Downloaded from genome.cshlp.org on June 5, 2026 . Published by Cold Spring Harbor Laboratory Press

GENOME RESEARCH

A naturally occurring variant of *MBD4* causes maternal germline hypermutation in primates

Alexandra M. Stendahl, Rashesh Sanghvi, Samuel Peterson, et al.

Genome Res. published online November 20, 2023
Access the most recent version at doi:10.1101/gr.277977.123

P<P Published online November 20, 2023 in advance of the print journal.

Accepted Manuscript Peer-reviewed and accepted for publication but not copyedited or typeset; accepted manuscript is likely to differ from the final, published version.

Open Access Freely available online through the *Genome Research* Open Access option.

Creative Commons License This manuscript is Open Access.This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International license), as described at http://creativecommons.org/licenses/by/4.0/.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**

To subscribe to *Genome Research* go to:
https://genome.cshlp.org/subscriptions

Published by Cold Spring Harbor Laboratory Press

Downloaded from genome.cshlp.org on June 5, 2026 . Published by Cold Spring Harbor Laboratory Press

A naturally occurring variant of *MBD4* causes maternal germline hypermutation in primates

Alexandra M. Stendahl[1,*], Rashesh Sanghvi[2,*], Samuel Peterson[1,*], Karina Ray[1], Ana C. Lima[1], Raheleh Rahbari[2,#], Donald F. Conrad[1,#]

1.Division of Genetics, Oregon National Primate Research Center, Beaverton, OR, USA

2. Cancer, Ageing and Somatic Mutation (CASM), Wellcome Trust Sanger Institute, Hinxton, UK

*These authors contributed equally

#These authors jointly supervised the project

Correspondence to: rr11@sanger.ac.uk, conradon@ohsu.edu

Downloaded from genome.cshlp.org on June 5, 2026 . Published by Cold Spring Harbor Laboratory Press

Abstract

As part of an ongoing genome sequencing project at the Oregon National Primate Research Center, we identified a rhesus macaque with a rare homozygous frameshift mutation in the gene Methyl-CpG binding domain 4 (*MBD4*). MBD4 is responsible for the repair of C>T deamination mutations at CpG dinucleotides and has been linked to somatic hypermutation and cancer predisposition in humans. We show here that MBD4-associated hypermutation also affects the germline: the 6 offspring of the *MBD4*-null dam have a 4-6-fold increase in *de novo* mutation burden. This excess burden was predominantly C>T mutations at CpG dinucleotides consistent with *MBD4* loss-of-function in the dam. There was also a significant excess of C>T at CpA sites, indicating an important, unappreciated role for MBD4 to repair deamination in CpA contexts. The *MBD4*-null dam developed sustained eosinophilia later in life, but we saw no other signs of neoplastic processes associated with *MBD4* loss-of-function in humans, nor any obvious disease in the hypermutated offspring. This work provides the first evidence for a genetic factor causing hypermutation in the maternal germline of a mammal, and adds to the very small list of naturally occurring variants known to modulate germline mutation rates in mammals.

Downloaded from genome.cshlp.org on June 5, 2026 . Published by Cold Spring Harbor Laboratory Press

Introduction

The DNA glycosylase MBD4 (Methyl-CpG binding domain 4) repairs C>T mutations through the base excision repair pathway by removing the thymine in a G-T mismatch (Pidugu et al. 2021). MBD4 is specifically active at repairing the spontaneous deamination of 5-methyl-cytosine to thymine, one of the most common somatic mutations in the genome (Sanders et al. 2018).

MBD4 is critical for genome stability and preventing mutations, and loss of function of *MBD4* has recently been associated with higher risk of MBD4-associated neoplasia syndrome, a multi-organ tumor predisposition syndrome (Palles et al. 2022). It is further associated with several specific cancers including adenomatous colorectal polyposis, acute myeloid leukemia, and uveal melanoma (Palles et al. 2022; Derrien et al. 2021; Millar et al. 2002). Additionally, many types of cancerous tumors are often found to contain mutations in the *MBD4* gene, especially gastric, endometrial, and pancreatic carcinomas (Wong et al. 2002). Congruent with the function of *MBD4*, C>T mutations are found in high incidence in gastrointestinal tumors of *MBD4*$^{-/-}$ mice, with most occurring at CpG dinucleotides (Wong et al. 2002).

As part of an ongoing genetic study at Oregon National Primate Research Center (ONPRC), we identified a rhesus macaque (ID: 26537) with a germline homozygous TC>C deletion in the *MBD4* gene at 2:147,059,371 (MMul 10 genome assembly, **Figure 1**). In this short report, we describe a germline hypermutation phenotype that we ascribe to this variant, and we analyze the patterns of de novo mutations across 6 offspring of 26537 to generate hypotheses about the timing and molecular basis for this effect.

Downloaded from genome.cshlp.org on June 5, 2026 . Published by Cold Spring Harbor Laboratory Press

Results

Sequencing of the parents of 26537 confirmed that the *MBD4* genotype is an inherited homozygous deletion (**Figure S1**). In macaques, the *MBD4* protein is 572 amino acids long and consists of a methyl binding domain at the N terminus and a glycosylase domain at the C terminus, which is involved in the DNA repair. The mutation in 26537 results in a frameshift (ENSMMUG00000012723:c.984delC,p.Gly329fs) leading to an isoleucine to serine substitution at position 330 (Ile330Ser) and an early truncation at the following amino acid (Ile331AUU). Monkey 26537 is the only known homozygote at ONPRC or in mGAP, the database of macaque genomes (currently 3,455 monkeys,(Bimber et al. 2019)). The allele frequency in mGAP is 0.0072. mGAP contains genotype data on macaques from over 10 centers around the US. While 41% of the individuals in mGAP are from ONPRC, 38/39 MBD4 heterozygotes were found in ONPRC animals, the one exception being an animal from Rocky Mountain Labs.

26537 was an Indian-origin rhesus macaque born at ONPRC in 2007. She gave birth to six offspring between the ages of 4 and 10, and was euthanized at age 14 based on her participation in an experimental protocol unrelated to this study. Through an ongoing genome sequencing study, we sequenced 26537's six offspring, and the sires of her offspring to obtain complete trios (**Figure 2, Table S1**). We also sequenced a large pedigree of macaques from a prolific male breeder (ID: 18607), including his 111 offspring and the 61 dams (**Table S1**). Three of 26537's offspring were sired by this prolific male breeder. Three of 18607's offspring were produced from matings with heterozygous carriers of the *MBD4* mutation (18210 and 18147), providing a limited opportunity to assess potential dominant or additive effects of the mutation.

In our analysis of the prolific male's pedigree, all of 26537's offspring were flagged due to the high number of *de novo* mutations (DNMs, **Supplementary Methods**). Germline *de novo* mutations are mutations which appear in an offspring without detection in the somatic cells of either parent. Humans have a germline mutation rate of 1.2 to 1.3 $\times 10^{-8}$ per position per generation (~60-70 *de novo* mutations), while macaques appear to have a mutation rate of 0.58 to 0.77 $\times 10^{-8}$ per site per generation (Conrad et al. 2011; Kong et al. 2012; Rahbari et al. 2016; Wang et al. 2020; Bergeron et al. 2021). Germline *de novo* mutations can originate as errors in DNA replication during gametogenesis or other processes. Males tend to pass on more *de novo* mutations to their offspring, which may be a result of the continuous germ cell division in spermatogenesis as opposed to oogenesis in females, which is completed before birth (Crow 2000). However, as Abascal et al. found that somatic mutation rate is more related to cell type than number of divisions, there may be additional mutational processes involved in *de*

Downloaded from genome.cshlp.org on June 5, 2026 . Published by Cold Spring Harbor Laboratory Press

novo mutations in sperm (Abascal et al. 2021). Older males are especially at risk of passing high numbers of mutations to their offspring, contributing approximately 1.5-2 mutations per year of age (Kong et al. 2012; Rahbari et al. 2016; Abascal et al. 2021; Moore et al. 2021) . While some studies do show a maternal age effect, it is considerably smaller, approximately 0.3-0.5 mutations per year (Kong et al. 2012; Wong et al. 2016; Jónsson et al. 2017).

In this study, we sequenced the genomes of 176 animals to a mean coverage of 23.6×. Adjusting for callable bases, our filtered de novo mutation callset from control trios produced a mutation rate estimate of 0.81×10^{-8} per bp per generation. While the average rhesus macaque carried 10.5 *de novo* mutations (stdev=4), offspring of 26537 carried 46-64 DNMs (**Table S1**). To formally evaluate the impact of the *MBD4* mutation, we modeled the mutation burden in all sequenced individuals, and separate terms for *MBD4* heterozygotes and homozygotes (**Methods**). We found significant terms for paternal age and MBD4 homozygosity, but not maternal age or *MBD4* heterozygote status (**Figures 3A and 3B**). The number of callable bases per genome, which we use as a proxy for power, was not significantly associated with mutation burden. There are only two heterozygous dams in our analysis, with 3 offspring between the two of them, so power was limited to see an effect of a single copy of the MBD4 frameshift allele on germline mutation.

Despite being sired by 4 different males, all six offspring of 26537 had extremely high numbers of C>T mutations at CpG locations (between 25-39) compared to the average of 1 mutation per monkey not birthed by 26537. Because MBD4 repairs C>T mutations at CpG locations, an enrichment in C>T mutations at CpG locations originating from the maternal haplotype is consistent with the *MBD4* mutation. Moreover, read-based phasing revealed a 16-23 fold increase in maternally-derived mutations in offspring of 26537 compared to control trios (**Figure 3C**, **Table S2**, average of 21.33 maternal mutations vs 1.43, $p=1 \times 10^{-95}$ Poisson regression), while there was no difference in paternally-derived mutations (5.0 vs 4.1, p=0.176).

To build further evidence implicating *MBD4* loss-of-function as the cause of the observed hypermutation, we screened 26537 for other potentially causal mutations in genes related to genome instability genes derived from literature (**Methods**, **Tables S3-S6**), assessing all other parents included in our study to provide context. While 23 genes were mutated in the cohort, the only uniquely mutated gene in 26537 was MBD4 (**Figure 4**).

Mutational signature analysis (Tate et al. 2019) of all 26537 offspring showed almost all the mutations were attributed to signature 1, while in the control trios only 20% of mutations were attributed to signature 1 and the rest (80%) were signature 5 (**Figure 5A**). This composition of mutation signatures is consistent with the mutation data in humans: somatic mutations from *MBD4*$^{-/-}$ individuals have an excess of signature 1

Downloaded from genome.cshlp.org on June 5, 2026 . Published by Cold Spring Harbor Laboratory Press

(Palles et al. 2022) , while germline mutations in typical humans are a mixture of signatures 1 and 5 (Moore et al. 2021). Mutation signature 1 has been attributed to cytosine deamination, and is primarily composed of C>T mutations. Indeed, the vast majority of mutational excess observed in the offspring of 26537 were C>T mutations (**Figure 5B**). While most cytosine deamination is thought to occur at CpGs, approximately 40% of the C>T mutations in hypermutator offspring were not in CpG. We used Poisson regression to test for enrichment of C>T mutations across trinucleotide contexts and observed a clear enrichment in three trinucleotides that did not contain CpG: ACA, CCA, GCA (**Figure 5C**). This indicates a clear pattern of C>T mutations in locations where a cytosine is followed by an adenine (CpA). We fit a generalized linear model to the C>T mutation counts in CpA across all trios, including MBD4 genotype and parental ages as covariates; the MBD4 -/- genotype was significantly associated ($p=3.5 \times 10^{-4}$) but not MBD4 +/- ($p=0.69$). While CpG methylation is maintained in all tissues over the lifespan of an animal, CpA methylation is observed primarily in fetal development and in adult neurons(Greenberg and Bourc'his 2019).

In order to understand whether the hypermutation phenotype or *MBD4* frameshift mutation had an effect on offspring or dam health, we reviewed electronic health records and necropsy information if present for these individuals. None of the hypermutated offspring appear to have any major health concerns. While alive, 26537 did not appear to exhibit any major health issues, aside from a single stillbirth, however tissue from the stillborn was not available for analysis. 26537 was necropsied at 14 years of age as part of another research protocol at ONPRC. The timing of the necropsy was based on the overall design of that study and did not reflect any health or fertility issues of 26537. There were no gross indications of neoplasia at necropsy, nor during histological analysis of brain, gastric tissues or lymph nodes.

At periodic intervals as part of their well check up, monkeys at ONPRC receive complete blood counts (CBCs) and complete metabolic panels. Four CBCs run on 26537 during the last year of her life revealed sustained eosinophilia, with an average measurement of 19.8% eosinophils (all other measures of the CBC and CMP were in the normal range during this time). We compared these high eosinophilic measures to 24,638 CBCs run on 4,699 unique animals at ONPRC since 2016. The average % EOS for animals tested over this time was 2.2%; only 71 tests (0.3%) reported higher EOS counts (**Figure 6**). Fecal parasitology excluded parasitemia as a possible cause for eosinophilia in 26537. Eosinophilia is sometimes observed in several cancers and cancer-related syndromes, especially eosinophilic leukemia and myelodysplastic syndrome (MDS). Both MDS and acute myeloid leukemia have been observed in humans with biallelic mutation of *MBD4*(Palles et al. 2022). These findings provide evidence that the *MBD4* mutations in 26537 potentially had functional consequences at a cellular level, beyond deficiencies in DNA repair.

Downloaded from genome.cshlp.org on June 5, 2026 . Published by Cold Spring Harbor Laboratory Press

Discussion

Through a large genome sequencing project at ONPRC, we identified six hypermutated monkeys, all originating from a dam with a homozygous mutation in the *MBD4* gene causing a premature stop codon. This discovery adds to the short list of naturally occurring genetic variants that have been found to modulate germline mutation rates, in humans, in mice, and now in rhesus macaques.

Hypermutation can be the result of both environmental factors, such as parental exposure to chemotherapy, as well as genetic factors, such as a knockout of a DNA repair gene. Germline hypermutation (a large excess of *de novo* mutations) is quite rare in humans. Kaplanis et al. found twelve individuals out of 21,879 human families with a hypermutated phenotype; only 2 of these could be attributed to genetic causes (Kaplanis et al. 2022). Both cases were born to fathers with a germline hypermutation phenotype, attributed to rare biallelic mutations in the DNA repair genes *XPC* and *MPG*. *MPG* encodes N-methylpurine DNA glycosylase, which, similar to MBD4, is involved in correction of deaminated purines by the base-excision repair pathway. Inherited variation in another DNA glycosylase, *MUTYH*, can increase somatic mutation rates in humans (Robinson et al. 2022), and in the germline in mice (Sasani et al. 2022). Notably, our finding of a maternal germline hypermutator phenotype appears to be the first clear evidence of genetic variants affecting germline mutation in a female mammal. It seems plausible that, as the male germline is subject to extensive methylation, and *MBD4* is strongly expressed in male germ cells (Mahyari et al. 2023), zygotic MBD4 loss-of-function in a male would also lead to an increased rate of germline mutation. Indeed, all of the paternal hypermutator alleles reported to date affect genes that show robust expression in human oocytes (Menezo et al. 2007), and, as such, would be expected to lead to hypermutation of the female germline as well. Because the baseline mutation rate in male primates (including humans) is much higher than in females, we expect that statistical power will continue to favor ascertainment of germline hypermutators from affected males when studying population samples, even for genetic variants that affect male and female germlines equally.

Many questions remain about how, and when during the lifecycle, the germline is vulnerable to DNA damage. The data presented here may hold some important clues regarding the timing of cytosine deamination and its repair in the maternal germline. The *de novo* mutations transmitted from 26537 to her offspring were present in the oocytes that generated each offspring. All offspring from 26537 were heterozygous for the *MBD4* mutation, meaning that they were formed by fertilization of an *MBD4*$^{-/-}$ oocyte by a *MBD4*$^{+/+}$ spermatozoa. In principle, the mutations in the *MBD4*$^{-/-}$ oocytes could have originated at any timepoint in the lifecycle of 26537, from the zygote to the fertilization of each *MBD4*$^{-/-}$ oocyte. Multiple observations in this study would seem to

Downloaded from genome.cshlp.org on June 5, 2026 . Published by Cold Spring Harbor Laboratory Press

restrict the plausible developmental window during which the hypermutator phenotype was active. First, there is no detectable effect of maternal age on the number of mutations in the six offspring of 26537 (**Figure 3B**). A poisson regression model fit to the mutation burden in hypermutator trios alone found a slightly negative but not significant slope ($\square_{age}$ = -0.02, p=0.55), and a model fit to the full data found no interaction between hypermutator status and age ($\square_{age*hypermutator_status}$ = -1.13, p=0.21). Due to the small number of offspring we can't exclude the possibility that there is a small age-related accumulation of mutation over time, but a clock-like steady accumulation of mutation throughout the life of 26537 cannot be the primary source of the excess de novo mutation. Second, there is minimal sharing of mutations among the offspring of 26537: 4 of 316 (1.2%) filtered mutations in the offspring were observed more than once (**Table S7**). This suggests that none of the mutations were present in the embryonic progenitor of all germ cells, and very few of the 316 mutations were present in the embryo at the time of the PGC specification. Finally, in addition to the expected excess of CpG mutations in the offspring of 26537, we observed a secondary enrichment in CpA contexts accounting for 28% of all calls in these children (**Figure 5C**). Non-CpG methylation, especially involving CpA, occurs genome-wide in oocytes of mice (Shirane et al. 2013) and human (Yu et al. 2017). Human oocytes show dramatic (>100%) increases in CpA methylation during oocyte maturation, while CpG methylation levels stay stable (Yu et al. 2017). Furthermore, *MBD4* is expressed in developing oocytes, and its expression is highly correlated to the oocyte expression levels of *DNMT3A*, a methyltransferase responsible for non-CpG methylation. Both *MBD4* and *DNMT3A* show a remarkable increase in transcription in human fetal ovary during 16 weeks gestation, when oocytes are first reaching the germinal vesicle stage (Galetzka et al. 2007). Taken together, we propose that the source of the germline hypermutation phenotype from *MBD4*-/- animals is not due to a loss of "maintenance" deamination repair that accumulates over the lifetime of the animal, but, instead due to a time-restricted window of damage, perhaps induced during *de novo* CpA methylation.

Methods

Genome sequencing and primary analysis. DNA was extracted from peripheral blood. Whole-genome sequencing libraries were constructed using a PCR-free method (Illumina). Libraries were assessed for quality and quantity by real time PCR and Tapestation (Agilent). Libraries were normalized and mixed, then sequenced with a 2 × 150bp paired end protocol on S4 flow cells using a NovaSeq 6000. Basecall files were converted to FASTQ files using bcl2fastq (Illumina). Reads were aligned to Mmul10 using BWA-MEM v.0.7.17-r1188 (Li 2013). The average mapped read depth of all genomes in the study was 23.6× (minimum 14.5×, maximum 35.8×, **Table S1**).

Downloaded from genome.cshlp.org on June 5, 2026 . Published by Cold Spring Harbor Laboratory Press

Mapping of sample IDs used in this study to Sequence Read Archive accessions and mGAP IDs are provided in **Table S8.**

De novo mutation identification and filtering. Initial candidate *de novo* mutations were identified using DeNovoGear (Ramu et al. 2013). We first selected sites with a probability of *de novo* single base pair mutation greater than 0.5. With this filter, DeNovoGear identified an average of 6,144 candidate mutations per individual (see **Table S1** for exact counts of mutations identified and filtered in each sample, and see **Supplemental Code** for exact filtering commands). These sites were further filtered to include sites where offspring and parent read depth was between 4 and 80 reads which removed an average of 3848 mutations. Further filtering included requiring offspring to have a variant allele frequency (VAF) between 35% and 70% for mutations on autosomes. Mutations on the X Chromosome should have a VAF between 35% and 70% for female offspring and greater than 70% for male offspring. Mutations on the Y Chromosome were excluded. Alternate allele reads must appear on both forward and reverse strands, with at least 2 reads supporting the alternate allele. This filtering eliminated an additional 888 mutations on average. We removed sites which fell in highly repetitive regions (UCSC simple repeats track) and in known segmental duplications (UCSC Segmental Duplication track) (Haeussler et al. 2019) which removed an average of 50 and 681 mutations, respectively. We removed mutations which were relatively common in the macaque population, as defined by having an allele frequency in mGAP greater than 0.01 which removed 587 mutations. We removed variants where either parent had more than one alternate allele read or greater than 10% variant allele frequency which removed 77 mutations on average. Finally, we used a binomial test with FDR correction (adjusted p-value < 0.05) to settle on our final set of *de novo* mutations which removed between 0-3 mutations.

Phasing de novo mutations. To confirm the hypermutation phenotype was a result of mutations coming from 26537, we phased mutations to either the maternal or paternal allele. Germline mutations were jointly called for each trio using HaplotypeCaller (Poplin et al.) and the *de novo* mutations were phased using PhaseMyDeNovo using default parameters (https://github.com/queenjobo/PhaseMyDeNovo). We were able to phase 50% of mutations to a single parent in hypermutators and 53% of mutations in non-hypermutators. Non-hypermutated animals had 74% of phaseable mutations phased to the paternal side, whereas the hypermutated offspring had 81% of phaseable mutations phased to the maternal side. The paternal-to-maternal ratio (□ value = 2.87) of the non-hypermutators was similar to that reported in literature (~3) (Wang et al. 2020), compared to 0.23 for hypermutators.

Mutation Signature Extraction. To determine the mutational processes resulting in the hypermutation in the six offspring of 26537, we compared the *de novo* mutation profiles by extracting mutational signatures using SigProfilerExtractor (v 1.1.7) and estimating

Downloaded from genome.cshlp.org on June 5, 2026 . Published by Cold Spring Harbor Laboratory Press

their contribution to the mutation burden (Islam et al. 2022). In order to accurately estimate signature proportions reflective of the rhesus genome, we renormalized the COSMIC v3.2 SBS human signature table based on the GRCh37 assembly to reflect the trinucleotide frequencies present in MMul10. Mutation signatures, as statistical summaries of context-dependent mutation, rely on the base composition of euchromatic DNA. The base composition of euchromatic DNA in GRCh37 is 99.9% the same as the newer GRCh38 assembly and subsequent releases; the other assembly differences involve changes to heterochromatic DNA (which we do not analyze here), a small amount of new sequence in assembly gaps, and the addition of small, alternate haplotypes at 178 loci, also not considered here (Guo et al. 2017). Thus an analysis based on a newer assembly should produce nearly identical results.

Linear modeling of mutation burden. We used poisson regression to model the effects of maternal age, paternal age, MBD4 genotype, and sequencing coverage on observed mutation counts. To account for differences in sequencing depth among trios, we calculated the number of callable bases for each sample using mosdepth (Pedersen and Quinlan 2018) with categories "NO_COVERAGE", "LOW_COVERAGE","CALLABLE", and "HIGH_COVERAGE" split as "--quantize 0:1:4:80:". The average number of callable bases per genome was 1.32 Gb. We find this number a reasonable proxy for power because normalizing the total number of mutations to the callable genome size produces a mutation rate estimate very similar to published values for macaque. We fit this model separately for the total number of mutations, as well as labeled mutation subsets: A>C, A>T, A>G, C>A, C>G, C>T, C>T in CpG context, C>A in CpG context. C>T in non-CpG context. To test for an effect of hypermutator status on the age-related accumulation of mutations, we fit a version of this model with an additional interaction term between hypermutator status and age.

Mapping a genetic cause of hypermutation in 26537. To identify a genetic cause for the hypermutation phenotype of 26537, we used two approaches. First, we used mGAP, the macaque genotype and phenotype database, to annotate allele frequencies for all variants in the joint VCF of individuals in the study. We then selected all coding variants from 26537 with mGAP minor allele frequency < 1%, homozygous in 26537, and not homozygous in the other trio parents. Sequencing alignments and gene annotations were then used to curate the resulting list, eliminating sequencing errors and annotation errors. Of the remaining variants, the only variant that affected a gene with a known function relating to DNA repair was MBD4.

Second, in order to confirm MBD4 as the most likely cause of the hypermutation phenotype, we carefully investigated 1378 candidate hypermutation genes (**Tables S3-S6**), chosen based on previous literature tying them to genome instability. These are: cancer predisposition genes in pediatric cancer (Zhang et al. 2015); DNA damage repair genes from the Cancer Genome Atlas (Knijnenburg et al. 2018); an in-house cancer

Downloaded from genome.cshlp.org on June 5, 2026 . Published by Cold Spring Harbor Laboratory Press

driver list ("Tarpey driver genes"); Cancer gene consensus genes from COSMIC (https://cancer.sanger.ac.uk/census); genes identified as involved in developmental disorders ("DDD_genes", https://www.deciphergenomics.org/ddd/ddgenes). The full list of genes interrogated is provided in Supplemental Table 3. The joint VCF of germline genotypes was annotated using VEP. We then extracted all mutations in these genes that were annotated by VEP as "HIGH" or "Moderate" impact, and identified gene candidates with biallelic damaging mutations in at least one dam or sire, which generated a short list of 23 genes (**Figure 4**).

Data Access

The whole-genome sequencing data generated in this study has been submitted to the NCBI BioProject database under BioProject number PRJNA382404.

The code and input data to reproduce the analyses and figures described here are available as **Supplemental Code** and from https://github.com/conradlab/StendahlEtAl .

Competing Interests Statement

The authors declare that they have no competing interests.

Acknowledgements

We thank Rhonda MacAllister and Carolyn Labriola for help with pathology and medical records, Katinka Vigh-Conrad for assistance with figures, and three anonymous reviewers for their comments. D.F.C. is supported by National Institute of Health Office of Directors (NIH/OD) Grant P51OD011092 (to the Oregon National Primate Research Center) and NHGRI grant U24HG012483.

Downloaded from genome.cshlp.org on June 5, 2026 . Published by Cold Spring Harbor Laboratory Press

References

Abascal F, Harvey LMR, Mitchell E, Lawson ARJ, Lensing SV, Ellis P, Russell AJC, Alcantara RE, Baez-Ortega A, Wang Y, et al. 2021. Somatic mutation landscapes at single-molecule resolution. *Nature* **593**: 405–410.

Bergeron LA, Besenbacher S, Bakker J, Zheng J, Li P, Pacheco G, Sinding M-HS, Kamilari M, Gilbert MTP, Schierup MH, et al. 2021. The germline mutational process in rhesus macaque and its implications for phylogenetic dating. *Gigascience* **10**. http://dx.doi.org/10.1093/gigascience/giab029.

Bimber BN, Yan MY, Peterson SM. 2019. mGAP: the macaque genotype and phenotype resource, a framework for accessing and interpreting macaque variant data, and identifying new models of …. *Biomed Chromatogr*. https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-019-5559-7.

Conrad DF, Keebler JEM, DePristo MA, Lindsay SJ, Zhang Y, Casals F, Idaghdour Y, Hartl CL, Torroja C, Garimella KV, et al. 2011. Variation in genome-wide mutation rates within and between human families. *Nat Genet* **43**: 712–714.

Crow JF. 2000. The origins, patterns and implications of human spontaneous mutation. *Nat Rev Genet* **1**: 40–47.

Derrien A-C, Rodrigues M, Eeckhoutte A, Dayot S, Houy A, Mobuchon L, Gardrat S, Lequin D, Ballet S, Pierron G, et al. 2021. Germline MBD4 Mutations and Predisposition to Uveal Melanoma. *J Natl Cancer Inst* **113**: 80–87.

Galetzka D, Weis E, Tralau T, Seidmann L, Haaf T. 2007. Sex-specific windows for high mRNA expression of DNA methyltransferases 1 and 3A and methyl-CpG-binding domain proteins 2 and 4 in human fetal gonads. *Mol Reprod Dev* **74**: 233–241.

Greenberg MVC, Bourc’his D. 2019. The diverse roles of DNA methylation in mammalian development and disease. *Nat Rev Mol Cell Biol* **20**: 590–607.

Guo Y, Dai Y, Yu H, Zhao S, Samuels DC, Shyr Y. 2017. Improvements and impacts of GRCh38 human reference on high throughput sequencing data analysis. *Genomics* **109**: 83–90.

Haeussler M, Zweig AS, Tyner C, Speir ML, Rosenbloom KR, Raney BJ, Lee CM, Lee BT, Hinrichs AS, Gonzalez JN, et al. 2019. The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res* **47**: D853–D858.

Islam SMA, Díaz-Gay M, Wu Y, Barnes M, Vangara R, Bergstrom EN, He Y, Vella M, Wang J, Teague JW, et al. 2022. Uncovering novel mutational signatures by de novo extraction with SigProfilerExtractor. *Cell Genom* **2**: None.

Jónsson H, Sulem P, Kehr B, Kristmundsdottir S, Zink F, Hjartarson E, Hardarson MT, Hjorleifsson KE, Eggertsson HP, Gudjonsson SA, et al. 2017. Parental influence on human germline de novo mutations in 1,548 trios from Iceland. *Nature* **549**: 519–522.

Downloaded from genome.cshlp.org on June 5, 2026 . Published by Cold Spring Harbor Laboratory Press

Kaplanis J, Ide B, Sanghvi R, Neville M, Danecek P, Coorens T, Prigmore E, Short P, Gallone G, McRae J, et al. 2022. Genetic and chemotherapeutic influences on germline hypermutation. *Nature* **605**: 503–508.

Knijnenburg TA, Wang L, Zimmermann MT, Chambwe N, Gao GF, Cherniack AD, Fan H, Shen H, Way GP, Greene CS, et al. 2018. Genomic and Molecular Landscape of DNA Damage Repair Deficiency across The Cancer Genome Atlas. *Cell Rep* **23**: 239-254.e6.

Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, Gudjonsson SA, Sigurdsson A, Jonasdottir A, Jonasdottir A, et al. 2012. Rate of de novo mutations and the importance of father's age to disease risk. *Nature* **488**: 471–475.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-bioGN]*. http://arxiv.org/abs/1303.3997.

Mahyari E, Vigh-Conrad KA, Daube C, Lima AC, Guo J, Carrell DT, Hotaling JM, Aston KI, Conrad DF. 2023. The Human Infertility Single-cell Testis Atlas (HISTA): An interactive molecular scRNA-Seq reference of the human testis. *bioRxiv* 2023.09.23.558896. https://www.biorxiv.org/content/biorxiv/early/2023/09/24/2023.09.23.558896 (Accessed October 18, 2023).

Menezo Y Jr, Russo G, Tosti E, El Mouatassim S, Benkhalifa M. 2007. Expression profile of genes coding for DNA repair in human oocytes using pangenomic microarrays, with a special focus on ROS linked decays. *J Assist Reprod Genet* **24**: 513–520.

Millar CB, Guy J, Sansom OJ, Selfridge J, MacDougall E, Hendrich B, Keightley PD, Bishop SM, Clarke AR, Bird A. 2002. Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. *Science* **297**: 403–405.

Moore L, Cagan A, Coorens THH, Neville MDC, Sanghvi R, Sanders MA, Oliver TRW, Leongamornlert D, Ellis P, Noorani A, et al. 2021. The mutational landscape of human somatic and germline cells. *Nature* **597**: 381–386.

Palles C, West HD, Chew E, Galavotti S, Flensburg C, Grolleman JE, Jansen EAM, Curley H, Chegwidden L, Arbe-Barnes EH, et al. 2022. Germline MBD4 deficiency causes a multi-tumor predisposition syndrome. *Am J Hum Genet* **109**: 953–960.

Pedersen BS, Quinlan AR. 2018. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* **34**: 867–868.

Pidugu LS, Bright H, Lin W-J, Majumdar C, Van Ostrand RP, David SS, Pozharski E, Drohat AC. 2021. Structural Insights into the Mechanism of Base Excision by MBD4. *J Mol Biol* **433**: 167097.

Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA, Kling DE, Gauthier LD, Levy-Moonshine A, Roazen D, et al. Scaling accurate genetic variant discovery to tens of thousands of samples. http://dx.doi.org/10.1101/201178.

Rahbari R, UK10K Consortium, Wuster A, Lindsay SJ, Hardwick RJ, Alexandrov LB, Al Turki S, Dominiczak A, Morris A, Porteous D, et al. 2016. Timing, rates and spectra of human germline mutation. *Nat Genet* **48**: 126–133.

Downloaded from genome.cshlp.org on June 5, 2026 . Published by Cold Spring Harbor Laboratory Press

Ramu A, Noordam MJ, Schwartz RS, Wuster A, Hurles ME, Cartwright RA, Conrad DF. 2013. DeNovoGear: de novo indel and point mutation discovery and phasing. *Nat Methods* **10**: 985–987.

Robinson PS, Thomas LE, Abascal F, Jung H, Harvey LMR, West HD, Olafsson S, Lee BCH, Coorens THH, Lee-Six H, et al. 2022. Inherited MUTYH mutations cause elevated somatic mutation rates and distinctive mutational signatures in normal human cells. *Nat Commun* **13**: 3949.

Sanders MA, Chew E, Flensburg C, Zeilemaker A, Miller SE, Al Hinai AS, Bajel A, Luiken B, Rijken M, Mclennan T, et al. 2018. MBD4 guards against methylation damage and germ line deficiency predisposes to clonal hematopoiesis and early-onset AML. *Blood* **132**: 1526–1534.

Sasani TA, Ashbrook DG, Beichman AC, Lu L, Palmer AA, Williams RW, Pritchard JK, Harris K. 2022. A natural mutator allele shapes mutation spectrum variation in mice. *Nature* **605**: 497–502.

Shirane K, Toh H, Kobayashi H, Miura F, Chiba H, Ito T, Kono T, Sasaki H. 2013. Mouse oocyte methylomes at base resolution reveal genome-wide accumulation of non-CpG methylation and role of DNA methyltransferases. *PLoS Genet* **9**: e1003439.

Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, Boutselakis H, Cole CG, Creatore C, Dawson E, et al. 2019. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res* **47**: D941–D947.

Wang RJ, Thomas GWC, Raveendran M, Harris RA, Doddapaneni H, Muzny DM, Capitanio JP, Radivojac P, Rogers J, Hahn MW. 2020. Paternal age in rhesus macaques is positively associated with germline mutation accumulation but not with measures of offspring sociability. *Genome Res* **30**: 826–834.

Wong E, Yang K, Kuraguchi M, Werling U, Avdievich E, Fan K, Fazzari M, Jin B, Brown AMC, Lipkin M, et al. 2002. Mbd4 inactivation increases Cright-arrowT transition mutations and promotes gastrointestinal tumor formation. *Proc Natl Acad Sci U S A* **99**: 14937–14942.

Wong WSW, Solomon BD, Bodian DL, Kothiyal P, Eley G, Huddleston KC, Baker R, Thach DC, Iyer RK, Vockley JG, et al. 2016. New observations on maternal age effect on germline de novo mutations. *Nat Commun* **7**: 10486.

Yu B, Dong X, Gravina S, Kartal Ö, Schimmel T, Cohen J, Tortoriello D, Zody R, Hawkins RD, Vijg J. 2017. Genome-wide, single-cell DNA methylomics reveals increased non-CpG methylation during human oocyte maturation. *Stem Cell Reports* **9**: 397–407.

Zhang J, Walsh MF, Wu G, Edmonson MN, Gruber TA, Easton J, Hedges D, Ma X, Zhou X, Yergeau DA, et al. 2015. Germline Mutations in Predisposition Genes in Pediatric Cancer. *N Engl J Med* **373**: 2336–2346.

Figure Captions

Figure 1. Identification of a biallelic frameshift mutation in MBD4 in 26537. Location of the mutation with respect to genome (Mmul10) and protein (ENSMMUG00000012723) coordinates.

Downloaded from genome.cshlp.org on June 5, 2026 . Published by Cold Spring Harbor Laboratory Press

Figure 2. Overview of families sequenced for the study.

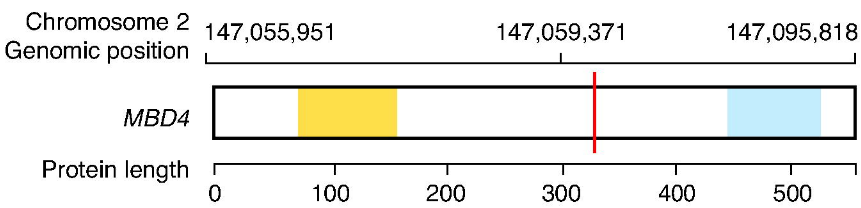
Figure 3. Children of 26537 show rates of germline *de novo* mutation (DNM) 4-6-fold higher than children from control parents. (A) relationship between paternal age and number of DNMs observed in offspring, (B) relationship between maternal age and number of DNMs observed in offspring. The observed paternal and maternal age effects seen in non-26537 offspring largely adhere to previous literature [9-11] . Best fit lines and numerical values within each panel correspond to a standard linear model fit to control (non hypermutator) data only. (C) Percentage of de novo mutations observed on paternally and maternally inherited chromosomes. Read-based phasing of DNMs shows an excess of maternally-derived DNMs in offspring of 26537 not observed in control offspring. The alpha estimate for each set of phased mutations is shown. Alpha=ratio of paternal:maternal mutations.

Figure 4. MBD4 is the only mutated DNA repair gene private to 26537.
To confirm MBD4 is the most likely candidate for the hypermutation phenotype, we searched for biallelic damaging mutations across 1,381 candidate genes for germline hypermutation derived from the literature (**Table S3**). We identified candidate mutations in 23 genes, shown here. *MBD4* was the only gene specifically mutated in 26537 and not other parents in the study.

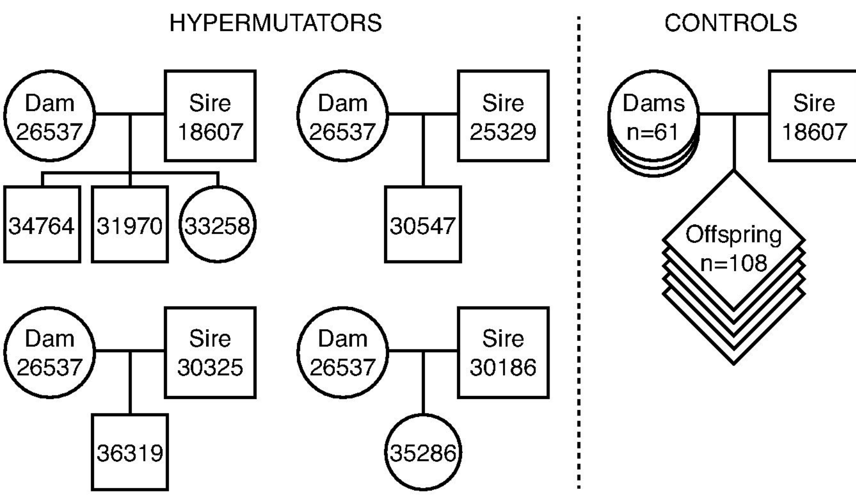
Figure 5. Mutation signature analysis of *MBD4* germline mutations. (A) Unlike germline mutational signatures in control individuals, which are mostly signature 5 with smaller proportion of signature 1, the majority of mutations in the hypermutated offspring were attributed to signature 1 with very small contribution from signature 5. (B) Of the 6 possible substitution types, C>T was the only based substitution with a significant excess in offspring of 26537 (Hypermutator) compared to offspring from control families (Non-hypermutator). Most of these C>T changes were in a CpG context, consistent with the known function of MBD4 in repairing spontaneous methyl-cytosine deamination. (C) The remaining excess of C>T changes were observed at CpA positions ($p=3 \times 10^{-4}$ for contrast of hypermutator vs control by poisson regression).

Figure 6. 26537 showed sustained eosinophilia at age 14. Four independent CBCs run over several weeks showed 16%-22.7% eosinophils (red arrows). These numbers were in the top 0.5% of over 25,000 CBCs run at ONPRC since 2016 (histogram).

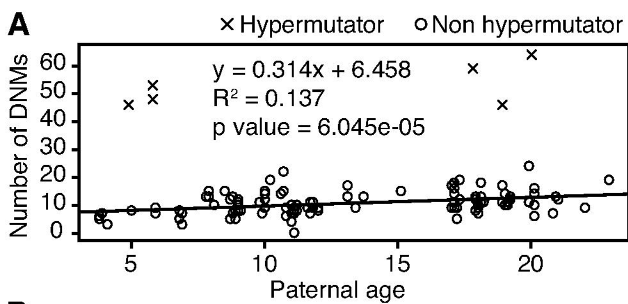
Downloaded from genome.cshlp.org on June 5, 2026 . Published by Cold Spring Harbor Laboratory Press

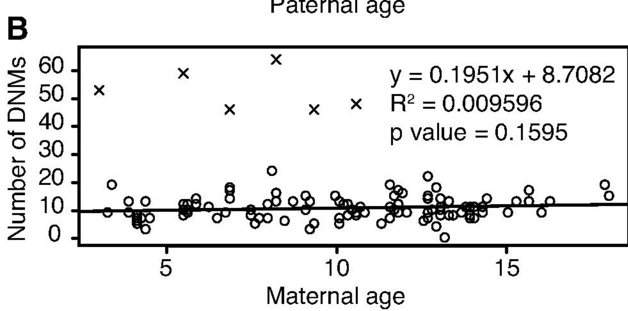


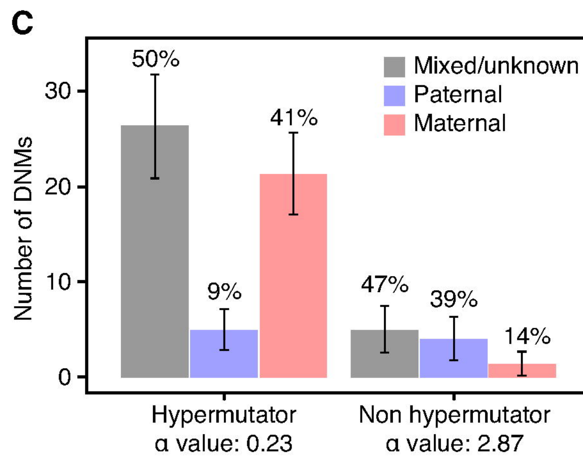
Downloaded from genome.cshlp.org on June 5, 2026 . Published by Cold Spring Harbor Laboratory Press



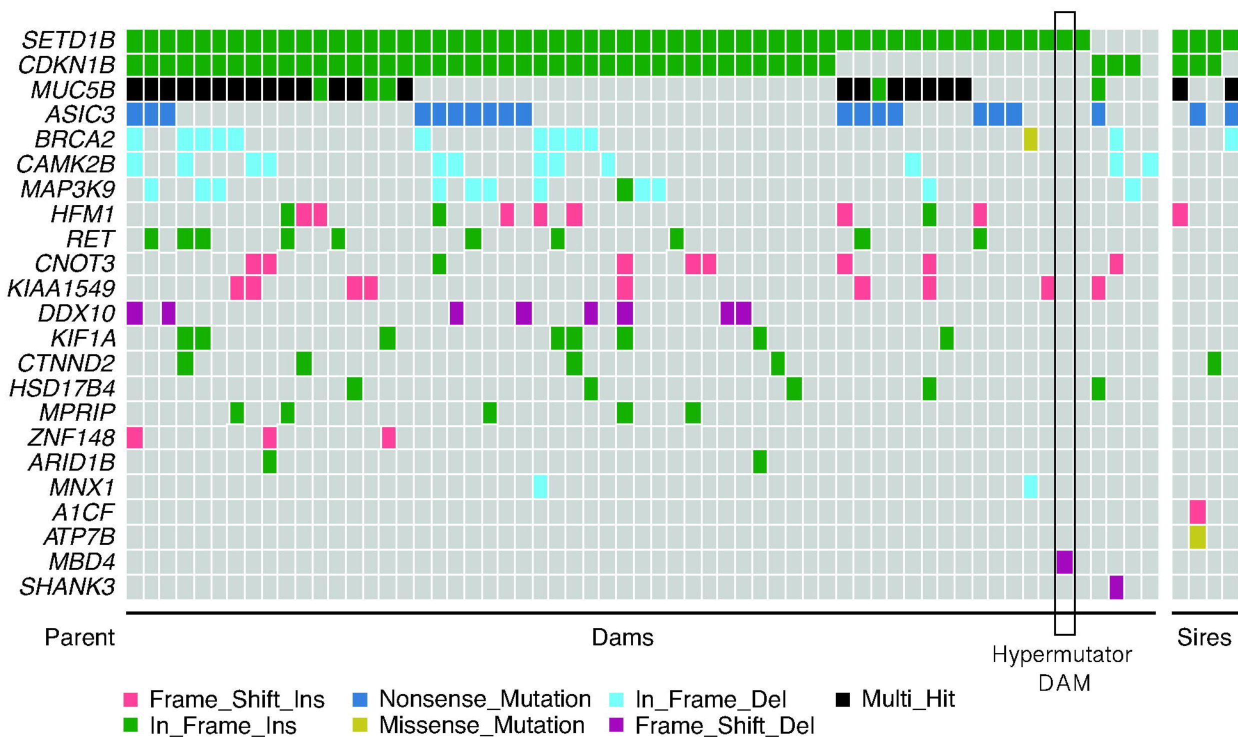
Downloaded from genome.cshlp.org on June 5, 2026 . Published by Cold Spring Harbor Laboratory Press



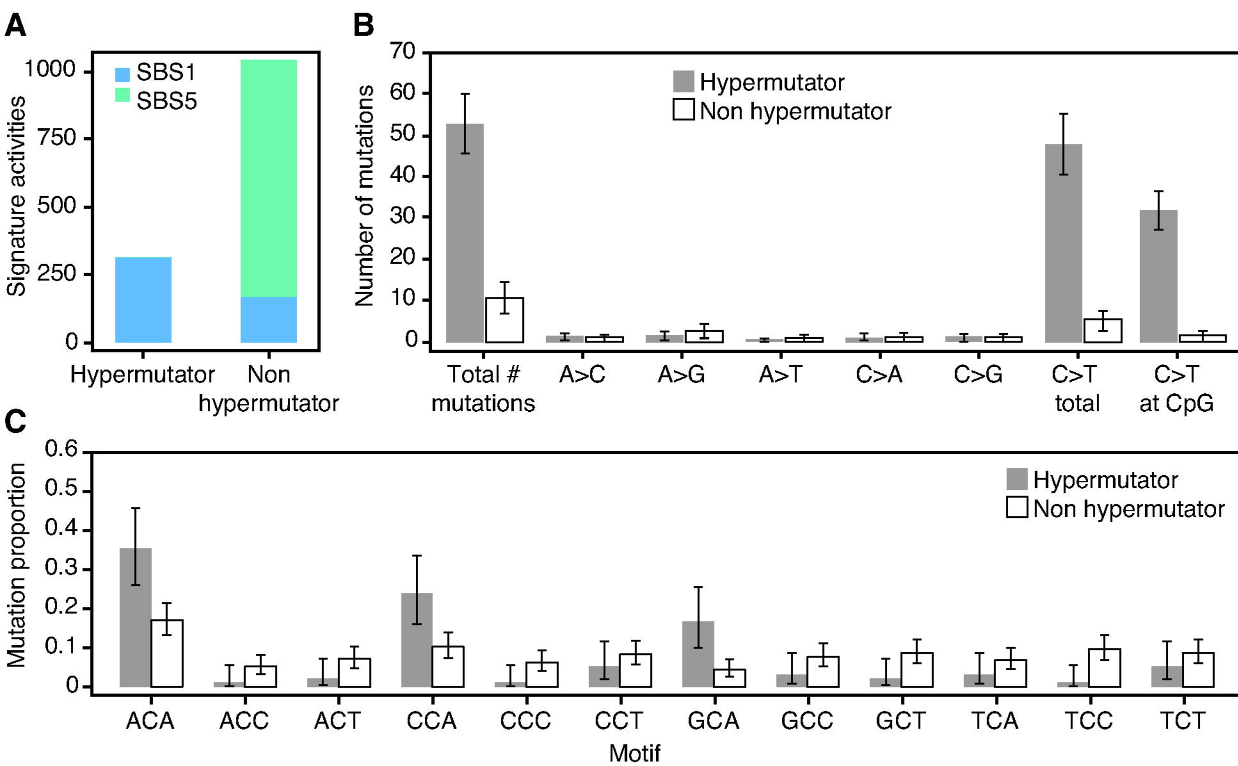


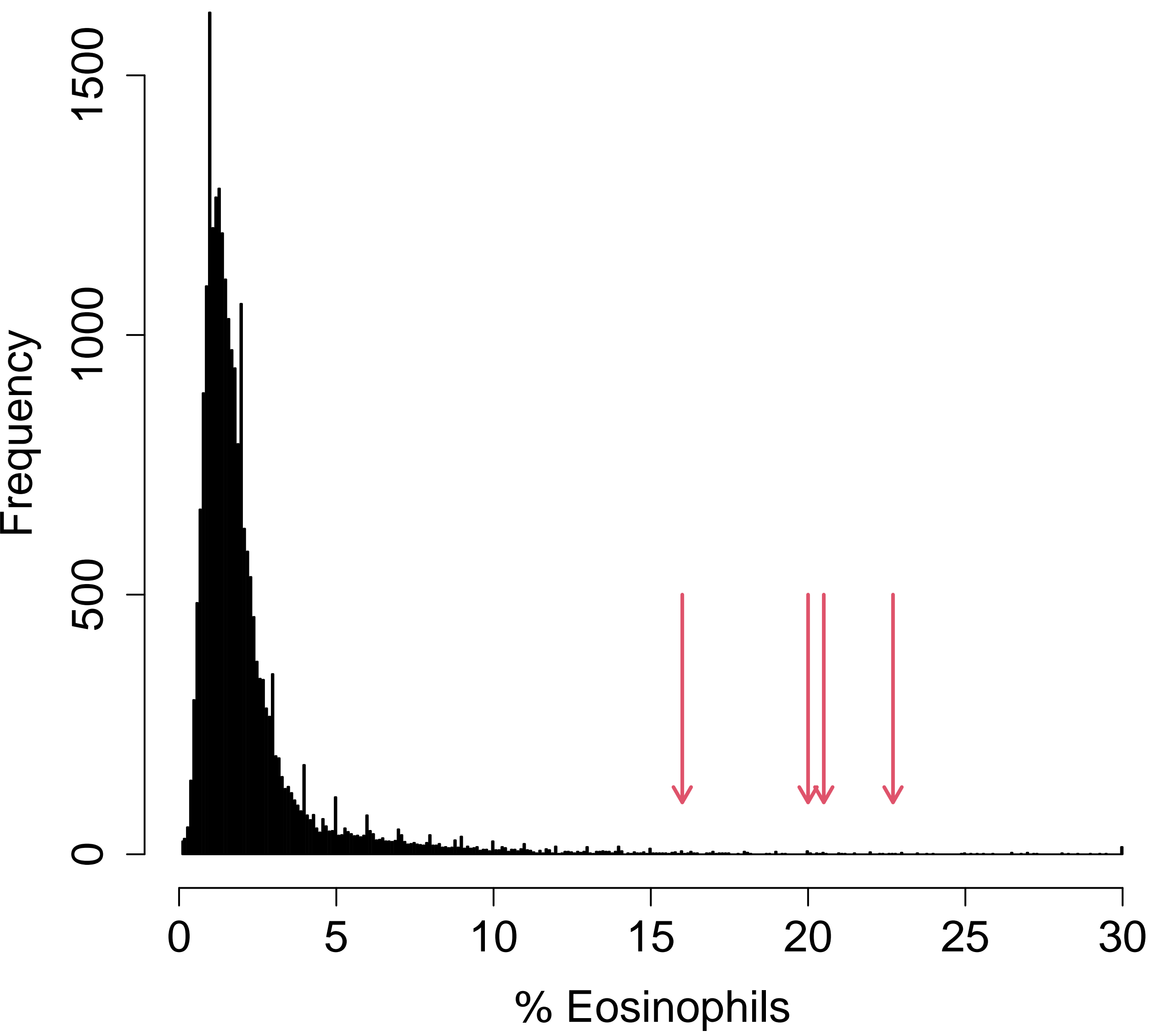


Downloaded from genome.cshlp.org on June 5, 2026 . Published by Cold Spring Harbor Laboratory Press



Downloaded from genome.cshlp.org on June 5, 2026 . Published by Cold Spring Harbor Laboratory Press




Frequency
0
500
1000
1500
0
5
10
15
20
25
30
% Eosinophils