

B



C





Supplementary Figure 1. Somatic mutations, age correlations, and CpG-biased mutational signatures in *MBD4*-deficient individuals. (A) Correlation between age and total number of somatic mutations in peripheral blood in three cohorts: *MBD4* c.612_615del heterozygous carriers (left), in-house $MBD4^{+/+}$ controls (middle), and TCGA blood-derived normals (right). No significant correlation was observed (*MBD4* c.612_615del heterozygous carriers: Pearson r = –0.5, p = 0.17; $MBD4^{+/+}$: r = 0, p = 0.99; TCGA: r = –0.06, p = 0.75), suggesting age alone does not account for the increased mutation burden in *MBD4* c.612_615del heterozygous carriers. Each point represents one donor; TCGA samples are color-coded by tumor type. (B) SBS96 mutational signature contribution in leukemic samples from the *MBD4* biallelic c.612_615del patient compared to 11 AML samples from the Beat AML cohort (***p < 0.001, Wilcoxon rank-sum test). (C) SBS96 signature proportion in MBD4 c.612_615del heterozygous carriers (n = 9), in-house $MBD4^{+/+}$ healthy donors (n = 19), and TCGA blood-derived normal samples (n = 36) from peripheral blood, showing significantly increased SBS96 contribution in *MBD4* c.612_615del heterozygous carriers compared to both controls (***p < 0.001, Wilcoxon rank-sum test).
Abbreviations: AML – Acute myeloid leukemia; SBS, Single Base Substitution; TCGA- The Cancer Genome Atlas; WT- wild type.