

Epistasis between mutator alleles contributes to germline mutation spectra variability in laboratory mice

Reviewed Preprint

Revised by authors after peer review.

[About eLife's process](#)

Reviewed preprint version 2

January 10, 2024 (this version)

Reviewed preprint version 1

July 20, 2023

Sent for peer review

May 12, 2023

Posted to preprint server

April 25, 2023

Thomas A. Sasani, Aaron R. Quinlan , Kelley Harris 

Department of Human Genetics, University of Utah • Department of Human Genetics, University of Utah;
Department of Biomedical Informatics, University of Utah • Department of Genome Sciences, University of Washington

 https://en.wikipedia.org/wiki/Open_access

 Copyright information

Abstract

Maintaining germline genome integrity is essential and enormously complex. Although many proteins are involved in DNA replication, proofreading, and repair [1], *mutator alleles* have largely eluded detection in mammals.

DNA replication and repair proteins often recognize sequence motifs or excise lesions at specific nucleotides. Thus, we might expect that the spectrum of *de novo* mutations — the frequencies of C>T, A>G, etc. — will differ between genomes that harbor either a mutator or wild-type allele. Previously, we used quantitative trait locus mapping to discover candidate mutator alleles in the DNA repair gene *Mutyh* that increased the C>A germline mutation rate in a family of inbred mice known as the BXDs [2,3].

In this study we developed a new method to detect alleles associated with mutation spectrum variation and applied it to mutation data from the BXDs. We discovered an additional C>A mutator locus on chromosome 6 that overlaps *Ogg1*, a DNA glycosylase involved in the same base-excision repair network as *Mutyh* [4]. Its effect depended on the presence of a mutator allele near *Mutyh*, and BXDs with mutator alleles at both loci had greater numbers of C>A mutations than those with mutator alleles at either locus alone. Our new methods for analyzing mutation spectra reveal evidence of epistasis between germline mutator alleles and may be applicable to mutation data from humans and other model organisms.

eLife assessment

By developing a novel method for detecting genetic variants associated with germline mutation spectrum variation, this **important** study identifies a new "mutator" locus in a population of inbred mouse strains, although the causal gene(s) and allele(s) within this locus remain uncertain. The authors further demonstrate that this new mutator locus interacts epistatically with a previously identified mutator allele on C>A mutation rate, showcasing the complexity of the genetic basis underlying variation in mutation rate and spectrum. Evidence for major findings in this paper is **convincing**, and the new method has the potential to be applicable to a variety of experimental systems and natural populations.

Introduction

Germline mutation rates reflect the complex interplay between DNA proofreading and repair pathways, exogenous sources of DNA damage, and life-history traits. For example, parental age is an important determinant of mutation rate variability; in many mammalian species, the number of germline *de novo* mutations observed in offspring increases as a function of paternal and maternal age [5,6,7,8,9]. Rates of germline mutation accumulation are also variable across human families [6,10], likely due to either genetic variation or differences in environmental exposures. Although numerous protein-coding genes contribute to the maintenance of genome integrity, genetic variants that increase germline mutation rates, known as *mutator alleles*, have proven difficult to discover in mammals.

The dearth of observed germline mutators in mammalian genomes is not necessarily surprising, since alleles that lead to elevated germline mutation rates would likely have deleterious consequences and be purged by negative selection if their effect sizes are large [11]. Moreover, germline mutation rates are relatively low, and direct mutation rate measurements require whole-genome sequencing data from both parents and their offspring. As a result, large-scale association studies — which have been used to map the contributions of common genetic variants to many complex traits — are not currently well-powered to investigate the polygenic architecture of germline mutation rates [12].

Despite these challenges, less traditional strategies have been used to identify a small number of mutator alleles in humans, macaques [13], and mice. By focusing on families with rare genetic diseases, a recent study discovered two mutator alleles that led to significantly elevated rates of *de novo* germline mutation in human genomes [14]. Other groups have observed mutator phenotypes in the germlines and somatic tissues of adults who carry cancer-predisposing inherited mutations in the POLE/POLD1 exonucleases [15,16]. Candidate mutator loci were also found by identifying human haplotypes from the Thousand Genomes Project with excess counts of derived alleles in genomic windows [17].

In mice, a germline mutator allele was recently discovered by sequencing a large family of inbred mice [2]. Commonly known as the BXDs, these recombinant inbred lines (RILs) were derived from either F2 or advanced intercrosses of C57BL/6J and DBA/2J, two laboratory strains that exhibit significant differences in their germline mutation spectra [3,18]. The BXDs were maintained via brother-sister mating for up to 180 generations, and each BXD therefore accumulated hundreds or thousands of germline mutations on a nearly-homozygous linear mosaic of parental *B* and *D* haplotypes. Due to their husbandry in a controlled laboratory setting, the BXDs were largely free from confounding by environmental heterogeneity, and the effects of selection on *de novo* mutations were attenuated by strict inbreeding [19].

In this previous study, whole-genome sequencing data from the BXD family were used to map a quantitative trait locus (QTL) for the C>A mutation rate [2]. Germline C>A mutation rates were nearly 50% higher in mice with *D* haplotypes at the QTL, likely due to genetic variation in the DNA glycosylase *Mutyh* that reduced the efficacy of oxidative DNA damage repair. Pathogenic variants of *Mutyh* also appear to act as mutators in human germline and somatic tissues [16,20]. Importantly, the QTL did not reach genome-wide significance in a scan for variation in overall germline mutation rates, which were only modestly higher in BXDs with *D* alleles, demonstrating the utility of mutation spectrum analysis for mutator allele discovery. Close examination of the mutation spectrum is likely to be broadly useful for detecting mutator alleles, as genes involved in DNA proofreading and repair often recognize particular sequence motifs or excise specific types of DNA lesions [21]. Mutation spectra are usually defined in terms of *k*-mer nucleotide context; the

1-mer mutation spectrum, for example, consists of 6 mutation types after collapsing by strand complement (C>T, C>A, C>G, A>T, A>C, A>G), while the 3-mer mutation spectrum contains 96 (each of the 1-mer mutations partitioned by trinucleotide context).

Although mutation spectrum analysis can enable the discovery of mutator alleles that affect the rates of specific mutation types, early implementations of this strategy have suffered from a few drawbacks. For example, performing association tests on the rates or fractions of every k -mer mutation type can quickly incur a substantial multiple testing burden. Since germline mutation rates are generally quite low, estimates of k -mer mutation type frequencies from individual samples can also be noisy and imprecise. In populations of RILs, inbreeding duration can also vary substantially; for example, some BXDs were inbred for only 20 generations, while others were inbred for nearly 200. As a result, the variance of individual k -mer mutation rate estimates in those populations will be higher than if all samples were inbred for the same duration. We were therefore motivated to develop a statistical method that could overcome the sparsity of *de novo* mutation spectra, eliminate the need to test each k -mer mutation type separately, and enable sensitive detection of alleles that influence the germline mutation spectrum.

Here, we present a new mutation spectrum association test, called “aggregate mutation spectrum distance,” that minimizes multiple testing burdens and mitigates the challenges of sparsity in *de novo* mutation datasets. We leverage this method to re-analyze germline mutation data from the BXD family and find compelling evidence for a second mutator allele that was not detected using previous approaches. The new allele appears to interact epistatically with the mutator that was previously discovered in the BXDs, further augmenting the C>A germline mutation rate in a subset of inbred mice. Our observation of epistasis suggests that mild DNA repair deficiencies can compound one another, as mutator alleles chip away at the redundant systems that collectively maintain germline integrity.

Results

A novel method for detecting mutator alleles

We developed a statistical method, termed “aggregate mutation spectrum distance” (AMSD), to detect loci that are associated with mutation spectrum variation in recombinant inbred lines (RILs) (**Figure 1** [↗](#); *Materials and Methods*). Our approach leverages the fact that mutator alleles often leave behind distinct and detectable impressions on the *mutation spectrum*, even if they increase the overall mutation rate by a relatively small amount. Given a population of haplotypes, we assume that each has been genotyped at the same collection of biallelic loci and that each harbors *de novo* mutations which have been partitioned by k -mer context (**Figure 1** [↗](#)). At every locus, we calculate a cosine distance between the aggregate mutation spectra of haplotypes that inherited either parental allele. Using permutation tests, we then identify loci at which those distances are larger than what we’d expect by random chance. To account for polygenic effects on the mutation process that might be shared between BXDs, we also regress the cosine distance at each marker against the genetic similarity between haplotype groups, and assess significance using the fitted residuals (which we call the “adjusted” cosine distances) (*Materials and Methods*).

Using simulated data, we find that our method’s power is primarily limited by the initial mutation rate of the k -mer mutation type affected by a mutator allele and the total number of *de novo* mutations used to detect it (**Figure 1-figure supplement 1**). Given 100 haplotypes with an average of 500 *de novo* germline mutations each, AMSD has approximately 90% power to detect a mutator allele that increases the C>A *de novo* mutation rate by as little as 20%. However, the approach has less than 20% power to detect a mutator of identical effect size that augments the C>G mutation rate, since C>G mutations are expected to make up a smaller fraction of all *de novo* germline mutations to begin with. Simulations also demonstrate that our approach is well-powered to detect

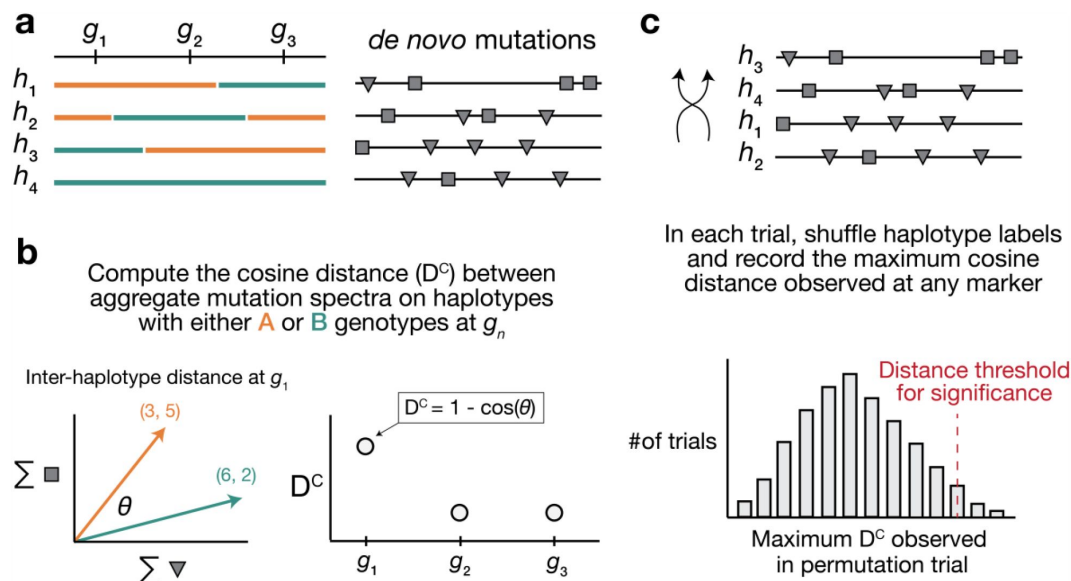


Figure 1

Overview of aggregate mutation spectrum distance method for discovering mutator alleles.

a) A population of four haplotypes has been genotyped at three informative markers (g_1 through g_3); each haplotype also harbors unique *de novo* germline mutations. In practice, *de novo* mutations are partitioned by *k*-mer context; for simplicity in this toy example, *de novo* mutations are simply classified into two possible mutation types (grey squares represent C>(A/T/G) mutations, while grey triangles represent A>(C/T/G) mutations). **b)** At each informative marker g_n , we calculate the total number of each mutation type observed on haplotypes that carry either parental allele (i.e., the aggregate mutation spectrum) using all genome-wide *de novo* mutations. For example, haplotypes with A (orange) genotypes at g_1 carry a total of three “triangle” mutations and five “square” mutations, and haplotypes with B (green) genotypes carry a total of six triangle and two square mutations. We then calculate the cosine distance between the two aggregate mutation spectra, which we call the “aggregate mutation spectrum distance.” Cosine distance can be defined as $1 - \cos(\theta)$, where θ is the angle between two vectors; in this case, the two vectors are the two aggregate spectra. We repeat this process for every informative marker g_n . **c)** To assess the significance of any distance peaks in b), we perform permutation tests. In each of *N* permutations, we shuffle the haplotype labels associated with the *de novo* mutation data, run a genome-wide distance scan, and record the maximum cosine distance encountered at any locus in the scan. Finally, we calculate the $1 - p$ percentile of the distribution of those maximum distances to obtain a genome-wide cosine distance threshold at the specified value of *p*.

large-effect mutator alleles (e.g., those that increase the mutation rate of a specific k -mer by 50%), even with a relatively small number of mutations per haplotype (Figure 1-figure supplement 1). Both AMSD and traditional quantitative trait locus (QTL) mapping have similar power to detect alleles that augment the rates of individual 1-mer mutation types (Figure 1-figure supplement 2), but AMSD has a number of potential advantages for mutator allele discovery. For example, we find that AMSD is better-powered than QTL mapping when the number of simulated *de novo* mutations is allowed to vary (by a factor of 20) across haplotypes (Figure 1-figure supplement 3) and when mutator allele frequencies are less than 50% (Figure 1-figure supplement 4). However, we also caution that many of the parameters used in our simulations are specific to the BXD mice (e.g., numbers of haplotypes, average numbers of mutations, expected allele frequencies at markers), and do not necessarily reflect the power of AMSD on other populations.

Re-identifying a mutator allele on chromosome 4 in the BXDs

We applied our aggregate mutation spectrum distance method to 117 BXDs (*Materials and Methods*) with a total of 65,552 *de novo* germline mutations [2]. Using mutation data that were partitioned by 1-mer nucleotide context, we discovered a locus on chromosome 4 that was significantly associated with mutation spectrum variation (Figure 2a; maximum adjusted cosine distance of 1.20×10^{-2} at marker ID rs27509845; position 118.28 Mbp in GRCm38/mm10 coordinates; 90% bootstrap confidence interval from 114.79 - 118.75 Mbp).

Using quantitative trait locus (QTL) mapping, we previously identified a nearly-identical locus on chromosome 4 that was significantly associated with the C>A germline mutation rate in the BXDs [2]. This locus overlapped 21 protein-coding genes that were annotated by the Gene Ontology as being involved in “DNA repair,” but only one of those genes contained nonsynonymous differences between the two parental strains: *Mutyh*. *Mutyh* encodes a protein involved in the base-excision repair of 8-oxoguanine (8-oxoG), a DNA lesion caused by oxidative damage, and prevents the accumulation of C>A mutations [4,22,23]. C>A germline mutation fractions are nearly 50% higher in BXDs that inherit *D* genotypes at marker ID rs27509845 (the marker at which we observed the highest adjusted cosine distance on chromosome 4) than in those that inherit *B* genotypes (Figure 3) [2].

An additional germline mutator allele on chromosome 6

After confirming that AMSD could recover the mutator locus overlapping *Mutyh*, we tested its ability to identify additional mutator loci in the BXDs. To eliminate potential confounding of the mutation spectrum landscape by the large-effect mutator locus on chromosome 4, we performed AMSD scans that were conditional on the presence of either *D* or *B* alleles at rs27509845. We hypothesized that such conditioning might reveal epistatic interactions between alleles at the chromosome 4 locus and mutator alleles elsewhere in the genome. Specifically, we divided the BXDs into those with either *D* ($n = 66$) or *B* ($n = 44$) genotypes at rs27509845 ($n = 7$ BXDs were heterozygous) and ran an aggregate mutation spectrum distance scan using each group separately (Figure 2b-c). We excluded the BXD68 RIL from these scans, since we previously found that BXD68 harbors a strain-private C>A mutator allele of even larger effect than that of a *D* allele at rs27509845 [2].

Using the BXDs with *D* genotypes at rs27509845, we identified a locus on chromosome 6 that was significantly associated with mutation spectrum variation (Figure 2b; maximum adjusted cosine distance of 3.69×10^{-3} at marker rs46276051; position 111.27 Mbp in GRCm38/mm10 coordinates; 90% bootstrap confidence interval from 95.01 - 114.02 Mbp). This signal was specific to BXDs with *D* genotypes at the rs27509845 locus, as we did not observe any new mutator loci after performing an AMSD scan using BXDs with *B* genotypes at rs27509845 (Figure 2c). The peak markers on chromosome 4 and 6 did not exhibit strong linkage disequilibrium ($R^2 = 4 \times 10^{-5}$). We

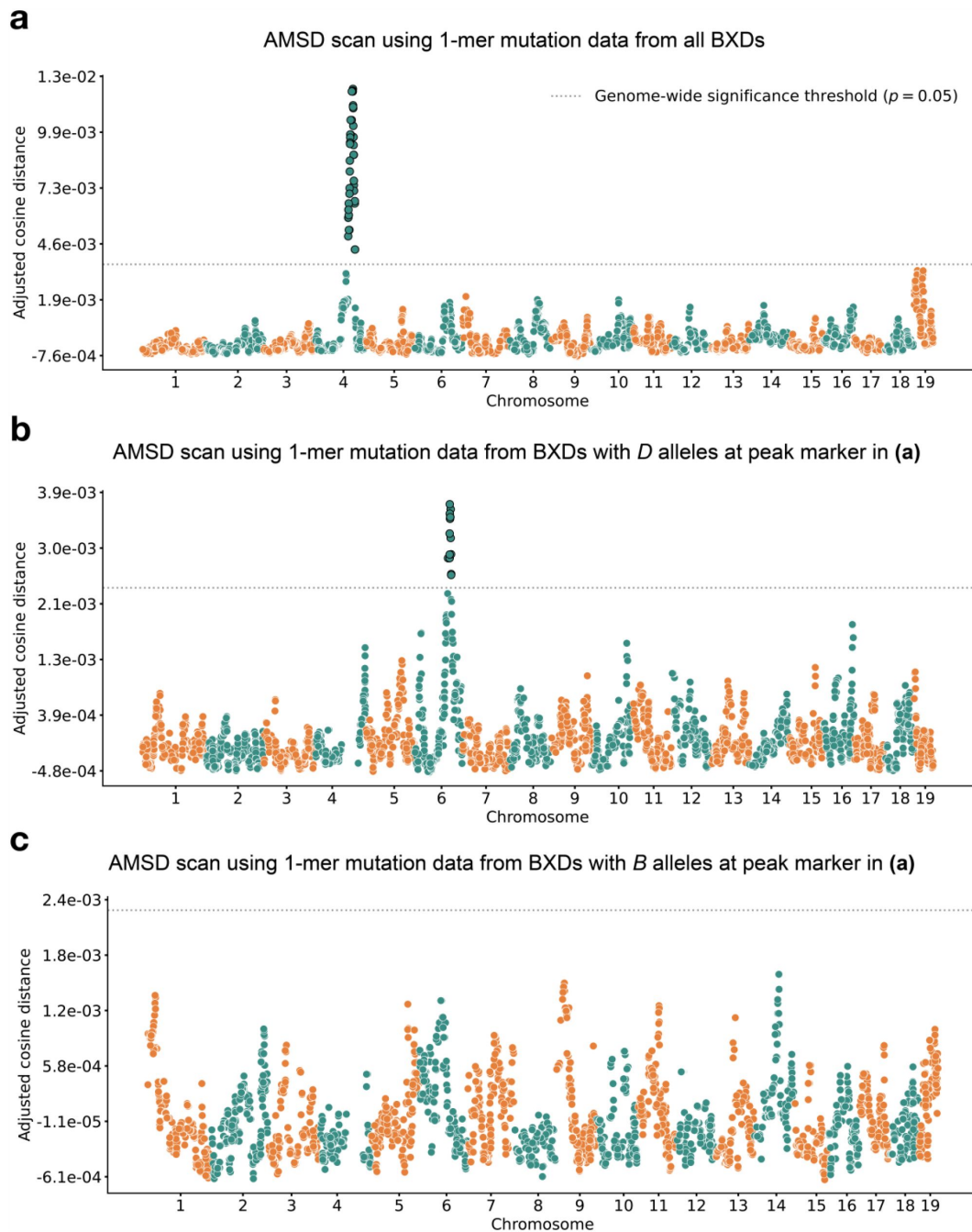


Figure 2

Results of aggregate mutation spectrum distance scans in the BXDs.

a) Adjusted cosine distances between aggregate 1-mer de novo mutation spectra on BXD haplotypes ($n = 117$ haplotypes; 65,552 total mutations) with either D or B alleles at 7,128 informative markers. Cosine distance threshold at $p = 0.05$ was calculated by performing 10,000 permutations of the BXD mutation data, and is shown as a dotted grey line. **b)** Adjusted cosine distances between aggregate 1-mer de novo mutation spectra on BXD haplotypes with D alleles at rs27509845 ($n = 66$ haplotypes; 42,171 total mutations) and either D or B alleles at 6,957 informative markers. Cosine distance threshold at $p = 0.05$ was calculated by performing 10,000 permutations of the BXD mutation data, and is shown as a dotted grey line. **c)** Adjusted cosine distances between aggregate 1-mer de novo mutation spectra on BXD haplotypes with B alleles at rs27509845 ($n = 44$ haplotypes; 22,645 total mutations) and either D or B alleles at 6,957 informative markers. Cosine distance threshold at $p = 0.05$ was calculated by performing 10,000 permutations of the BXD mutation data, and is shown as a dotted grey line.

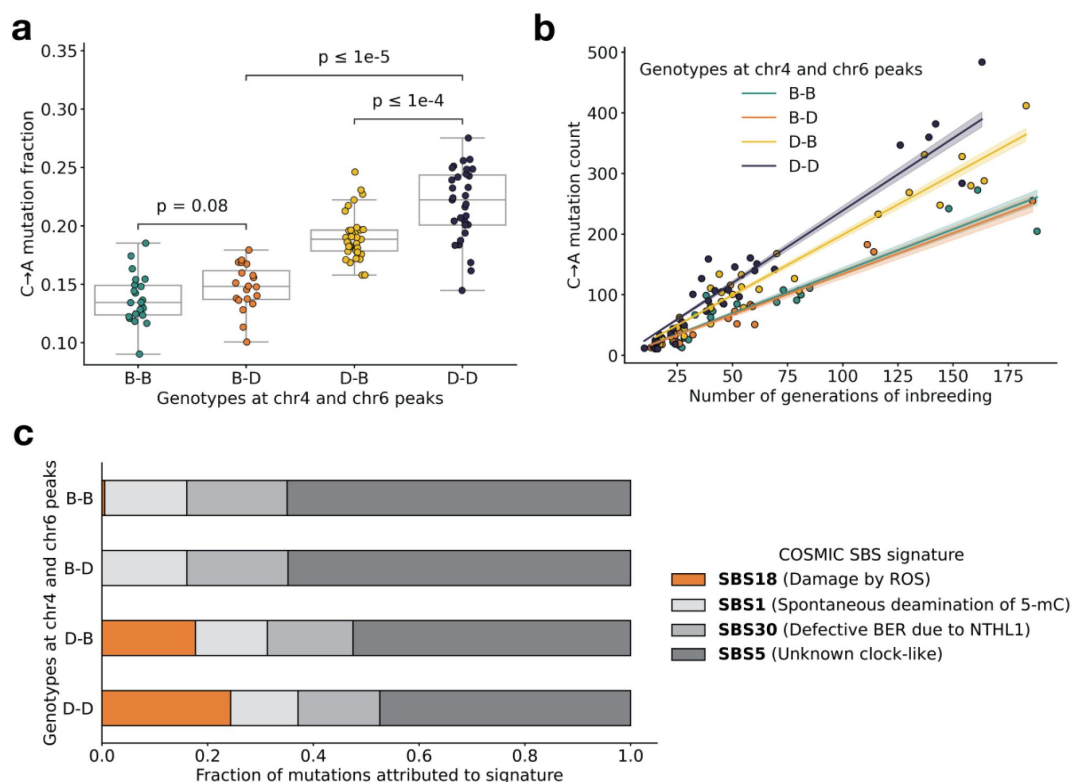


Figure 3

BXD mutation spectra are affected by alleles at both mutator loci.

a) C>A de novo germline mutation fractions in BXDs with either D or B genotypes at markers rs27509845 (chr4 peak) and rs46276051 (chr6 peak). Distributions of C>A mutation fractions were compared with two-sided Mann-Whitney U-tests; annotated p-values are uncorrected. B-B vs. B-D comparison: U-statistic = 149.0, $p = 7.58e-2$; B-D vs D-D comparison: U-statistic = 21.0, $p = 2.61e-8$; D-B vs D-D comparison: U-statistic = 232.5, $p = 6.99e-5$. **b)** The count of C>A de novo germline mutations in each BXD was plotted against the number of generations for which it was inbred. Lines represent predicted C>A counts in each haplotype group from a generalized linear model (Poisson family, identity link), and shading around each line represents the 95% confidence interval. **c)** Germline mutations in each BXD were assigned to COSMIC SBS mutation signatures using SigProfilerExtractor [29]. After grouping BXDs by their genotypes at rs27509845 and rs46276051, we calculated the fraction of mutations in each group that was attributed to each signature. The proposed etiologies of each mutation signature are: SBS1 (spontaneous deamination of methylated cytosine nucleotides at CpG contexts), SBS5 (unknown, clock-like signature), SBS18 (damage by reactive oxygen species, related to SBS36 and defective base-excision repair due to loss-of-function mutations in MUTYH), and SBS30 (defective base-excision repair due to NTHL1 mutations).

also performed QTL scans for the fractions of each 1-mer mutation type using the same mutation data, but none produced a genome-wide significant log-odds score at any locus (Figure 2-figure supplement 1; *Materials and Methods*).

We queried the region surrounding the top marker on chromosome 6 (+/- the 90% bootstrap confidence interval) and discovered 64 protein-coding genes, of which four were annotated with a Gene Ontology (GO) [24,25] term related to “DNA repair”: *Fancd2*, *Ogg1*, *Setmar*, and *Rad18*. None of the remaining genes were annotated with a cellular function that would obviously contribute to a germline mutator phenotype; however, many of these GO annotations are imperfect and/or incomplete. Although we focus our analysis on DNA repair genes, it remains possible that other genes within the confidence interval could underlie the C>A mutator phenotype we identified in the BXDs.

Of the annotated DNA repair genes within the confidence interval, two harbored nonsynonymous differences between the parental C57BL/6J and DBA/2J strains (Table 1). *Ogg1* encodes a key member of the base-excision repair response to oxidative DNA damage (a pathway that also includes *Mutyh*), and in mice *Setmar* encodes a SET domain-containing histone methyltransferase; both *Ogg1* and *Setmar* are expressed in mouse gonadal cells. Because the bootstrap can exhibit poor coverage in QTL mapping studies [26], we also scanned an interval +/- 5 Mbp from the peak AMSD marker on chromosome 6 for additional candidate genes. Although the choice of a 10 Mbp interval is somewhat arbitrary, the interval does contain a plausible candidate: *Mbd4*, a protein-coding gene involved in base excision repair that also harbors a non-synonymous difference between the BXD parental strains and is expressed in mouse gonads (Table 1).

We also considered the possibility that expression quantitative trait loci (eQTLs), rather than nonsynonymous mutations, could contribute to the C>A mutator phenotype associated with the locus on chromosome 6. Using GeneNetwork [27] we mapped eQTLs for the four aforementioned DNA repair genes (as well as *Mbd4*) in a number of tissues, though we did not have access to expression data from germline cells. Notably, *D* alleles near the cosine distance peak on chromosome 6 were significantly associated with decreased *Ogg1* expression in kidney, liver, hippocampus, and gastrointestinal tissues (Table supplement 1). Although these cis-eQTLs are challenging to interpret (given their tissue specificity and our lack of access to germline expression data), the presence of strong-effect cis-eQTLs for *Ogg1* suggests that the C>A mutator phenotype observed in the BXDs may be mediated by regulatory, rather than protein-altering, variants.

Finally, we queried a dataset of structural variants (SVs) identified via high-quality, long-read assembly of inbred laboratory mouse strains [28] and found 176 large insertions or deletions (>100 bp) within the 90% bootstrap confidence interval around the cosine distance peak on chromosome 6; none overlapped the exonic sequences of protein-coding genes. One protein-coding gene involved in DNA repair (*Rad18*) harbored an intronic deletion within the interval on chromosome 6 (chr6:112,629,618-112,636,619); however, additional experimental evidence will be needed to probe the functional impact of this structural variant.

Evidence of epistasis between germline mutator alleles

Next, we more precisely characterized the effects of the chromosome 4 and 6 mutator alleles on mutation spectra in the BXDs. To pinpoint the mutation type(s) that underlied the significant cosine distance peak on chromosome 6, we compared the aggregate counts of each 1-mer mutation type (plus CpG>TpG) on BXD haplotypes with *D* genotypes at rs27509845 and either *D* or *B* genotypes at rs46276051. We found that C>A mutations were significantly enriched on BXD haplotypes with *D* genotypes at the chromosome 6 mutator locus, relative to those with *B* genotypes (χ^2 statistic = 85.36, $p = 2.48\text{e-}20$). On average, C>A germline mutation fractions were significantly higher in BXDs with *D* alleles at both mutator loci than in BXDs with *D* alleles at either locus alone (Figure 3a and 3-figure supplement 1). Among BXDs with *B* alleles at the locus

Gene name	Ensembl transcript name	Nucleotide change	Amino acid change	Position in GRCm38/mm10 coordinates	PhyloP conservation score	SIFT prediction
Setmar	ENSMUST000000049246	C>T	p.Leu103Phe	chr6:108,075,853	0.422	0.0 (intolerant/deleterious)
Setmar	ENSMUST000000049246	T>G	p.Ser273Arg	chr6:108,076,365	-0.355	0.3 (tolerant/benign)
Ogg1	ENSMUST000000032406	A>G	p.Thr95Ala	chr6:113,328,510	-0.016	0.84 (tolerant/benign)
Mbd4	ENSMUST000000032469	C>T	p.Asp129Asn	chr6:115,849,644	2.28	0.02 (intolerant/deleterious)

Table 1

Nonsynonymous mutations in DNA repair genes near the chr6 peak

overlapping *Mutyh*, those with *D* alleles on chromosome 6 did not exhibit significantly elevated C>A mutation fractions (**Figure 3a**). After controlling for inbreeding duration, we observed that C>A *de novo* mutation counts were always highest in BXDs with *D* alleles at both mutator loci (**Figure 3b**). After 100 generations of inbreeding, BXDs with *D* alleles at both mutator loci were predicted to have 238.9 C>A mutations (95% CI: 231.4 - 246.4), about 20% more than the 199.0 mutations (95% CI: 193.3 - 204.7) predicted in those with *D* and *B* alleles at the chromosome 4 and chromosome 6 loci, respectively (**Figure 3b**).

We also used SigProfilerExtractor [29] to assign the germline mutations in each BXD to single-base substitution (SBS) mutation signatures from the COSMIC catalog [30]. Mutation signatures often reflect specific exogenous or endogenous sources of DNA damage, and the proportions of mutations attributable to particular SBS signatures can suggest a genetic or environmental etiology. The SBS1, SBS5, and SBS30 mutation signatures were active in nearly all BXDs, regardless of genotypes at the chromosome 4 and 6 mutator loci (**Figure 3c**). However, the SBS18 signature, which is dominated by C>A mutations and likely reflects unrepaired DNA damage from reactive oxygen species, was almost exclusively active in mice with *D* alleles at the chromosome 4 locus; the highest SBS18 activity was observed in mice with *D* alleles at both mutator loci (**Figure 3c**). SBS18 activity was lowest in mice with *D* alleles at the chromosome 6 mutator locus alone (**Figure 3c**), further demonstrating that *D* alleles at this locus are not sufficient to cause a mutator phenotype.

To more formally test for statistical epistasis, we fit a generalized (Poisson) linear model predicting counts of C>A mutations in each BXD as a function of genotypes at rs27509845 and rs46276051 (the markers with the largest adjusted cosine distance at the two mutator loci); the model also accounted for differences in inbreeding duration and sequencing coverage between the BXDs (*Materials and Methods*). A model that included an interaction term between genotypes at the two markers fit the data significantly better than a model including only additive effects ($p = 7.92 \times 10^{-7}$; *Materials and Methods*), indicating that the combined effects of *D* genotypes at both loci exceeded the sum of marginal effects of *D* genotypes at either locus alone.

To explore the effects of the two mutator loci in other inbred laboratory mice, we also compared the germline mutation spectra of Sanger Mouse Genomes Project (MGP) strains [31]. Dumont [18] previously identified germline mutations that were private to each of the 29 MGP strains; these private variants likely represent recent *de novo* mutations (**Figure 3**—figure supplement 2). Only two of the MGP strains possess *D* genotypes at both the chromosome 4 and chromosome 6 mutator loci: DBA/1J and DBA/2J. As before, we tested for epistasis in the MGP strains by fitting two linear models predicting C>A mutation counts as a function of genotypes at the two mutator loci. A model incorporating an interaction term did not fit the MGP data significantly better than a model with additive effects alone ($p = 0.806$), so we are unable to confirm the signal of epistasis; however, this may be due to the smaller number of MGP strains with *de novo* germline mutation data.

Some of the candidate mutator alleles are segregating in wild mice

To determine whether the candidate mutator alleles on chromosome 6 were segregating in natural populations, we queried previously published sequencing data generated from 67 wild-derived mice [32]. Although these data are not a comprehensive sampling of the genetic diversity in wild mice, they include three subspecies of *Mus musculus*, as well as the outgroup *Mus spretus*. We found that the *Ogg1* *D* allele was segregating at an allele frequency of 0.259 in *Mus musculus domesticus*, the species from which C57BL/6J and DBA/2J derive the majority of their genomes [33], and was fixed in *Mus musculus musculus*, *Mus musculus castaneus*, and the outgroup *Mus spretus* (**Figure 3-figure supplement 3**). The *Setmar* p.Ser273Arg *D* allele was also present at an allele frequency of 0.37 in *Mus musculus domesticus*, while *D* alleles at the *Setmar* p.Leu103Phe

variant were not observed in any wild *Mus musculus domesticus* animals. *D* alleles at the *Mbd4* p.Asp129Asn variant were also absent from all wild mouse populations (Figure 3-figure supplement 3).

Discussion

Epistasis between germline mutator alleles

We have identified a locus on chromosome 6 that amplifies a C>A germline mutator phenotype in the BXDs, a family of inbred mice derived from the laboratory strains DBA/2J and C57BL/6J. DBA/2J (*D*) alleles at this locus have no significant effect on C>A mutation rates in mice that also harbor “wild-type” C57BL/6J (*B*) alleles at a previously discovered mutator locus on chromosome 4 [2]. However, mice with *D* alleles at *both* loci have even higher mutation rates than those with *D* alleles at the chromosome 4 mutator locus alone (Figure 3). Epistatic interactions between mutator alleles have been previously documented in yeast [34] and in human cell lines [35], but never to our knowledge in a whole-animal context.

Importantly, we discovered epistasis between germline mutator alleles in an unnatural population of model organisms that have been inbred by brother-sister mating in a highly controlled laboratory environment [3]. This breeding setup has likely attenuated the effects of natural selection on all but the most deleterious alleles [19], and may have facilitated the fixation of large-effect mutator alleles that would be less common in wild mice. Without fine-mapping the chromosome 6 mutator allele, however, we are unable to trace its origin to either a captive breeding colony of laboratory mice or a wild, outbreeding *Mus musculus* population. If the mutator allele on chromosome 6 has even a weak deleterious fitness effect, there might be a greater likelihood that it persisted in captivity. Indeed, if purifying selection is required to keep mutation rates low, mutational pressure might cause mutation rates to rise in just a few generations of relaxed selection. This dynamic may explain the recent discovery of a large-effect mutator allele in a rhesus macaque research colony [13], as well as the observation that domesticated animals tend to have higher mutation rates than those in the wild [36]. Although we have not conclusively fine-mapped the chromosome 6 mutator locus to a causal variant, we argue that nonsynonymous or regulatory variants in the DNA glycosylase *Ogg1* are the best candidates.

Protein-coding genes that may underlie the chromosome 6 mutator locus

Five protein-coding genes involved in DNA repair overlap the C>A mutator locus on chromosome 6: *Ogg1*, a glycosylase that excises the oxidative DNA lesion 8-oxoguanine (8-oxoG) [4], *Setmar*, a histone methyltransferase involved in non-homologous end joining (NHEJ) of double-stranded breaks (DSBs) [37,38], *Fancd2*, and *Rad18*. One other DNA repair gene, *Mbd4*, lies just outside of the 90% bootstrap confidence interval on chromosome 6 (but within a 10 Mbp interval around the peak AMSD marker). We are unable to conclusively determine that one or more of these genes harbors a causal variant underlying the observed C>A mutator phenotype, but we believe that *Ogg1* is the most plausible candidate. *Ogg1* is a member of the same base-excision repair pathway as *Mutyh* (the gene that likely underlies the chromosome 4 mutator locus), contains a nonsynonymous fixed difference between the C57BL/6J and DBA/2J parental strains, and appears to be regulated by cis-eQTLs across a number of tissues within the BXD cohort.

The C57BL/6J and DBA/2J *Setmar* coding sequences differ by two missense variants (Table 1), one of which is predicted to be deleterious by *in silico* tools. The primate *SETMAR* ortholog is involved in NHEJ of double-strand breaks, but its role in DNA repair appears to depend on the function of both a SET methyltransferase domain and a *Mariner*-family transposase domain

[38 [↗](#), 39 [↗](#), 40 [↗](#)]. Since the murine *Setmar* ortholog lacks the latter element, and because primate *SETMAR* is involved in a DNA repair process that is not expected to affect the rate of C>A mutations, we believe it is unlikely to underlie the epistatic interaction between the chromosome 4 and 6 mutator loci in the BXDs (*Appendix 1*). Moreover, we did not observe any significant cis-eQTLs for *Setmar* across a variety of tissues in the BXD cohort (Table supplement 1). None of the remaining DNA repair genes (*Fancd2* or *Rad18*) contains a nonsynonymous fixed difference between the C57BL/6J and DBA/2J parental strains, and none appear to be regulated by cis-eQTLs that would feasibly lead to a germline C>A mutator phenotype (Table supplement 1); the only significant cis-eQTLs we observed for genes other than *Ogg1* were those at which *D* alleles actually led to *increased* expression.

An *Ogg1* mutator allele might impair the excision of 8-oxoguanine lesions

Ogg1 is a member of the same base-excision repair (BER) pathway as *Mutyh*, the protein-coding gene we previously implicated as harboring mutator alleles at the locus on chromosome 4 [4 [↗](#)]. Each of these genes has a distinct role in the BER response to oxidative DNA damage, and thereby the prevention of C>A mutations [41 [↗](#), 42 [↗](#)]. Following damage by reactive oxygen species, *Ogg1* is able to recognize and remove 8-oxoguanine lesions that are base-paired with cytosine nucleotides; once 8-oxoG is excised, other members of the BER pathway are mobilized to restore a proper G:C base pair at the site. If an 8-oxoG lesion is not removed before the cell enters S-phase, adenine can be mis-incorporated opposite 8-oxoG during DNA replication [41 [↗](#)]. If this occurs, *Mutyh* can excise the mispaired adenine, leaving a one-nucleotide gap that is processed and filled with a cytosine by other BER proteins. The resulting C:8-oxoG base pair can then be “returned” to *Ogg1* for excision and repair. Defects in the BER response to oxidative damage lead to significantly elevated rates of C>A mutation. For example, triple-knockout (KO) mice lacking *Ogg1*, *Mutyh*, and *Mth1* (which encodes an enzyme that prevents 8-oxo-dGTP from being incorporated during DNA synthesis [43 [↗](#)]) accumulate a 100-fold excess of 8-oxoG in their gonadal cells [42 [↗](#)]. Almost 99% of *de novo* germline mutations in the *Ogg1/Mutyh/Mth1* triple KO mice are C>A transversions, demonstrating the clear role of 8-oxoG repair in preventing C>A mutation. Additionally, missense mutations and loss-of-heterozygosity in *Ogg1* have been associated with increased risk of human cancer [44 [↗](#), 45 [↗](#)], and copy-number losses of either *Ogg1* or *Mutyh* are linked to elevated rates of spontaneous C>A mutation in human neuroblastoma [46 [↗](#)].

Nonsynonymous mutations may underlie the chromosome 6 mutator phenotype

The p.Thr95Ala *Ogg1* missense variant is not predicted to be deleterious by the *in silico* tool SIFT [47 [↗](#)], and occurs at a nucleotide that is not particularly well-conserved across mammalian species (Table 1 [↗](#)). We also observe that the *D* allele at p.Thr95Ala is segregating at an allele frequency of approximately 26% among wild-derived *Mus musculus domesticus* animals, and is fixed in other wild populations of *Mus musculus musculus*, *Mus musculus castaneus*, and *Mus spretus*. Although we would expect *a priori* that *Ogg1* deficiency should lead to increased 8-oxoG accumulation and elevated C>A mutation rates, these lines of evidence suggest that p.Thr95Ala is not highly deleterious on its own, and might only exert a detectable effect on the BER gene network when *Mutyh* function is also impaired (or *Mutyh* expression is decreased). It is also possible that *D* alleles at *Ogg1* lead to a very subtle increase in C>A mutation rates, and we are simply underpowered to detect such a small mutation rate effect in the BXDs.

No indication of causal structural variation or mobile element insertions near the chromosome 6 mutator locus

Although we argue above that *Ogg1* is likely the the best candidate gene to explain the new BXD C>A mutator phenotype, we cannot conclusively determine that the p.Thr95Ala missense mutation is a causal allele. We previously hypothesized that *Mutyh* missense mutations on *D* haplotypes were responsible for the large-effect C>A mutator phenotype we observed in the BXDs [2]. However, subsequent long-read assemblies of several inbred laboratory mouse strains revealed that this mutator phenotype might be caused by a ~5 kbp mobile element insertion (MEI) within the first intron of *Mutyh* [28], which is associated with significantly reduced expression of *Mutyh* in embryonic stem cells. We queried the new high-quality assemblies for evidence of mobile elements or other large structural variants (SVs) in the region surrounding the mutator locus on chromosome 6, but found no similarly compelling evidence that either SVs or MEIs might underlie the mutator phenotype described in this study.

Expression quantitative trait loci (eQTLs) might mediate germline mutator phenotypes in the BXDs

We observed strong-effect cis-eQTLs for *Ogg1* expression across a number of tissues in the BXDs (Table supplement 1). In each of these tissue types, *D* genotypes were associated with decreased expression of *Ogg1*. As mentioned above, new evidence from long-read genome assemblies has demonstrated that an intronic mobile element insertion in *Mutyh* may be responsible for decreased *Mutyh* expression, and therefore higher C>A mutation rates, in BXDs with *D* haplotypes at the chromosome 4 mutator locus [28]. Taken together, these results raise the exciting possibility that the mutator loci on both chromosome 4 and chromosome 6 lead to increased C>A mutation rates by lowering the expression of DNA repair genes in the same base-excision repair network. Moreover, the effect of the mutator locus on chromosome 6 may be conditional on the expression of the mutator near *Mutyh*, suggesting that complex genetic interactions can underlie germline mutator phenotypes in mammalian genomes.

Mbd4 may buffer the effects of *Mutyh* mutator alleles by triggering apoptosis

As mentioned in the *Results*, *Mbd4* lies just outside of the 90% bootstrap confidence interval on chromosome 6, but within a 10 Mbp window surrounding the AMSD peak. Due to the uncertainties associated with bootstrap confidence intervals in QTL mapping [26], we have included a discussion of the evidence supporting *Mbd4* as a causal gene below.

Unlike the *Ogg1* p.Thr95Ala mutation, the p.Asp129Asn variant in *Mbd4* resides within an annotated protein domain (the *Mbd4* methyl-CpG binding domain), occurs at a nucleotide and amino acid residue that are both well-conserved, and is predicted to be deleterious by SIFT [47] (Table 1). A missense mutation that affects the homologous amino acid in humans (p.Asp142Gly in GRCh38/hg38) is also present on a single haplotype in the Genome Aggregation Database (gnomAD) [48] and is predicted by SIFT and Polyphen [49] to be “deleterious” and “probably_damaging” in human genomes, respectively.

One puzzling observation is that loss-of-function mutations in *Mbd4* are not typically associated with C>A mutator phenotypes. Instead, *Mbd4* deficiency is usually implicated in C>T mutagenesis at CpG sites, and we did not detect an excess of C>T mutations in BXDs with *D* alleles at the chromosome 6 mutator locus (Figure 3-figure supplement 1). However, loss of function mutations in *Mbd4* have also been shown to exacerbate the effects of exogenous DNA damage agents. For example, mouse embryonic fibroblasts that harbor homozygous *Mbd4* knockouts fail to undergo apoptosis following treatment with a number of chemotherapeutics and mutagenic compounds [50]. Most of these exogenous mutagens cause DNA damage that is normally repaired by

mismatch repair (MMR) machinery, but murine intestinal cells with biallelic *Mbd4* LOF mutations also showed a reduced apoptotic response to gamma irradiation, which is repaired independently of the MMR gene *Mlh1* [51]. Homozygous loss of *Mbd4* function also leads to accelerated intestinal tumor formation in mice that harbor an *Apc* allele that predisposes them to intestinal neoplasia [52], and mice with biallelic truncations of the *Mbd4* coding sequence exhibit modestly increased mutation rates in colon cancer cell lines, including increased C>A mutation rates in certain lines [53].

Together, these lines of evidence suggest that *Mbd4* can modulate sensitivity to many types of exogenous mutagens, potentially through its role in determining whether cells harboring DNA damage should undergo apoptosis [50, 51]. We speculate that in mice with deficient 8-oxoguanine repair — caused by a mutator allele in *Mutyh*, for example — reactive oxygen species (ROS) could cause accumulation of DNA damage in the germline. If those germ cells harbor fully functional copies of *Mbd4*, they might be able to trigger apoptosis and partially mitigate the effects of a *Mutyh* mutator allele. However, mice with reduced activity of both *Mbd4* and *Mutyh* may have a reduced ability to initiate cell death in response to DNA damage; as a result, their germ cells may accumulate even higher levels of ROS-mediated damage, leading to substantially elevated germline C>A mutation rates.

We anticipate that future experimental work will be able to more conclusively establish a mechanistic explanation for the epistatic interaction between mutator loci described in this paper.

Strengths and limitations of the aggregate mutation spectrum distance approach

Our aggregate mutation spectrum distance (AMSD) approach was able to identify a mutator allele that escaped notice using quantitative trait locus (QTL) mapping. To more systematically compare the power of AMSD and QTL mapping, we performed simulations under a variety of possible parameter regimes. Overall, we found that AMSD and QTL mapping have similar power to detect mutator alleles on haplotypes that each harbor tens or hundreds of *de novo* germline mutations (Figure 1-figure supplement 2). Nonetheless, only AMSD was able to discover the mutator locus on chromosome 6 in the BXDs, demonstrating that it outperforms QTL mapping in certain experimental systems. For example, simulations demonstrate that AMSD enjoys greater power than QTL mapping when haplotypes carry highly variable numbers of mutations that can be leveraged for mutator mapping (Figure 1-figure supplement 3). Because the BXDs were generated in six breeding epochs over a period of nearly 40 years, the oldest lines have accumulated orders of magnitude more mutations than the youngest lines; these younger BXDs have much noisier mutation spectra as a result. While approaches for QTL mapping typically weight the phenotypic measurements of each sample equally, AMSD compares the *aggregate* mutation spectra of haplotypes at every locus, a property that likely increased its power to detect mutators in the BXD dataset.

Another benefit of the AMSD approach is that it obviates the need to perform separate association tests for every possible *k*-mer mutation type, and therefore the need to adjust significance thresholds for multiple tests. Since AMSD compares the complete mutation spectrum between haplotypes that carry either allele at a site, it would also be well-powered to detect a mutator allele that exerted a coordinated effect on multiple *k*-mer mutation types (e.g., increased the rates of both C>T and C>A mutations).

However, the AMSD method suffers a handful of drawbacks when compared to QTL mapping. Popular QTL mapping methods (such as R/qt12 [54]) use linear models to test associations between genotypes and phenotypes, enabling the inclusion of additive and interactive covariates, as well as kinship matrices, in QTL scans. Although we have developed methods to account for inter-sample relatedness in the AMSD approach (*Materials and Methods*), they are not as flexible

as similar methods in QTL mapping software. Additionally, the AMSD method assumes that mutator alleles affect a subset of k -mer mutation types; if a mutator allele increased the rates of all mutation types equally on haplotypes that carried it, AMSD would be unable to detect it.

Discovering mutator alleles in other experimental systems

Our discovery of a second BXD mutator allele underscores the power of recombinant inbred lines (RILs) as a resource for dissecting the genetic architecture of germline mutation rates. Large populations of RILs exist for many model organisms, and we anticipate that as whole-genome sequencing becomes cheaper and cheaper, the AMSD method could be useful for future mutator allele discovery outside of the BXDs. At the same time, RILs are a finite resource that require enormous investments of time and labor to construct. If germline mutator alleles are only detectable in these highly unusual experimental populations, we are unlikely to discover more than a small fraction of the mutator alleles that may exist in nature.

In natural, outbreeding populations, selection on germline mutator alleles will likely prevent large-effect mutators from reaching high allele frequencies, but a subset may be detectable by sequencing a sufficient number of human trios [56]. Since germline mutators often seem to exert their effects on a small number of mutation types, mutation spectrum analyses may have greater power to detect the genes that underlie heritable mutation rate variation, even if each gene has only a modest effect on the overall mutation rate per generation.

Thousands of human pedigrees have been sequenced in an effort to precisely estimate the rate of human *de novo* germline mutation [5,6,55], and as family sequencing has become a more common step in the diagnosis of many congenital disorders, these datasets are growing on a daily basis. AMSD could potentially be applied to these large cohorts of two- or three-generation families. We propose that by pooling sparse mutation counts across many individuals who share the same candidate mutator allele, even a subtle mutator signal might rise above the noise of *de novo* germline mutation rate estimates. We note that the aggregate mutation spectrum distance approach will require modification before it can be successfully applied to cohorts of outbred, sexually-reproducing individuals. AMSD assumes that individuals harbor one of two possible genotypes at each marker, and does not yet account for heterozygous genotypes. As a result, our method is currently applicable only to resources like the BXD RILs, in which individuals have been inbred for sufficiently long that effectively all genotypes are homozygous. However, as human trio datasets become larger and larger in the coming years, rare mutator alleles may eventually be represented in as many individuals as the high-frequency variants present in the BXDs, thereby enabling more systematic mapping of elusive mutator phenotypes.

Materials and Methods

Identifying *de novo* germline mutations in the BXDs

The BXD resource currently comprises a total of 152 recombinant inbred lines (RILs). BXDs were derived from either F2 or advanced intercrosses, and subsequently inbred by brother-sister mating for up to 180 generations [3]. BXDs were generated in distinct breeding “epochs,” which were each initiated with a distinct cross of C57BL/6J and DBA/2J parents; epochs 1, 2, 4, and 6 were derived from F2 crosses, while epochs 3 and 5 were derived from advanced intercrosses [3]. Previously, we analyzed whole-genome sequencing data from the BXDs and identified candidate *de novo* germline mutations in each line [2]. A detailed description of the methods used for DNA extraction, sequencing, alignment, and variant processing, as well as the characteristics of the *de novo* mutations, are available in a previous manuscript [2].

Briefly, we identified private single-nucleotide mutations in each BXD that were absent from all other BXDs, as well as from the C57BL/6J and DBA/2J parents. We required each private variant to be meet the following criteria:

- genotyped as either homozygous or heterozygous for the alternate allele, with at least 90% of sequencing reads supporting the alternate allele
- supported by at least 10 sequencing reads
- Phred-scaled genotype quality of at least 20
- must not overlap regions of the genome annotated as segmental duplications or simple repeats in GRCm38/mm10
- must occur on a parental haplotype that was inherited by at least one other BXD at the same locus; these other BXDs must be homozygous for the reference allele at the variant site

A new approach to discover germline mutator alleles

Calculating aggregate mutation spectrum distance

We developed a new approach to discover loci that affect the germline *de novo* mutation spectrum in biparental RILs ([Figure 1](#)).

We assume that a collection of haplotypes has been genotyped at informative markers, and that *de novo* germline mutations have been identified on each haplotype.

At each informative marker, we divide haplotypes into two groups based on the parental allele that they inherited. We then compute a *k*-mer mutation spectrum using the aggregate mutation counts in each haplotype group. The *k*-mer mutation spectrum contains the frequency of every possible *k*-mer mutation type in a collection of mutations, and can be represented as a vector of size $6 \times 4^{k-1}$ after collapsing by strand complement. For example, the 1-mer mutation spectrum is a 6-element vector that contains the frequencies of C>T, C>G, C>A, A>G, A>T, and A>C mutations. Since C>T transitions at CpG nucleotides are often caused by a distinct mechanism (spontaneous deamination of methylated cytosine), we expand the 1-mer mutation spectrum to include a separate category for CpG>TpG mutations [57].

At each marker, we then compute the cosine distance between the two aggregate spectra. The cosine distance between two vectors **A** and **B** is defined as

$$D^c = 1 - \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|}$$

where $\|\mathbf{A}\|$ and $\|\mathbf{B}\|$ are the L^2 (or Euclidean) norms of **A** and **B**, respectively. The cosine distance metric has a number of favorable properties for comparing mutation spectra. Since it adjusts for the magnitude of the two input vectors, cosine distance can be used to compare two spectra with unequal total mutation counts (even if those total counts are relatively small).

Additionally, by calculating the cosine distance between mutation spectra, we avoid the need to perform separate comparisons of mutation counts at each individual *k*-mer mutation type.

Inspired by methods from QTL mapping [54,58], we use permutation tests to establish genome-wide cosine distance thresholds. In each of *N* permutation trials, we randomly shuffle the per-haplotype mutation data such that haplotype labels no longer correspond to the correct mutation counts. Using the shuffled mutation data, we perform a genome-wide scan as described

above, and record the maximum cosine distance observed at any locus. After N permutations (usually 10,000), we compute the $1 - p$ percentile of the distribution of maximum statistics, and use that percentile value as a genome-wide significance threshold (for example, at $p = 0.05$).

Estimating confidence intervals around AMSD peaks

If we identified an adjusted cosine distance peak on a particular chromosome, we used a bootstrap resampling approach [59] to estimate confidence intervals. In each of $N = 10,000$ trials, we resampled the mutation spectrum data and corresponding marker genotypes (on the chromosome of interest) with replacement. Using those resampled spectra and genotypes, we performed an aggregate mutation spectrum distance scan on the chromosome of interest and recorded the position of the marker with the largest adjusted cosine distance value. We then defined a 90% confidence interval by finding two marker locations between which 90% of all N bootstrap samples produced a peak cosine distance value. In other words, we estimated the bounds of the 90% confidence interval by finding the markers that defined the 5th and 95th percentiles of the distribution of maximum adjusted cosine distance values across N bootstrap trials. We note, however, that the bootstrap can exhibit poor performance in QTL mapping studies [26]; namely, bootstrap confidence intervals tend to be larger than those estimated using either a “LOD drop” method or a Bayes credible interval, and can exhibit poorer-than-expected coverage (a measure of whether the confidence interval contains the true QTL location).

Accounting for relatedness between strains

We expect each BXD to derive approximately 50% of its genome from C57BL/6J and 50% from DBA/2J. As a result, every pair of BXDs will likely have identical genotypes at a fraction of markers. Pairs of more genetically similar BXDs may also have more similar mutation spectra, potentially due to shared polygenic effects on the mutation process. Therefore, at a given marker, if the BXDs that inherited D alleles are more genetically dissimilar from those that inherited B alleles (considering all loci throughout the genome in our measurement of genetic similarity), we might expect the aggregate mutation spectra in the two groups to also be more dissimilar.

We implemented a simple approach to account for these potential issues of relatedness. At each marker g_i , we divide BXD haplotypes into two groups based on the parental allele they inherited. As before, we first compute the aggregate mutation spectrum in each group of haplotypes and calculate the cosine distance between the two aggregate spectra (D_i^C). Then, within each group of haplotypes, we calculate the allele frequency of the D allele at every marker along the genome to obtain a vector of length n , where n is the number of genotyped markers. To quantify the genetic similarity between the two groups of haplotypes, we calculate the Pearson correlation coefficient r_i between the two vectors of marker-wide D allele frequencies.

Put another way, at every marker g_i along the genome, we divide BXD haplotypes into two groups and compute two metrics: D_i^C (the cosine distance between the two groups' aggregate spectra) and r_i (the correlation between genome-wide D allele frequencies in the two groups). To control for the potential effects of genetic similarity on cosine distances, we regress $(D_1^C, D_2^C, \dots, D_n^C)$ on $(r_1, r_2, \dots, r_n)$ for all n markers using an ordinary least-squares model. We then use the residuals from the fitted model as the “adjusted” cosine distance values for each marker. If genome-wide genetic similarity between haplotypes perfectly predicts cosine distances at each marker, these residuals will all be 0 (or very close to 0). If genome-wide genetic similarity has no predictive power, the residuals will simply represent the difference between the observed cosine distance at a single marker and the marker-wide mean of cosine distances.

Accounting for BXD population structure due to breeding epochs

The current BXD family was generated in six breeding “epochs.” As discussed previously, each epoch was initiated with a distinct cross of C57BL/6J and DBA/2J parents; BXDs in four of the epochs were generated following F2 crosses of C57BL/6J and DBA/2J, and BXDs in the other two were generated following advanced intercrosses. Due to this breeding approach the BXD epochs differ from each other in a few important ways. For example, BXDs derived in epochs 3 and 5 (i.e., from advanced intercross) harbor larger numbers of fixed recombination breakpoints than those from epochs 1, 2, 4, and 6 [3]. Although the C57BL/6J and DBA/2J parents used to initialize each epoch were completely inbred, they each possessed a small number unique *de novo* germline mutations that were subsequently inherited by many of their offspring. A number of these “epoch-specific” variants have also been linked to phenotypic variation observed between BXDs from different epochs [3,60,61,62].

To account for potential population structure, as well as these epoch-specific effects, we introduced the ability to perform stratified permutation tests in the aggregate mutation spectrum distance approach. Normally, in each of N permutations we shuffle the per-haplotype mutation spectrum data such that haplotype labels no longer correspond to the correct mutation spectra (i.e., shuffle mutation spectra *across* epochs). In the stratified approach, we instead shuffle per-haplotype mutation data *within* epochs, preserving epoch structure while still enabling mutation spectra permutations. We used this epoch-aware approach for all permutation tests presented in this manuscript.

Implementation and source code

The aggregate mutation spectrum distance method was implemented in Python, and relies heavily on the following Python libraries: numpy, pandas, matplotlib, scikit-learn, pandera, seaborn, and numba [63,64,65,66,67,68,69].

The code underlying AMSD, as well as documentation of the method, is available on GitHub (<https://github.com/quinlan-lab/proj-mutator-mapping>). We have also deposited a reproducible Snakemake [70] workflow for running reproducing all analyses and figures presented in the manuscript.

Simulations to assess the power of the aggregate mutation spectrum distance approach

We performed a series of simple simulations to estimate our power to detect alleles that affect the germline mutation spectrum using the aggregate mutation spectrum distance method.

Simulating genotypes

First, we simulate genotypes on a population of haplotypes at a collection of sites. We define a matrix G of size (s, h) , where s is the number of sites and h is the number of haplotypes. We assume that every site is biallelic, and that the minor allele frequency at every site is 0.5. For every entry G_{ij} , we take a single draw from a uniform distribution in the interval $[0,0.1,0]$. If the value of that draw is less than 0.5, we assign the value of G_{ij} to be 1. Otherwise, we assign the value of G_{ij} to be 0.

Defining expected mutation type probabilities

Next, we define a vector of 1-mer mutation probabilities:

$$P = (0.29, 0.17, 0.12, 0.075, 0.1, 0.075, 0.17)$$

These probabilities sum to 1 and roughly correspond to the expected frequencies of C>T, CpG>TpG, C>A, C>G, A>T, A>C, and A>G *de novo* germline mutations in mice, respectively [9]. If we are simulating the 3-mer mutation spectrum, we modify the vector of mutation probabilities P to be length 96, and assign every 3-mer mutation type a value of $\frac{P_c}{16}$, where P_c is the probability of the “central” mutation type associated with the 3-mer mutation type. In other words, each of the 16 possible NCN>NTN 3-mer mutation types would be assigned a mutation probability of $\frac{P_c}{16} = \frac{0.46}{16} = 0.02875$. We then generate a vector of lambda values by scaling the mutation probabilities by the number of mutations we wish to simulate (m):

$$\lambda = Pm$$

We also create a second vector of lambda values (λ'), in which we multiply the λ value of a single mutation type by the mutator effect size e .

Rather than simulate the same mean number of mutations (m) on every haplotype, we also performed a series of simulations in which the mean number of mutations on each haplotype was allowed to vary. The BXD RILs were inbred for variable numbers of generations, and each BXD therefore accumulated a variable number of *de novo* germline mutations [2]. To more closely approximate the BXD haplotypes, we performed simulations in which the number of mutations (m) on each haplotype was drawn from a uniform distribution from m to $20m$. In other words, we created a vector of mutation counts M containing h evenly-spaced integers from m to $20m$, where h is the number of simulated haplotypes. Thus, if we simulated between 100 and 2,000 mutations on 50 haplotypes, the i th entry of M would be $100 + \frac{(2,000-100)}{50} i$. Each haplotype’s mean number of mutations was then assigned by looking up the haplotype’s index i in M .

In our simulations, we assume that genotypes at a single site (the “mutator locus”) are associated with variation in the mutation spectrum. That is, at a single site s_i , all of the haplotypes with 1 alleles should have elevated rates of a particular mutation type and draw their mutation counts from λ' , while all of the haplotypes with 0 alleles should have “wild-type” rates of that mutation type and draw their mutation counts from λ . We therefore pick a random site s_i to be the “mutator locus,” and identify the indices of haplotypes in G that were assigned 1 alleles at s_i . We call these indices h_{mut} .

Simulating mutation spectra

To simulate the mutation spectrum on our toy population of haplotypes, we define a matrix C of size (h, n) , where $n = 6 \times 4^{k-1}$ (or if $k = 1$ and we include CpG>TpG mutations, $6 \times 4^{k-1} + 1$).

Then, we populate the matrix C separately for *mutator* and *wild-type* haplotypes. For every row i in the matrix (i.e., for every haplotype), we first ask if i is in h_{mut} (that is, if the haplotype at index i was assigned a 1 allele at the “mutator locus”). If so, we set the values of C_i to be the results of a single Poisson draw from λ' . If row i is not in h_{mut} , we set the values of C_i to be the results of a single Poisson draw from λ .

Assessing power to detect a simulated mutator allele using AMSD

For each combination of parameters (number of simulated haplotypes, number of simulated markers, mutator effect size, etc.), we run 100 independent trials. In each trial, we simulate the genotype matrix G and the mutation counts C . We calculate a “focal” cosine distance as the cosine distance between the aggregate mutation spectra of haplotypes with either genotype at s_i (the site at which we artificially simulated an association between genotypes and mutation spectrum variation). We then perform an aggregate mutation spectrum distance scan using $N = 1,000$

permutations. If fewer than 5% of the N permutations produced a cosine distance greater than or equal to the focal distance, we say that the approach successfully identified the mutator allele in that trial.

Assessing power to detect a simulated mutator allele using quantitative trait locus (QTL) mapping

Using simulated data, we also assessed the power of traditional quantitative trait locus (QTL) mapping to detect a locus associated with mutation spectrum variation. As described above, we simulated both genotype and mutation spectra for a population of haplotypes under various conditions (number of mutations per haplotype, mutator effect size, etc.). Using those simulated data, we used R/qtl2 [54] to perform a genome scan for significant QTL as follows; we assume that the simulated genotype markers are evenly spaced (in physical Mbp coordinates) on a single chromosome. First, we calculate the fraction of each haplotype's *de novo* mutations that belong to each of the $6 \times 4^{k-1}$ possible k -mer mutation types. We then convert the simulated genotypes at each marker to genotype probabilities using the `calc_genoprob` function in R/qtl2, with `map_function = "c-f"` and `error_prob = 0`. For every k -mer mutation type, we use genotype probabilities and per-haplotype mutation fractions to perform a scan for QTL with the `scan1` function; to make the results more comparable to those from the AMSD method, we do not include any covariates or kinship matrices in these QTL scans. We then use the `scan1perm` function to perform 1,000 permutations of the per-haplotype mutation fractions and calculate log-odds (LOD) thresholds for significance. We consider the QTL scan to be "successful" if it produces a LOD score above the significance threshold (defined using $\alpha = \frac{0.05}{7}$) for the marker at which we simulated an association with mutation spectrum variation.

In our simulations, we augment the mutation rate of a single k -mer mutation type on haplotypes carrying the simulated mutator allele. However, in an experimental setting, we would not expect to have *a priori* knowledge of the mutation type affected by the mutator. Thus, by using an alpha threshold of 0.05 in our simulations, we would likely over-estimate the power of QTL mapping for detecting the mutator. Since we would need to perform 7 separate QTL scans (one for each 1-mer mutation type plus CpG>TpG) in an experimental setting, we calculate QTL LOD thresholds at a Bonferroni-corrected alpha value of $\alpha = \frac{0.05}{7}$.

Applying the aggregate mutation spectrum distance method to the BXDs

We downloaded previously-generated BXD *de novo* germline mutation data from the GitHub repository associated with our previous manuscript, which was also archived at Zenodo [2,71,72], and downloaded a CSV file of BXD genotypes at ~7,300 informative markers from GeneNetwork [27,73]. We also downloaded relevant metadata about each BXD from the manuscript describing the updated BXD resource [3]. These files are included in the GitHub repository associated with this manuscript.

As in our previous manuscript [2], we included mutation data from a subset of the 152 BXDs in our aggregate mutation spectrum distance scans. Specifically, we removed BXDs that were backcrossed to a C57BL/6J or DBA/2J parent at any point during the inbreeding process (usually, in order to rescue that BXD from inbreeding depression [3]). We also removed BXD68 from our genome-wide scans, since we previously discovered a hyper-mutator phenotype in that line; the C>A germline mutation rate in BXD68 is over 5 times the population mean, likely due to a private deleterious nonsynonymous mutation in *Mutyh* [2]. In our previous manuscript, we removed any BXDs that had been inbred for fewer than 20 generations, as it takes approximately 20 generations of strict brother-sister mating for an RIL genome to become >98% homozygous [74]. As a result, any potential mutator allele would almost certainly be either fixed or lost after 20

generations; if fixed, the allele would remain linked to any excess mutations it causes for the duration of subsequent inbreeding. In other words, the *de novo* mutations present in the genome of a “young” BXD (i.e., a BXD that was inbred for fewer than 20 generations) would not reflect a mutator allele’s activity as strongly as the mutations present in the genome of a much older BXD. This presented a challenge when we used quantitative trait locus mapping to discover mutator alleles in our previous manuscript, since the phenotypes (i.e., C>A mutation rates) of young and old BXDs were weighted equally; thus, we simply removed the younger BXDs from our analysis to avoid using their especially noisy mutation spectra. Since AMSD computes an *aggregate* mutation spectrum using all BXDs that inherited a particular allele at a locus, and can overcome the sparsity and noise of individual mutation spectra, we chose to include these younger BXDs in our genome-wide scans in this study.

In total, we included 117 BXDs in our genome-wide scans.

Identifying candidate single-nucleotide mutator alleles overlapping the chromosome 6 peak

We investigated the region implicated by our aggregate mutation spectrum distance approach on chromosome 6 by subsetting the joint-genotyped BXD VCF file (European Nucleotide Archive accession PRJEB45429 [75]) using bcftools [76]. We defined the candidate interval surrounding the cosine distance peak on chromosome 6 as the 90% bootstrap confidence interval (extending from approximately 95 Mbp to 114 Mbp). To predict the functional impacts of both single-nucleotide variants and indels on splicing, protein structure, etc., we annotated variants in the BXD VCF using the following snpEff [77] command:

```
java -Xmx16g -jar /path/to/snpEff/jarfile GRCm38.75 /path/to/bxd/vcf >
/path/to/uncompressed/output/vcf
```

and used cyvcf2 [78] to iterate over the annotated VCF file in order to identify nonsynonymous fixed differences between the parental C57BL/6J and DBA/2J strains.

Identifying candidate structural variant alleles overlapping the chromosome 6 peak

We downloaded summary VCFs containing insertion, deletion and inversion structural variants (identified via high-quality, long-read assembly of inbred laboratory mouse strains [28]) from the Zenodo link associated with the Ferraj et al. manuscript: <https://doi.org/10.5281/zenodo.7644286>.

We then downloaded a TSV file containing RefSeq gene predictions in GRCm39/mm39 from the UCSC Table Browser [79], and used the bx-python library [80] to intersect the interval spanned by each structural variant with the intervals spanned by the txStart and txEnd of every RefSeq entry. We queried all structural variants within the 90% bootstrap confidence interval on chromosome 6.

Extracting mutation signatures

We used SigProfilerExtractor (v.1.1.21) [30] to extract mutation signatures from the BXD mutation data. After converting the BXD mutation data to the “matrix” input format expected by SigProfilerExtractor, we ran the sigProfilerExtractor method as follows:

```
# install the mm10 mouse reference data
genInstall.install('mm10')

# run mutation signature extraction
sig.sigProfilerExtractor(
  'matrix',
  /path/to/output/directory,
  /path/to/input/mutations,
  maximum_signatures=10,
  nmf_replicates=100,
  opportunity_genome="mm10",
)
```

Comparing mutation spectra between Mouse Genomes Project strains

We downloaded mutation data from a previously published analysis [18] (Supplementary File 1, Excel Table S3) that identified strain-private mutations in 29 strains that were originally whole-genome sequenced as part of the Sanger Mouse Genomes (MGP) project [31]. When comparing counts of each mutation type between MGP strains that harbored either *D* or *B* alleles at the chromosome 4 or chromosome 6 mutator loci, we adjusted mutation counts by the number of callable A, T, C, or G nucleotides in each strain as described previously [2].

Querying GeneNetwork for eQTLs at the mutator locus

We used the online GeneNetwork resource [27], which contains array- and RNA-seq-derived expression measurements in a wide variety of tissues, to find *cis*-eQTLs for the DNA repair genes we implicated under the cosine distance peak on chromosome 6. On the GeneNetwork homepage (genenetwork.org), we selected the “BXD Family” **Group** and used the **Type** dropdown menu to select each of the specific expression datasets described in Table 2. In the **Get Any** text box, we then entered the listed gene name and clicked **Search**. After selecting the appropriate trait ID on the next page, we used the **Mapping Tools** dropdown to run Hayley-Knott regression [81] with default parameters: 1,000 permutations, interval mapping, no cofactors, and WGS-based genotypes (2022).

If we discovered a significant *cis*-eQTL for the gene of interest (that is, a locus on chromosome 6 with an LRS greater than or equal to the “significant LRS” genome-wide threshold), we then performed a second genome-wide association test for the trait of interest using GEMMA [82] with the following parameters: WGS-based marker genotypes, a minor allele frequency threshold of 0.05, and leave-one-chromosome-out (LOCO). By using both Haley-Knott regression and GEMMA, we could first discover loci that exceeded a genome-wide LRS threshold, and then more precisely estimate the effect of those loci on gene expression [83].

The exact names of the expression datasets we used for each tissue are shown in Table 2 below:

Calculating the frequencies of candidate mutator alleles in wild mice

To determine the frequencies of the *Ogg1* and *Setmar* nonsynonymous mutations in other populations of mice, we queried a VCF file containing genome-wide variation in 67 wild-derived mice from four species of *Mus* [32]. We calculated the allele frequency of each nonsynonymous

Tissue name	Complete name of GeneNetwork expression data
Kidney	Mouse kidney M430v2 Sex Balanced (Aug06) RMA
Gastrointestinal	UTHSC Mouse BXD Gastrointestinal Affy MoGene 1.0 ST Gene Level (Apr14) RMA
Hematopoietic stem cells	UMCG Stem Cells ILM6v1.1 (Apr09) transformed
Spleen	UTHSC Affy MoGene 1.0 ST Spleen (Dec10) RMA
Liver	UTHSC BXD Liver RNA-Seq Avg (Oct19) TPM Log2
Heart	NHLBI BXD All Ages Heart RNA-Seq (Nov20) TMP Log2 **
Hippocampus	Hippocampus Consortium M430v2 (Jun06) RMA

Table 2

Names of gene expression datasets used for each tissue type on GeneNetwork

mutation in each of the four species or subspecies (*Mus musculus domesticus*, *Mus musculus musculus*, *Mus musculus castaneus*, and *Mus spretus*), including genotypes that met the following criteria:

- supported by at least 10 sequencing reads
- Phred-scaled genotype quality of at least 20

Testing for epistasis between the two mutator loci

To test for statistical epistasis between the mutator loci on chromosome 4 and chromosome 6, we modeled C>A mutation rates in the BXDs as a function of genotypes at either locus:

$$\log \lambda^i = \beta_0 + \beta_1 X_1^i + \beta_2 X_2^i + \log A^i$$

Here, λ^i represents the count of C>A mutations in the i th BXD. We assume that the count of C>A mutations follows a Poisson distribution with a mean equal to the C>A mutation rate (expressed per base pair, per generation) multiplied by product of two terms: the total number of generations of inbreeding and the total number of base pairs accessible to variant calling. The BXDs differ in both their durations of inbreeding and the proportions of their genomes that were sequenced to sufficient depth, which influences the number of mutations we observe in each BXD. We therefore model C>A mutation counts as *rates* by including a $\log A^i$ term – sometimes referred to as an “offset” – which represents the product of the number of “callable” cytosine/guanine nucleotides in each BXD (i.e., the total number of cytosines/guanines covered by at least 10 sequencing reads) and the number of generations for which the BXD was inbred. The offset is assumed to have a coefficient of 1. The X_1^i and X_2^i terms represent BXD genotypes in the i th BXD at markers rs27509845 and rs46276051 (the markers with peak cosine distances on chromosomes 4 and 6 in the two aggregate mutation spectrum distance scans); *B* genotypes are coded as 0 and *D* genotypes are coded as 1. We limited our analysis to the $n = 108$ BXDs that were homozygous at both loci, allowing us to model genotypes at either locus as binary categorical variables. In the R statistical language, we defined the model as follows:

```
m1 <- glm(Count ~ offset(log(ADJ_AGE)) + Genotype_A + Genotype_B, data = data, family=poisson())
```

In this model, Count is the count of C>A *de novo* mutations observed in each BXD. The Genotype_A and Genotype_B terms are analogous to the X_1 and X_2 terms in the mathematical model above, and ADJ_AGE is analogous to the A term.

To explicitly test if the effects of *D* genotypes at the chromosome 6 locus depended on the presence of *D* genotypes at the chromosome 4 locus, we fit a second model incorporating a multiplicative interaction between X_1 and X_2 :

$$\log \lambda^i = \beta_0 + \beta_1 X_1^i + \beta_2 X_2^i + \beta_{12} X_1^i X_2^i + \log A^i$$

Here, the β_{12} term capture the interaction effect between genotypes at the two mutator loci. If both X_1^i and X_2^i are equal to 1 (i.e., genotypes at both the chromosome 4 and chromosome 6 loci are *D*), the β_{12} term will capture the resulting effect on C>A mutation counts. We can interpret a significantly non-zero β_{12} value as evidence for a non-additive effect of *D* genotypes at both loci.

We defined this model in the R statistical language as follows:

```
m2 <- glm(Count ~ offset(log(ADJ_AGE)) + Genotype_A * Genotype_B, data = data, family=poisson())
```

An identical model can also be written by explicitly specifying both the additive and interaction effects:

```
m2 <- glm(Count ~ offset(log(ADJ_AGE)) + Genotype_A + Genotype_B +  
Genotype_A:Genotype_B, data = data, family=poisson())
```

Using analysis of variance (ANOVA), we compared the model incorporating an interaction effect to the one with only additive effects:

```
anova(m2, m1, test="Chisq")
```

If model m2 is a significantly better fit to the data than m1, we can reject the null hypothesis that the effect of *D* genotypes at both markers is equal to the sum of the marginal effects of *D* genotypes at either rs27509845 or rs46276051. In other words, if m2 is a better fit than m1, then the combined effect of *D* genotypes at both markers is non-additive, and indicative of statistical epistasis.

We tested for statistical epistasis in the Sanger Mouse Genomes Project (MGP) strains using a nearly identical approach. In this analysis, we fit two models as follows:

```
m1 <- glm(Count ~ offset(log(CALLABLE_C)) + Genotype_A * Genotype_B,  
data = data, family=poisson())
```

```
m2 <- glm(Count ~ offset(log(CALLABLE_C)) + Genotype_A + Genotype_B,  
data = data, family=poisson())
```

where Count is the count of strain-private C>A mutations observed in each MGP strain [18]. The CALLABLE_C term represents the total number of cytosine and guanine nucleotides that were accessible for mutation calling in each strain, and the Genotype_A and Genotype_B terms represent MGP genotypes at the chromosome 4 and chromosome 6 mutator loci, respectively. We compared the two models using ANOVA as described above.

Acknowledgements

We thank Robert W. Williams (University of Tennessee Health Sciences Center) and Don F. Conrad (Oregon Health & Science University) for very helpful comments and feedback on a draft of this manuscript.

References

1. Chatterjee Nimrat, Walker Graham C (2017) **Mechanisms of DNA damage, repair, and mutagenesis** *Environmental and molecular mutagenesis* <https://doi.org/10.1002/em.22087>
2. Sasani Thomas A, Ashbrook David G, Beichman Annabel C, Lu Lu, Palmer Abraham A, Williams Robert W, Pritchard Jonathan K, Harris Kelley (2022) **A natural mutator allele shapes mutation spectrum variation in mice** *Nature* <https://doi.org/10.1038/s41586-022-04701-5>
3. Ashbrook David G *et al.* (2021) **A platform for experimental precision medicine: The extended BXD mouse family** *Cell systems* <https://doi.org/10.1016/j.cels.2020.12.002>
4. David Sheila S, O'Shea Valerie L, Kundu Sucharita (2007) **Base-excision repair of oxidative DNA damage** *Nature* <https://doi.org/10.1038/nature05978>
5. Jónsson Hákon *et al.* (2017) **Parental influence on human germline de novo mutations in 1,548 trios from Iceland** *Nature* <https://doi.org/10.1038/nature24018>
6. Sasani Thomas A, Pedersen Brent S, Gao Ziyue, Baird Lisa, Przeworski Molly, Jorde Lynn B, Quinlan Aaron R (2019) **Large, three-generation human families reveal post-zygotic mosaicism and variability in germline mutation accumulation** *eLife* <https://doi.org/10.7554/elife.46922>
7. Wang Richard J, Raveendran Muthuswamy, Harris RAlan, Murphy William J, Lyons Leslie A, Rogers Jeffrey, Hahn Matthew W (2022) **De novo Mutations in Domestic Cat are Consistent with an Effect of Reproductive Longevity on Both the Rate and Spectrum of Mutations** *Molecular biology and evolution* <https://doi.org/10.1093/molbev/msac147>
8. Wu Felix L, Strand Alva I, Cox Laura A, Ober Carole, Wall Jeffrey D, Moorjani Priya, Przeworski Molly (2020) **A comparison of humans and baboons suggests germline mutation rates do not track cell divisions** *PLoS biology* <https://doi.org/10.1371/journal.pbio.3000838>
9. Lindsay Sarah J, Rahbari Raheleh, Kaplanis Joanna, Keane Thomas, Hurles Matthew E (2019) **Similarities and differences in patterns of germline mutation between mice and humans** *Nature communications* <https://doi.org/10.1038/s41467-019-12023-w>
10. Rahbari Raheleh *et al.* (2015) **Timing, rates and spectra of human germline mutation** *Nature genetics* <https://doi.org/10.1038/ng.3469>
11. Lynch Michael, Ackerman Matthew S, Gout Jean-Francois, Long Hongan, Sung Way, Thomas WKelley, Foster Patricia L (2016) **Genetic drift, selection and the evolution of the mutation rate** *Nature reviews. Genetics* <https://doi.org/10.1038/nrg.2016.104>
12. Kessler Michael D *et al.* (2020) **De novo mutations across 1,465 diverse genomes reveal mutational insights and reductions in the Amish founder population** *Proceedings of the National Academy of Sciences of the United States of America* <https://doi.org/10.1073/pnas.1902766117>

13. Stendahl Alexandra M, Sanghvi Rashesh, Peterson Samuel, Ray Karina, Lima Ana C, Rahbari Raheleh, Conrad Donald F (2023) **A naturally occurring variant of *MBD4* causes maternal germline hypermutation in primates** *Cold Spring Harbor Laboratory* <https://doi.org/gr3ghz>
14. Kaplanis Joanna *et al.* (2022) **Genetic and chemotherapeutic influences on germline hypermutation** *Nature* <https://doi.org/10.1038/s41586-022-04712-2>
15. Robinson Philip S *et al.* (2021) **Increased somatic mutation burdens in normal human cells due to defective DNA polymerases** *Nature genetics* <https://doi.org/10.1038/s41588-021-00930-y>
16. Sherwood Kitty *et al.* (2023) **Germline de novo mutations in families with Mendelian cancer syndromes caused by defects in DNA repair** *Nature communications* <https://doi.org/10.1038/s41467-023-39248-0>
17. Seoighe Cathal, Scally Aylwyn (2017) **Inference of Candidate Germline Mutator Loci in Humans from Genome-Wide Haplotype Data** *PLoS genetics* <https://doi.org/10.1371/journal.pgen.1006549>
18. Dumont Beth L (2019) **Significant Strain Variation in the Mutation Spectra of Inbred Laboratory Mice** *Molecular biology and evolution* <https://doi.org/10.1093/molbev/msz026>
19. Halligan Daniel L, Keightley Peter D (2009) **Spontaneous Mutation Accumulation Studies in Evolutionary Genetics** *Annual Review of Ecology, Evolution, and Systematics* <https://doi.org/dvrjz8>
20. Robinson Philip S *et al.* (2022) **Inherited MUTYH mutations cause elevated somatic mutation rates and distinctive mutational signatures in normal human cells** *Nature communications* <https://doi.org/10.1038/s41467-022-31341-0>
21. Carlson Jedidiah, DeWitt William S, Harris Kelley (2020) **Inferring evolutionary dynamics of mutation rates through the lens of mutation spectrum variation** *Current opinion in genetics & development* <https://doi.org/10.1016/j.gde.2020.05.024>
22. Viel Alessandra *et al.* (2017) **A Specific Mutational Signature Associated with DNA 8-Oxoguanine Persistence in MUTYH-defective Colorectal Cancer** *EBioMedicine* <https://doi.org/10.1016/j.ebiom.2017.04.022>
23. Pilati Camilla *et al.* (2017) **Mutational signature analysis identifies MUTYH deficiency in colorectal cancers and adrenocortical carcinomas** *The Journal of pathology* <https://doi.org/10.1002/path.4880>
24. Consortium The Gene Ontology *et al.* (2000) **Gene ontology: tool for the unification of biology** *Nature genetics* <https://doi.org/10.1038/75556>
25. research Nucleic acids (2021) **The Gene Ontology resource: enriching a GOld mine. Nucleic acids research (2021-01-08)** <https://www.ncbi.nlm.nih.gov/pubmed/33290552> DOI: 10.1093/nar/gkaa1113 PMID: 33290552 PMCID: PMC7779012 *The Gene Ontology resource: enriching a GOld mine* <https://doi.org/10.1093/nar/gkaa1113>
26. Manichaikul Ani, Dupuis Josée, Sen Saunak, Broman Karl W (2006) **Poor performance of bootstrap confidence intervals for the location of a quantitative trait locus** *Genetics* <https://doi.org/10.1534/genetics.106.061549>

27. Mulligan Megan K, Mozhui Khyobeni, Prins Pjotr, Williams Robert W (2017) **GeneNetwork: A Toolbox for Systems Genetics** *Methods in molecular biology (Clifton, N.J)* https://doi.org/10.1007/978-1-4939-6427-7_4
28. Ferraj Ardian *et al.* (2023) **Resolution of structural variation in diverse mouse genomes reveals chromatin remodeling due to transposable elements** *Cell Genomics* <https://doi.org/10.1016/j.celgen.2023.100776>
29. Islam SMAshiquil *et al.* (2022) **Uncovering novel mutational signatures by** *Cell genomics* <https://doi.org/10.1016/j.xgen.2022.100179>
30. Tate John G *et al.* (2019) **COSMIC: the Catalogue Of Somatic Mutations In Cancer** *Nucleic acids research* <https://doi.org/10.1093/nar/gky1015>
31. Keane Thomas M *et al.* (2011) **Mouse genomic variation and its effect on phenotypes and gene regulation** *Nature* <https://doi.org/10.1038/nature10413>
32. Harr Bettina *et al.* (2016) **Genomic resources for wild populations of the house mouse, *Mus musculus* and its close relative *Mus spretus*** *Scientific data* <https://doi.org/10.1038/sdata.2016.75>
33. Yang Hyuna, Bell Timothy A, Churchill Gary A, de Villena Fernando Pardo-Manuel (2007) **On the subspecific origin of the laboratory mouse** *Nature genetics* <https://doi.org/10.1038/ng2087>
34. Heck Julie Akiko, Argueso Juan Lucas, Gemici Zekeriyya, Reeves Richard Guy, Bernard Ann, Aquadro Charles F, Alani Eric (2006) **Negative epistasis between natural variants of the *Saccharomyces cerevisiae* MLH1 and PMS1 genes results in a defect in mismatch repair** *Proceedings of the National Academy of Sciences of the United States of America* <https://doi.org/10.1073/pnas.0510998103>
35. Petljak Mia *et al.* (2022) **Mechanisms of APOBEC3 mutagenesis in human cancer cells** *Nature* <https://doi.org/10.1038/s41586-022-04972-y>
36. Bergeron Lucie A *et al.* (2023) **Evolution of the germline mutation rate across vertebrates** *Nature* <https://doi.org/10.1038/s41586-023-05752-y>
37. Fnu Sheema, Williamson Elizabeth A, De Haro Leyma P, Brenneman Mark, Wray Justin, Shaheen Montaser, Radhakrishnan Krishnan, Lee Suk-Hee, Nickoloff Jac A, Hromas Robert (2010) **Methylation of histone H3 lysine 36 enhances DNA repair by nonhomologous end-joining** *Proceedings of the National Academy of Sciences of the United States of America* <https://doi.org/10.1073/pnas.1013571108>
38. Lee Suk-Hee, Oshige Masahiko, Durant Stephen T, Rasila Kanwaldeep Kaur, Williamson Elizabeth A, Ramsey Heather, Kwan Lori, Nickoloff Jac A, Hromas Robert (2005) **The SET domain protein Metnase mediates foreign DNA integration and links integration to nonhomologous end-joining repair** *Proceedings of the National Academy of Sciences of the United States of America* <https://doi.org/10.1073/pnas.0503676102>
39. Kim Hyun-Suk, Chen Qiujia, Kim Sung-Kyung, Nickoloff Jac A, Hromas Robert, Georgiadis Millie M, Lee Suk-Hee (2014) **The DDN catalytic motif is required for Metnase functions in non-homologous end joining (NHEJ) repair and replication restart** *The Journal of biological chemistry* <https://doi.org/10.1074/jbc.M113.533216>

40. Beck Brian D, Lee Sung-Sook, Williamson Elizabeth, Hromas Robert A, Lee Suk-Hee (2011) **Biochemical characterization of metnase's endonuclease activity and its role in NHEJ repair** *Biochemistry* <https://doi.org/10.1021/bi200333k>
41. Markkanen Enni (2017) **Not breathing is not an option: How to deal with oxidative DNA damage** *DNA repair* <https://doi.org/10.1016/j.dnarep.2017.09.007>
42. Ohno Mizuki, Sakumi Kunihiro, Fukumura Ryutaro, Furuichi Masato, Iwasaki Yuki, Hokama Masaaki, Ikemura Toshimichi, Tsuzuki Teruhisa, Gondo Yoichi, Nakabeppu Yusaku (2014) **8-oxoguanine causes spontaneous de novo germline mutations in mice** *Scientific reports* <https://doi.org/10.1038/srep04689>
43. Sakumi K, Furuichi M, Tsuzuki T, Kakuma T, Kawabata S, Maki H, Sekiguchi M (1993) **Cloning and expression of cDNA for a human enzyme that hydrolyzes 8-oxo-dGTP, a mutagenic substrate for DNA synthesis** *The Journal of biological chemistry*
44. Mahjabeen Ishrat, Masood Nosheen, Baig Ruqia Mehmood, Sabir Maimoona, Inayat Uzma, Malik Faraz Arshad, Kayani Mahmood Akhtar (2012) **Novel mutations of OGG1 base excision repair pathway gene in laryngeal cancer patients** *Familial cancer* <https://doi.org/10.1007/s10689-012-9554-2>
45. Chevillard S, Radicella JP, Levalois C, Lebeau J, Poupon MF, Oudard S, Dutrillaux B, Boiteux S (1998) **Mutations in OGG1, a gene involved in the repair of oxidative DNA damage, are found in human lung and kidney tumours** *Oncogene* <https://doi.org/10.1038/sj.onc.1202096>
46. Boogaard L van den *et al.* (2021) **Defects in 8-oxo-guanine repair pathway cause high frequency of C > A substitutions in neuroblastoma Marlinde** *Proceedings of the National Academy of Sciences* <https://doi.org/grtcs9>
47. Ng Pauline C, Henikoff Steven (2003) **SIFT: Predicting amino acid changes that affect protein function** *Nucleic acids research* <https://doi.org/10.1093/nar/gkg509>
48. Karczewski Konrad J *et al.* (2020) **The mutational constraint spectrum quantified from variation in 141,456 humans** *Nature* <https://doi.org/10.1038/s41586-020-2308-7>
49. Adzhubei Ivan A, Schmidt Steffen, Peshkin Leonid, Ramensky Vasily E, Gerasimova Anna, Bork Peer, Kondrashov Alexey S, Sunyaev Shamil R (2010) **A method and server for predicting damaging missense mutations** *Nature methods* <https://doi.org/10.1038/nmeth0410-248>
50. Cortellino Salvatore *et al.* (2003) **The base excision repair enzyme MED1 mediates DNA damage response to antitumor drugs and is associated with mismatch repair system integrity** *Proceedings of the National Academy of Sciences of the United States of America* <https://doi.org/10.1073/pnas.2334585100>
51. Sansom Owen James, Zabkiewicz Joanna, Bishop Stefan Mark, Guy Jackie, Bird Adrian, Clarke Alan Richard (2003) **MBD4 deficiency reduces the apoptotic response to DNA-damaging agents in the murine small intestine** *Oncogene* <https://doi.org/10.1038/sj.onc.1206850>
52. Millar Catherine B, Guy Jacky, Sansom Owen J, Selfridge Jim, MacDougall Eilidh, Hendrich Brian, Keightley Peter D, Bishop Stefan M, Clarke Alan R, Bird Adrian (2002) **Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice** *Science (New York, N.Y.)* <https://doi.org/10.1126/science.1073354>

53. Bader SA, Walker M, Harrison DJ (2007) **A human cancer-associated truncation of MBD4 causes dominant negative impairment of DNA repair in colon cancer cells** *British journal of cancer* <https://doi.org/10.1038/sj.bjc.6603592>
54. Broman Karl W, Gatti Daniel M, Simecek Petr, Furlotte Nicholas A, Prins Pjotr, Sen Saunak, Yandell Brian S, Churchill Gary A (2018) **R/qtl2: Software for Mapping Quantitative Trait Loci with High-Dimensional Data and Multiparent Populations** *Genetics* <https://doi.org/10.1534/genetics.118.301595>
55. Werling Donna M *et al.* (2018) **An analytical framework for whole-genome sequence association studies and its implications for autism spectrum disorder** *Nature genetics* <https://doi.org/10.1038/s41588-018-0107-y>
56. Milligan William R, Amster Guy, Sella Guy (2022) **The impact of genetic modifiers on variation in germline mutation rates within and among human populations** *Genetics* <https://doi.org/10.1093/genetics/iyac087>
57. Arnheim Norman, Calabrese Peter (2009) **Understanding what determines the frequency and pattern of human germline mutations** *Nature reviews. Genetics* <https://doi.org/10.1038/nrg2529>
58. Churchill GA, Doerge RW (1994) **Empirical threshold values for quantitative trait mapping** *Genetics* <https://doi.org/10.1093/genetics/138.3.963>
59. Visscher PM, Thompson R, Haley CS (1996) **Confidence intervals in QTL mapping by bootstrapping** *Genetics* <https://doi.org/10.1093/genetics/143.2.1013>
60. Stafford Alexandra M, Reed Cheryl, Baba Harue, Walter Nicole Ar, Mootz John Rk, Williams Robert W, Neve Kim A, Fedorov Lev M, Janowsky Aaron J, Phillips Tamara J (2019) **aar1 gene variants have a causal role in methamphetamine intake and response and interact with Oprm1** *eLife* <https://doi.org/10.7554/elife.46472>
61. Mulligan Megan K *et al.* (2019) **Identification of a Functional Non-coding Variant in the GABA** *Frontiers in genetics* <https://doi.org/10.3389/fgene.2019.00188>
62. Lu Ye, Zhou Diana, Lu Hong, Xu Fuyi, Yue Junming, Tong Jianping, Lu Lu (2019) **Investigating a downstream gene of Molecular vision**
63. Harris Charles R *et al.* (2020) **Array programming with NumPy** *Nature* <https://doi.org/ghbzf2>
64. (2023) **pandas-dev/pandas: Pandas The Pandas Development Team Zenodo** <https://doi.org/ggt8bh>
65. Hunter John D (2007) **Matplotlib: A 2D Graphics Environment** *Computing in Science & Engineering* <https://doi.org/drbjhg>
66. Pedregosa Fabian *et al.* (2011) **Scikit-learn: Machine Learning in Python** *Journal of Machine Learning Research*
67. Bantilan Niels (2020) **pandera: Statistical Data Validation of Pandas Dataframes** *Proceedings of the Python in Science Conference* <https://doi.org/grr54q>
68. Waskom Michael (2021) **seaborn: statistical data visualization** *Journal of Open Source Software* <https://doi.org/gjqn3g>

69. Lam Siu Kwan, Pitrou Antoine, Seibert Stanley (2015) **Numba** *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC* <https://doi.org/gf3nks>
70. Mölder Felix *et al.* (2021) **Sustainable data analysis with Snakemake** *F1000Research* <https://doi.org/gjjkwv>
71. Sasani Tom (2023) **A natural mutator allele shapes mutation spectrum variation in mice**
72. tomsasani/bxd_mutator_manuscript: Final figure generation updates prior to publication Tom Sasani (2022) **tomsasani/bxd_mutator_manuscript: Final figure generation updates prior to publication** Tom Sasani Zenodo (2022-02-01) grrwv8 DOI: 10.5281/zenodo.5941048 <https://doi.org/grrwv8>
73. **BXD Genotype / WebQTL**
74. **Genetics and Probability in Animal Breeding Experiments**
75. ENA Browser **ENA Browser** <https://www.ebi.ac.uk/ena/browser/view/PRJEB45429>
76. Danecek Petr *et al.* (2021) **Twelve years of SAMtools and BCFtools** *GigaScience* <https://doi.org/10.1093/gigascience/giab008>
77. Cingolani Pablo, Platts Adrian, Wang Le Lily, Coon Melissa, Nguyen Tung, Wang Luan, Land Susan J, Lu Xiangyi, Ruden Douglas M (2012) **A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3** *Fly* <https://doi.org/10.4161/fly.19695>
78. Pedersen Brent S, Quinlan Aaron R (2017) **cyvcf2: fast, flexible variant analysis with Python** *Bioinformatics (Oxford, England)* <https://doi.org/10.1093/bioinformatics/btx057>
79. Karolchik Donna, Hinrichs Angela S, Furey Terrence S, Roskin Krishna M, Sugnet Charles W, Haussler David, Kent WJames (2004) **The UCSC Table Browser data retrieval tool** *Nucleic acids research* <https://doi.org/10.1093/nar/gkh103>
80. (2023) **bx-python** Taylor Lab at Johns Hopkins University
81. Haley CS, Knott SA (1992) **A simple regression method for mapping quantitative trait loci in line crosses using flanking markers** *Heredity* <https://doi.org/10.1038/hdy.1992.131>
82. Girolami A, Santarossa A, Martinelli S, Sartori MT, Visentin I (1987) **Study on a new chromogenic substrate for the prothrombin time determination** *Folia haematologica (Leipzig, Germany : 1928)*
83. Watson Pamela M, Ashbrook David G (2020) **GeneNetwork: a continuously updated tool for systems genetics analyses** *Cold Spring Harbor Laboratory* <https://doi.org/gsnpr5>
84. Cordaux Richard, Udit Swalpa, Batzer Mark A, Feschotte Cédric (2006) **Birth of a chimeric primate gene by capture of the transposase gene from a mobile element** *Proceedings of the National Academy of Sciences of the United States of America* <https://doi.org/10.1073/pnas.0601161103>
85. Tellier Michael, Chalmers Ronald (2019) **The roles of the human SETMAR (Metnase) protein in illegitimate DNA recombination and non-homologous end joining repair** *DNA repair* <https://doi.org/10.1016/j.dnarep.2019.06.006>

Article and author information

Thomas A. Sasani

Department of Human Genetics, University of Utah
ORCID iD: [0000-0003-2317-1374](https://orcid.org/0000-0003-2317-1374)

Aaron R. Quinlan

Department of Human Genetics, University of Utah; Department of Biomedical Informatics, University of Utah

For correspondence: aquinlan@genetics.utah.edu
ORCID iD: [0000-0003-1756-0859](https://orcid.org/0000-0003-1756-0859)

Kelley Harris

Department of Genome Sciences, University of Washington
For correspondence: harriske@uw.edu
ORCID iD: [0000-0003-0302-2523](https://orcid.org/0000-0003-0302-2523)

Copyright

© 2023, Sasani et al.

This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

Editors

Reviewing Editor

Ziyue Gao

University of Pennsylvania, Philadelphia, United States of America

Senior Editor

Detlef Weigel

Max Planck Institute for Biology Tübingen, Tübingen, Germany

Reviewer #1 (Public Review):

The mutation rate and spectrum have been found to differ between populations as well as across individuals within the same population. Hypothesizing that some of the observed variation has a genetic basis, the authors of this paper have made important contributions in the past few years in identifying genetic variants that modify mutation rate or spectrum in natural populations. This paper makes one significant step further by developing a new method for mapping genetic variants associated with the mutation spectrum, which reveals new biological insights.

Using traditional quantitative trait locus (QTL) mapping in the BXD mouse recombinant inbred lines (RILs), the authors of this paper previously identified a genetic locus associated with C>A mutation rate. However, this approach has limited power, as it suffers from multiple testing burden as well as noise in the "observed mutation spectrum phenotype" due to rarity and randomness of mutation events. To overcome these limitations, the authors developed a new method that they named "aggregate mutation spectrum distance" (AMSD),

which in short measures the difference in the aggregate mutation spectrum between two groups of individuals with distinct genotypes at a specific genomic locus. With this new approach, they recover the previously reported candidate mutator locus (near *Mutyh* gene) and identify a new candidate locus that modifies the C>A mutation rate on only the mutator allele genetic background at the *Mutyh* locus. Using more rigorous statistical testing, the authors show convincingly synergistic epistatic effects between the mutator alleles at the two loci.

Overall, the analyses presented are well done and provide convincing evidence for the major findings, including the new candidate mutator locus and its epistatic interaction with the *Mutyh* locus. The new AMSD method introduced is innovative and outperforms traditional QTL mapping under most conditions, as demonstrated by extensive simulations. I identify no major issues with this paper and think it is very well written.

One of the major advantages of the AMSD method over QTL mapping is alleviation of the multiple testing burden, as one comparison tests for any changes in the mutation spectrum, including simultaneous, small changes in the relative abundance of multiple mutation types. The flip side of this advantage of AMSD is that, when a significant association is detected, it is not immediately clear which mutation type is driving the signal. To narrow the signal to specific candidate mutation type(s), additional analyses are needed, such as testing for differential proportions of each mutation type between individuals with or without the candidate mutator allele. However, such analysis may be less powerful when the mutator allele leads to small changes in the relative abundance of multiple mutation types. This will be an area of improvement for future studies.

- <https://doi.org/10.7554/eLife.89096.2.sa2>

Reviewer #2 (Public Review):

In this paper Sasani, Quinlan and Harris present a new method for identifying genetic factors affecting germline mutation, which is particularly applicable to genome sequence data from mutation accumulation experiments using recombinant inbred lines. These are experiments where laboratory organisms are crossed and repeatedly inbred for many generations, to build up a substantial number of identifiable germline mutations. The authors apply their method to such data from mice, and identify two genetic factors at two separate genetic loci. Clear evidence of such factors has been difficult to obtain, so this is an important finding. They further show evidence of an epistatic interaction between these factors (meaning that they do not act independently in their effects on the germline mutation process). This is exciting because such interactions are difficult to detect and few if any other examples have been studied.

The authors present a careful comparison of their method to another similar approach, quantitative trait locus (QTL) analysis, and demonstrate that in situations such as the one analysed it has greater power to detect genetic factors with a certain magnitude of effect. They also test the statistical properties of their method using simulated data and permutation tests. Overall the analysis is rigorous and well motivated, and the methods explained clearly.

The main limitation of the approach is that it is difficult to see how it might be applied beyond the context of mutation accumulation experiments using recombinant inbred lines. This is because the signal it detects, and hence its power, is based on the number of extra accumulated mutations linked to (i.e. on the same chromosome as) the mutator allele. In germline mutation studies of wild populations the number of generations involved (and hence the total number of mutations) is typically small, or else the mutator allele becomes unlinked from the mutations it has caused (due to recombination), or is lost from the

population altogether (due to chance or perhaps selection against its deleterious consequences).

Nevertheless, accumulation lines are a common and well established experimental approach to studying mutation processes in many organisms, so the new method could have wide application and impact on our understanding of this fundamental biological process.

The evidence presented for an epistatic interaction is convincing, and the authors suggest some plausible potential mechanisms for how this interaction might arise, involving the DNA repair machinery and based on previous studies of the proteins implicated. However as with all such findings, given the higher degree of complexity of the proposed model it needs to be treated with greater caution, perhaps until replicated in a separate dataset or demonstrated in follow-up experiments exploring the pathway itself.

- <https://doi.org/10.7554/eLife.89096.2.sa1>

Reviewer #3 (Public Review):

Sasani et al. develop and implement a new method for mutator allele discovery in the BXD mouse population. This new method, termed "aggregate mutation spectrum distance" or AMSD, carries several notable strengths, including the ability to aggregate de novo mutations across individuals to reduce data sparsity and to combine mutation rate frequencies across multiple nucleotide contexts into a single estimate. As demonstrated by simulations, this method is better suited to mutator discovery under certain scenarios, as compared to conventional QTL or association mapping. Overall, the theoretical premise of the AMSD method is judged to be both strong and innovative, and the methodology could be extended to other species and populations to enable discovery of additional mutator alleles.

The authors then apply their method to the BXD mouse recombinant inbred mapping population. As proof-of-principle, they first successfully re-identify a known mutator locus in this population on chr4. Next, to assess possible genetic interactions involving this known mutator, Sasani et al. condition on the chr4 mutator genotype and reimplement the AMSD scan. This strategy led them to identify a second locus on chr6 that interacts epistatically with the chr4 locus; mice with "D" alleles at both loci exhibit a significantly increased burden of C>A de novo mutations, even though mice with the D allele at the chr6 locus alone show no appreciable increase in the C>A mutation fraction. This exciting discovery not only adds to the catalog of known mutator alleles, but also reveals key aspects of mutator biology and reinforces the hypothesis that segregating variants in genes associated with DNA repair influence germline mutation spectra.

Despite a high level of overall enthusiasm for this work, there are some limitations to the AMSD method. However, it is my judgement that the authors present a balanced summary of the strengths and weaknesses of their method in the revised manuscript. I also think that the authors' conclusions may actually somewhat undersell the scientific impact of their findings. As the authors note, few mutation rate modifiers have been identified in mammals. This is potentially because large- and moderate-effect modifiers are rapidly selected against due to their deleterious effects, but could also be due to pervasive epistasis wherein modifiers are only expressed on certain "permissive" genetic backgrounds, such as the chr6 locus the authors discover in this paper. The potential background dependence of mutator expression could partially shelter it from the action of selection, allowing the allele persist in populations. This discovery has significant implications for our understanding of mutation rate evolution, but only earns a cursory mention in the paper.

- <https://doi.org/10.7554/eLife.89096.2.sa0>

Author Response

The following is the authors' response to the original reviews.

Reviewer 1 (Public Review):

1. The name of the new method "inter-haplotype distance" is more confusing than helpful, as the haplotype information is not critical for implementing this method. First, the mutation spectrum is aggregated genome-wide regardless of the haplotypes where the mutations are found. Second, the only critical haplotype information is that at the focal site (i.e., the locus that is tested for association): individuals are aggregated together when they belong to the same "haplotype group" at the focal site. However, for the classification step, haplotype information is not really necessary: individuals can be grouped based on their genotypes at the given locus (e.g., AA vs AB). As the authors mentioned, this method can be potentially applied to other mutation datasets, where haplotype information may well be unavailable. I hope the authors can reconsider the name and remove the term "haplotype" (perhaps something like "inter-genotype distance"?) to avoid giving the wrong impression that haplotype information is critical for applying this method.

We appreciate the reviewer's concern about the name of our method. The reviewer is correct that haplotype information is not critical for our method to work, and as a result we've decided to simply rename the approach to "aggregate mutation spectrum distance" (abbreviated AMSD). For simplicity, we refer to the method as IHD throughout our responses to reviewers, but the revised manuscript now refers to AMSD.

1. The biggest advantage of the IHD method over QTL mapping is alleviation of the multiple testing burden, as one comparison tests for any changes in the mutation spectrum, including simultaneous, small changes in the relative abundance of multiple mutation types. Based on this, the authors claim that IHD is more powerful to detect a mutator allele that affects multiple mutation types. Although logically plausible, it is unclear under what quantitative conditions IHD can actually have greater power over QTL. It will be helpful to support this claim by providing some simulation results.

This comment prompted us to do a more detailed comparison of IHD vs. QTL power under conditions that are more similar to those observed in the BXD cohort. While preparing the original manuscript, we assumed that IHD might have greater power than QTL mapping in a population like the BXDs because some recombinant inbred lines have accumulated many more germline mutations than others (see Figure 1 in Sasani et al. 2022, Nature). In a quantitative trait locus scan (say, for the fraction of C>A mutations in each line) each BXD's mutation data would be weighted equally, even if a variable number of mutations was used to generate the phenotype point estimate in each line.

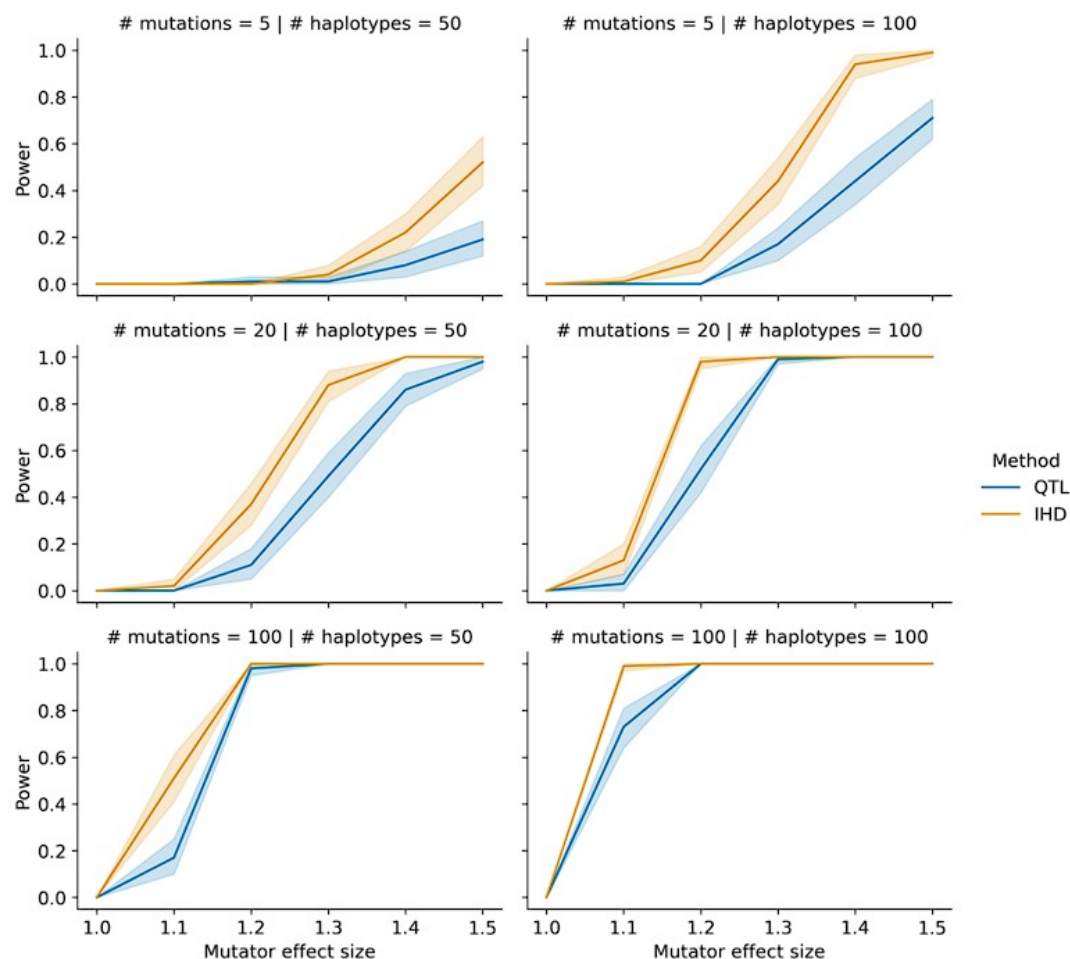
To address this, we performed a new series of simulations in which the average number of mutations per haplotype was allowed to vary. At the low end, some BXDs accumulated as few as 100 total germline mutations, while others have accumulated as many as 2,000. Thus, instead of simulating a mean number of mutations on each simulated haplotype, we allowed the mean number of mutations per haplotype to vary from N to 20N. By simulating a variable count of mutations on each haplotype, we could more easily test the benefits of comparing aggregate, rather than individual, mutation spectra between BXDs.

In these updated simulations, we find that IHD routinely outperforms QTL mapping under a range of parameter choices (see Author Response image 1). Since IHD aggregates the mutation spectra of all haplotypes with either B or D alleles at each locus in the genome, the

method is much less sensitive to individual haplotypes with low mutation counts. We include a mention of these updated simulations on lines 135-138 and describe the updated simulations in greater detail in the Materials and Methods (lines 705-715).

Author response image 1.

Power of IHD and QTL mapping on simulated haplotypes with variable counts of mutations. We simulated germline mutations on the specified number of haplotypes (as described in the manuscript) but allowed the total number of mutations per haplotype to vary by a factor of 20.



1. The flip side of this advantage of IHD is that, when a significant association is detected, it is not immediately clear which mutation type is driving the signal. Related to this, it is unclear how the authors reached the point that "...the C>A mutator phenotype associated with the locus on chromosome 6", when they only detected significant IHD signal at rs46276051 (on Chr6), when conditioning on D genotypes at the rs27509845 (on Chr4) and no significant signal for any 1-mer mutation type by traditional mapping. The authors need to explain how they deduced that C>A mutation is the major source of the signal. In addition, beyond C>A mutations, can mutation types other than C>A contribute to the IHD signal at rs46276051? More generally, I hope the authors can provide some guidelines on how to narrow a significant IHD signal to specific candidate mutation type(s) affected, which will make the method more useful to other researchers.

We thank the reviewer for pointing out this gap in our logic. We omitted specific instructions for narrowing down an IHD signal to specific mutation type(s) for a few reasons. First, this can be addressed using mutational signature analysis methods that are in widespread use. For example, upon identifying one or more candidate mutator loci, we can enter the mutation spectra of samples with each possible mutator genotype into a program (e.g., SigProfilerExtractor) to determine which combinations of mutation types occur proportionally more often in the genomes that harbor mutators (see Figure 3c in our manuscript). A second approach for narrowing down an IHD signal, highlighted in Figure 3a (and now described in the text of the Results section at lines 256-261), is to simply test which mutation type proportion(s) differ significantly between groups of samples with and without a candidate mutator (for example, with a Chi-square test of independence for each mutation type).

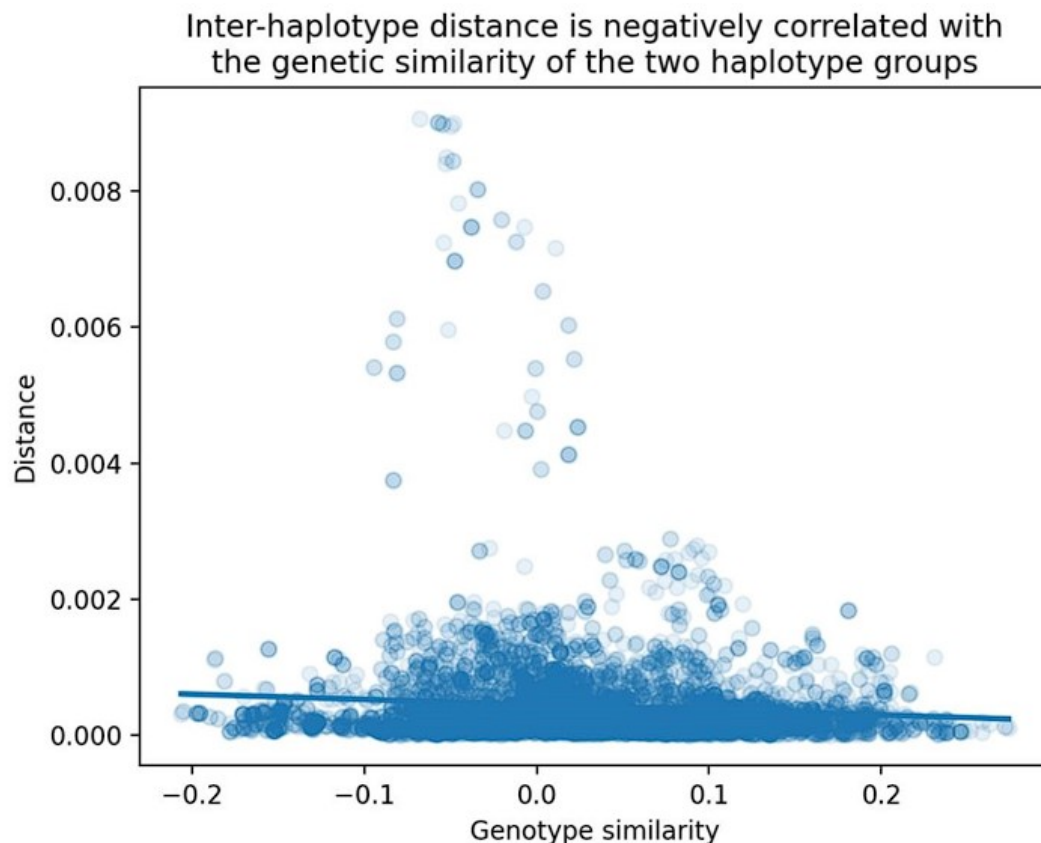
Although this second approach incurs a multiple testing burden, the burden is offset somewhat by using IHD to identify mutator loci, rather than performing association tests for every possible mutation type to begin with. Although Figure 3a only shows the significant difference in C>A fraction among BXDs with different mutator locus genotypes, Figure 3-figure supplement 1 shows the complete set of 1-mer spectrum comparisons. It is possible that this second approach would not prove very useful in the case of a mutator with a “flat” signature (i.e., a mutator that slightly perturbs the rates of many different mutation types), but in our case it clearly shows which mutation type is affected.

1. To account for differential relatedness between the inbred lines, the authors regressed the cosine distance between the two aggregate mutation spectra on the genome-wide genetic similarity and took the residual as the adjusted test metric. What is the value of the slope from this regression? If significantly non-zero, this would support a polygenic architecture of the mutation spectrum phenotype, which could be interesting. If not, is this adjustment really necessary? In addition, is the intercept assumed to be zero for this regression, and does such an assumption matter? I would appreciate seeing a supplemental figure on this regression.

The reviewer raises a good question. We find that the slope of the “distance vs. genetic similarity” regression is significantly non-zero, though the slope estimate itself is small. A plot of cosine distance vs. genome-wide genetic similarity (using all BXDs) is shown below in Author response image 2:

Author response image 2.

Relationship between cosine distance and genetic similarity in the BXDs. As described in the Materials and Methods, we computed two values at each marker in the BXDs: 1) the cosine distance between the aggregate mutation spectra of BXDs with either B or D genotypes at the marker, and 2) the correlation between genome-wide D allele frequencies in BXDs with either B or D genotypes at the marker. We then regressed these two values across all genome-wide markers.

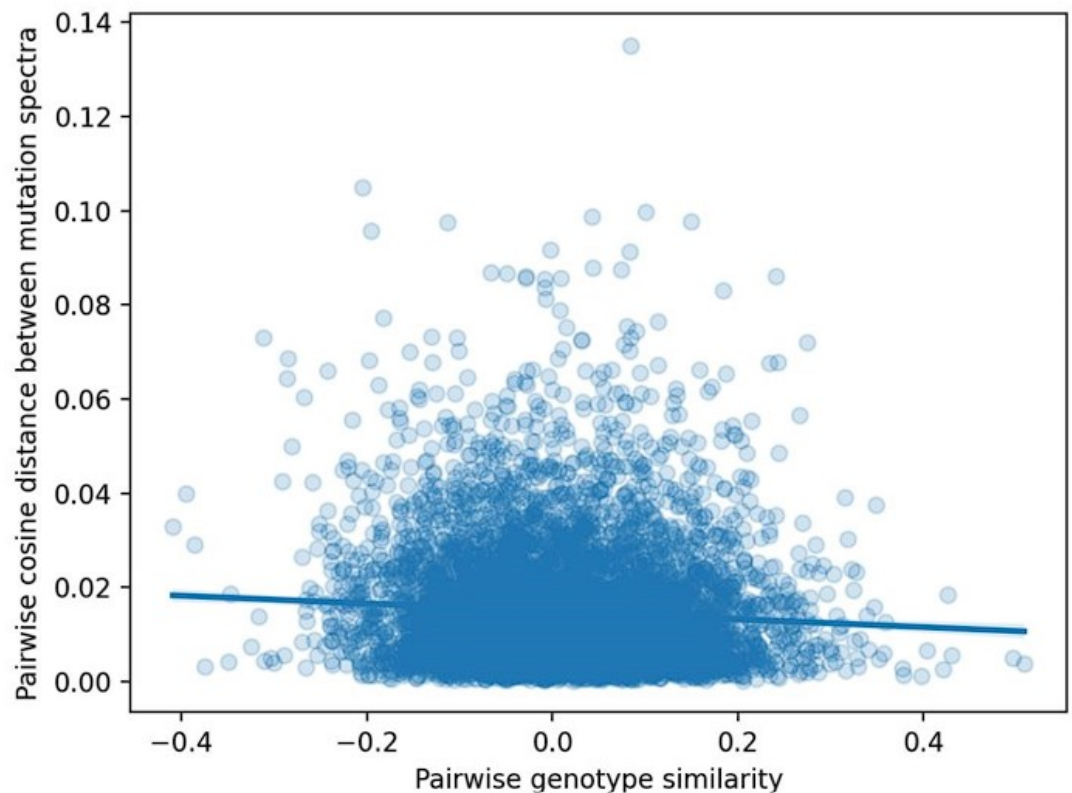


This result indicates that if two groups of BXDs (one with D genotypes and one with B genotypes at a given locus) are more genetically similar, their mutation spectra are also more similar. Since the regression slope estimate is significantly non-zero ($p < 2.2e-16$), we believe that it's still worth using residuals as opposed to raw cosine distance values. This result also suggests that there may be a polygenic effect on the mutation spectrum in the BXDs.

We have also generated a plot showing the cosine distance between the mutation spectra of every possible pair of BXDs, regressed against the genetic similarity between each of those pairs (Author Response image 3). Here, the potential polygenic effects on mutation spectra similarity are perhaps more obvious.

Author response image 3.

Pairwise cosine distance between BXD mutation spectra as a function of genetic similarity. We computed two values for every possible pair of $n = 117$ BXDs: 1) the cosine distance between the samples' individual 1-mer mutation spectra and 2) the correlation coefficient between the samples' genome-wide counts of D alleles.



Private Comments

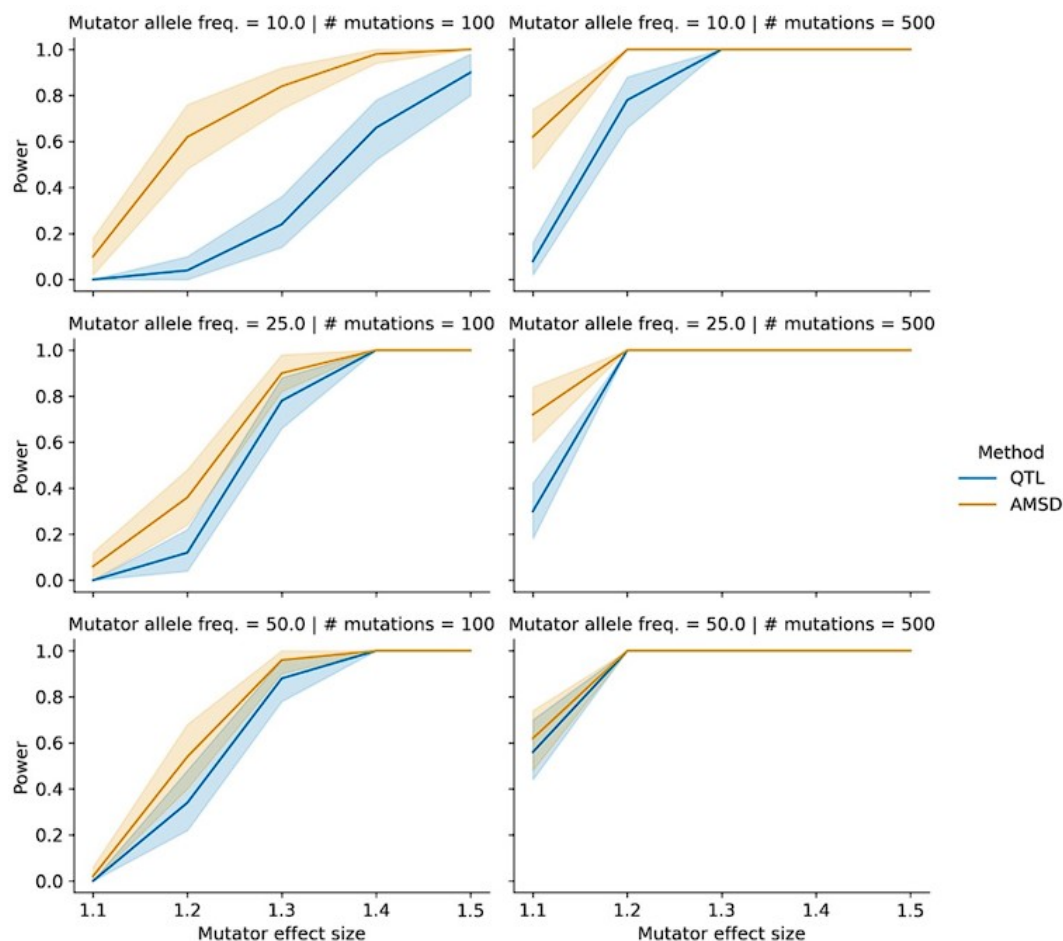
1. It will also be useful to see how the power of IHD and QTL mapping depend on the allele frequency of the mutator allele and the sample size, as mutator alleles are likely rare or semi-rare in natural populations (such as the human de novo mutation dataset that the authors mentioned).

This is another good suggestion. In general, we'd expect the power of both IHD and QTL mapping to decrease as a function of mutator allele frequency. At the same time, we note that the power of these scans should mostly depend on the absolute number of carriers of the mutator allele and less on its frequency. In the BXD mouse study design, we observe high frequency mutators but also a relatively small sample size of just over 100 individuals. In natural human populations, mutator frequencies might be orders of magnitude smaller, but sample sizes may be orders of magnitude larger, especially as new cohorts of human genomes are routinely being sequenced. So, we expect to have similar power to detect a mutator segregating at, say, 0.5% frequency in a cohort of 20,000 individuals, as we would to detect a mutator segregating at 50% frequency in a dataset of 200 individuals.

To more formally address the reviewer's concern, we performed a series of simulations in which we simulated a population of 100 haplotypes. We assigned the same average number of mutations to each haplotype but allowed the allele frequency of the mutator allele to vary between 0.1, 0.25, and 0.5. The results of these simulations are shown in Author response image 4 and reveal that AMSD tends to have greater power than QTL mapping at lower mutator allele frequencies. We now mention these simulations in the text at lines 135-138 and include the simulation results in Figure 1-figure supplement 4.

Author response image 4.

Power of AMSD and QTL mapping on simulated haplotypes with variable marker allele frequencies. We simulated germline mutations on the specified number of haplotypes (as described in the manuscript), but simulated genotypes at the mutator allele such that "A" alleles were at the specified allele frequency.



1. In the Methods section of "testing for epistasis between the two mutator loci", it will be helpful to explicitly lay out the model and assumptions in mathematical formulae, in addition to the R scripts. For example, are the two loci considered independent when their effects on mutation rate is multiplicative or additive? Given the R scripts provided, it seems that the two loci are assumed to have multiplicative effects on the mutation rate, and that the mutation count follows a Poisson distribution with mean being the mutation rate times ADJ_AGE (i.e., the mutation opportunity times the number of generations of an inbred line). However, this is not easily understandable for readers who are not familiar with R language. In addition, I hope the authors can be more specific when discussing the epistatic interaction between the two loci by explicitly saying "synergistic effects beyond multiplicative effects on the C>A mutation rate".

The reviewer raises a good point about the clarity of our descriptions of tests for epistasis. We have now added a more detailed description of these tests in the section of the Materials and Methods beginning at line 875. We have also added a statement to the text at lines 289-291: "the combined effects of D genotypes at both loci exceed the sum of marginal effects of D

genotypes at either locus alone.” We hope that this will help clarify the results of our tests for statistical epistasis.

Reviewer 2 (Public Review):

1. The main limitation of the approach is that it is difficult to see how it might be applied beyond the context of mutation accumulation experiments using recombinant inbred lines. This is because the signal it detects, and hence its power, is based on the number of extra accumulated mutations linked to (i.e. on the same chromosome as) the mutator allele. In germline mutation studies of wild populations the number of generations involved (and hence the total number of mutations) is typically small, or else the mutator allele becomes unlinked from the mutations it has caused (due to recombination), or is lost from the population altogether (due to chance or perhaps selection against its deleterious consequences).

The reviewer is correct that as it currently exists, IHD is mostly limited to applications in recombinant inbred lines (RILs) like the BXDs. This is due to the fact that IHD assumes that each diploid sample harbors one of two possible genotypes at a particular locus and ignores the possibility of heterozygous genotypes for simplicity. In natural, outbreeding populations, this assumption will obviously not hold. However, as we plan to further iterate on and improve the IHD method, we hope that it will be applicable to a wider variety of experimental systems in the future. We have added additional caveats about the applicability of our method to other systems in the text at lines 545-550.

Private Comments

1. On p. 8, perhaps I've misunderstood but it's not clear in what way the SVs identified were relevant to the samples used in this dataset - were the founder strains assembled? Is there any chance that additional SVs were present, e.g. de novo early in the accumulation line?

Our description of this structural variation resource could have been clearer. The referenced SVs were identified in Ferraj et al. (2023) by generating high-quality long read assemblies of inbred laboratory mice. Both DBA/2J and C57BL/6J (the founder strains for the BXD resource) were included in the Ferraj et al. SV callset. We have clarified our description of the callset at lines 247-248.

It is certainly possible that individual BXD lines have accumulated de novo structural variants during inbreeding. However, these "private" SVs are unlikely to produce a strong IHD association signal (via linkage to one of the ~7,000 markers) at either the chromosome 4 or chromosome 6 locus, since we only tested markers that were at approximately 50% D allele frequency among the BXDs.

1. On p. 13, comparing the IHD and QTL approaches, regarding the advantage of the former in that it detects the combined effect of multiple k-mer mutation types, would it not be straightforward to aggregate counts for different types in a QTL setting as well?

The mutation spectrum is a multi-dimensional phenotype (6-dimensional if using the 1-mer spectrum, 96-dimensional if using the 3-mer spectrum, etc.). Most QTL mapping methods use linear models to test for associations between genotypes and a 1-dimensional phenotype (e.g., body weight, litter size). In the past, we used QTL mapping to test for associations between genotypes and a single element of the mutation spectrum (e.g., the rate of C>A mutations), but there isn't a straightforward way to aggregate or collapse the mutation spectrum into a

1dimensional phenotype that retains the information contained within the full 1-mer or 3-mer spectrum. For that reason, we developed the "aggregate mutation spectrum" approach, as it preserves information about the complete mutation spectrum in each group of strains.

The reviewer is correct that we could also aggregate counts of different mutation types to, say, perform a QTL scan for the load of a specific mutational signature. For example, we could first perform standard mutational signature analysis on our dataset and then test for QTLs associated with each signature that is discovered. However, this approach would not solve the second problem that our method is designed to solve: the appropriate weighting of samples based on how many mutations they contain.

1. pp. 15-16: In the discussion of how you account for relatedness between strains, I found the second explanation (on p. 16) much clearer. It would be interesting to know how much variance was typically accounted for by this regression?

As shown in the response to Reviewer 1, genotype similarity between genotype groups (i.e., those with either D or B genotypes at a marker) generally explains a small amount of variance in the cosine distance between those groups ($R^2 \approx 0.007$). However, since the slope term in that regression is significantly non-zero, correcting for this relationship should still improve our power relative to using raw cosine distance values that are slightly confounded by this relationship.

1. Similarly, in the section on Applying the IHD method to the BXDs (pp. 18-19), I think this description was very useful, and some or all of this description of the experiment (and how the DNMs in it arise) could profitably be moved to the introduction.

We appreciate the reviewer's feedback about the details of the BXD cohort. Overall, we feel the description of the BXDs in the Introduction (at lines 65-73) is sufficient to introduce the cohort, though we now add some additional detail about variability in BXD inbreeding duration (at lines 89-93) to the Introduction as well, since it is quite relevant to some of the new simulation results presented in the manuscript.

1. A really minor one, not sure if this is for the journal or the authors, but it would be much better to include both page and line numbers in any version of an article for review. My pdf had neither!

We apologize for the lack of page/line numbers in the submitted PDF. We have now added line numbers to the revised version of the manuscript.

Reviewer 3 (Public Review):

1. Under simulated scenarios, the authors' new IHD method is not appreciably more powerful than conventional QTL mapping methods. While this does not diminish the rigor or novelty of the authors findings, it does temper enthusiasm for the IHD method's potential to uncover new mutators in other populations or datasets. Further, adaptation of this methodology to other datasets, including human trios or multigenerational families, will require some modification, which could present a barrier to broader community uptake. Notably, BXD mice are (mostly) inbred, justifying the authors consideration of just two genotype states at each locus, but this decision prevents out-of-the-box application to outbred populations and human genomic datasets. Lastly, some details of the IHD method are not clearly spelled out in the paper. In particular, it is unclear whether differences in BXD strain relatedness due to the breeding epoch structure are fully accounted for in permutations. The method's name - inter-haplotype distance - is also somewhat misleading, as it seems to imply that *de novo* mutations are aggregated at the scale of sub-chromosomal haplotype blocks, rather than across the whole genome.

The reviewer raises very fair concerns. As mentioned in response to a question from Reviewer 1, we performed additional simulation experiments that demonstrate the improved power of IHD (as compared to QTL mapping) in situations where mutation counts are variable across haplotypes or when mutator alleles are present at allele frequencies <50% (see Author response image 2 and 3, as well as new supplements to Figure 1 in the manuscript). However, the reviewer is correct that the IHD method is not applicable to collections of outbred individuals (that is, individuals with both heterozygous and homozygous genotypes), which will limit its current applications to datasets other than recombinant inbred lines. We have added a mention of these limitations to the Results at lines 138-141 and the Discussion at lines 545-550, but plan to iterate on the IHD method and introduce new features that enable its application to other datasets. We have also explicitly stated that we account for breeding epochs in our permutation tests in the Materials and Methods at lines 670-671. Both Reviewer 1 and Reviewer 3 raised concerns about the name of our method, and we have therefore changed “inter-haplotype distance” to “aggregate mutation spectrum distance” throughout the manuscript.

1. *Nominating candidates within the chr6 mutator locus requires an approach for defining a credible interval and excluding/including specific genes within that interval as candidates. Sasani et al. delimit their focal window to 5Mb on either side of the SNP with the most extreme P-value in their IHD scan. This strategy suffers from several weaknesses. First, no justification for using 10 Mb window, as opposed to, e.g., a 5 Mb window or a window size delimited by a specific threshold of P-value drop, is given, rendering the approach rather ad hoc. Second, within their focal 10Mb window, the authors prioritize genes with annotated functions in DNA repair that harbor protein coding variants between the B6 and D2 founder strains. While the logic for focusing on known DNA repair genes is sensible, this locus also houses an appreciable number of genes that are not functionally annotated, but could, conceivably, perform relevant biological roles. These genes should not be excluded outright, especially if they are expressed in the germline. Further, the vast majority of functional SNPs are non-coding, (including the likely causal variant at the chr4 mutator previously identified in the BXD population). Thus, the author's decision to focus most heavily on coding variants is not well-justified. Sasani et al. dedicate considerable speculation in the manuscript to the likely identity of the causal variant, ultimately favoring the conclusion that the causal variant is a predicted deleterious missense variant in Mbd4. However, using a 5Mb window centered on the peak IHD scan SNP, rather than a 10Mb window, Mbd4 would be excluded. Further, SNP functional prediction accuracy is modest [e.g., PMID 28511696], and exclusion of the missense variant in Ogg1 due its benign prediction is potentially premature, especially given the wealth of functional data implicating Ogg1 in C>A mutations in house mice. Finally, the DNA repair gene closest to the peak IHD SNP is Rad18, which the authors largely exclude as a candidate.*

We agree that the use of a 10 Mb window, rather than an empirically derived confidence interval, is a bit arbitrary and ad hoc. To address this concern, we have implemented a bootstrap resampling approach (Visscher et al. 1996, Genetics) to define confidence intervals surrounding IHD peaks. We have added a description of the approach to the Materials and Methods at lines 609-622, but a brief description follows. In each of N trials (here, N = 10,000), we take a bootstrap sample of the BXD phenotype and genotype data with replacement. We then perform an IHD scan on the chromosome of interest using the bootstrap sample and record the position of the marker with the largest cosine distance value (i.e., the "peak" marker). After N trials, we calculate the 90% confidence interval of bootstrapped peak marker locations; in other words, we identify the locations of two genotyped markers, between which 90% of all bootstrap trials produced an IHD peak. We note that bootstrap confidence intervals can exhibit poor "coverage" (a measure of how often the confidence intervals include the "true" QTL location) in QTL mapping studies (see Manichaikul et al. 2006, Genetics), but feel that the bootstrap is more reasonable than simply defining an ad hoc interval around an IHD peak.

The new 90% confidence interval surrounding the IHD peak on chromosome 6 is larger than the original (ad hoc) 10 Mbp window, now extending from around 95 Mbp to 114 Mbp. Notably, the new empirical confidence interval excludes Mbd4. We have accordingly updated our Results and Discussion sections to acknowledge the fact that Mbd4 no longer resides within the confidence interval surrounding the IHD peak on chromosome 6 and have added additional descriptions of genes that are now implicated by the 90% confidence interval. Given the uncertainties associated with using bootstrap confidence intervals, we have retained a brief discussion of the evidence supporting Mbd4 in the Discussion but focus primarily on Ogg1 as the most plausible candidate.

The reviewer raises a valid concern about our treatment of non-DNA repair genes within the interval surrounding the peak on chromosome 6. We have added more careful language to

the text at lines 219-223 to acknowledge the fact that non-annotated genes in the confidence interval surrounding the chromosome 6 peak may play a role in the epistatic interaction we observed.

The reviewer also raises a reasonable concern about our discussions of both Mbd4 and Ogg1 as candidate genes in the Discussion. Since Mbd4 does not reside within the new empirical bootstrap confidence interval on chromosome 6 and given the strong prior evidence that Ogg1 is involved in C>A mutator phenotypes (and is in the same gene network as Mutyh), we have reframed the Discussion to focus on Ogg1 as the most plausible candidate gene (see lines 357360).

Using the GeneNetwork resource, we also more carefully explored the potential effects of noncoding variants on the C>A mutator phenotype we observed on chromosome 6. We have updated the Results at lines 240-246 and the Discussion at line 439-447 to provide more evidence for regulatory variants that may contribute to the C>A mutator phenotype. Specifically, we discovered a number of strong-effect cis-eQTLs for Ogg1 in a number of tissues, at which D genotypes are associated with decreased Ogg1 expression. Given new evidence that the original mutator locus we discovered on chromosome 4 harbors an intronic mobile element insertion that significantly affects Mutyh expression (see Ferraj et al. 2023, Cell Genomics), it is certainly possible that the mutator phenotype associated with genotypes on chromosome 6 may also be mediated by regulatory, rather than coding, variation.

1. Additionally, some claims in the paper are not well-supported by the author's data. For example, in the Discussion, the authors assert that "multiple mutator alleles have spontaneously arisen during the evolutionary history of inbred laboratory mice" and that "... mutational pressure can cause mutation rates to rise in just a few generations of relaxed selection in captivity". However, these statements are undercut by data in this paper and the authors' prior publication demonstrating that a number of candidate variants are segregating in natural mouse populations. These variants almost certainly did not emerge de novo in laboratory colonies, but were inherited from their wild mouse ancestors. Further, the wild mouse population genomic dataset used by the authors falls far short of comprehensively sampling wild mouse diversity; variants in laboratory populations could derive from unsampled wild populations.

The reviewer raises a good point. In our previous publication (Sasani et al. 2022, Nature), we hypothesized that Mutyh mutator alleles had arisen in wild, outbreeding populations of *Mus musculus*, and later became fixed in inbred strains like DBA/2J and C57BL/6J. However, in the current manuscript, we included a statement about mutator alleles "spontaneously arising during the evolutionary history of inbred laboratory mice" to reflect new evidence (from Ferraj et al. 2023, Cell Genomics) that the mutator allele we originally identified in Mutyh may not be wild derived after all. Instead, Ferraj et al. suggest that the C>A mutator phenotype we originally identified is caused by an intronic mobile element insertion (MEI) that is present in DBA/2J and a handful of other inbred laboratory strains. Although this MEI may have originally occurred in a wild population of mice, we wanted to acknowledge the possibility that both the original Mutyh mutator allele, as well as the new mutator allele(s) we discovered in this manuscript, could have arisen during the production and inbreeding of inbred laboratory lines. We have also added language to the Discussion at lines 325-327 to acknowledge that the 67 wild mice we analyzed do not comprise a comprehensive picture of the genetic diversity present in wild-derived samples.

We have added additional language to the Discussion at lines 349-357 in which we acknowledge that the chromosome 6 mutator allele might have originated in either laboratory or wild mice and elaborate on the possibility that mutator alleles with deleterious fitness consequences may be more likely to persist in inbred laboratory colonies.

1. Finally, the implications of a discovering a mutator whose expression is potentially conditional on the genotype at a second locus are not raised in the Discussion. While not a weakness per se, this omission is perceived to be a missed opportunity to emphasize what, to this reviewer, is one of the most exciting impacts of this work. The potential background dependence of mutator expression could partially shelter it from the action of selection, allowing the allele persist in populations. This finding bears on theoretical models of mutation rate evolution and may have important implications for efforts to map additional mutator loci. It seems unfortunate to not elevate these points.

We agree and have added additional discussion of the possibility that the C>A mutator phenotypes in the BXDs are a result of interactions between the expression of two DNA repair genes in the same base-excision network to the Discussion section at lines 447-449.

Private comments

1. The criteria used to determine or specify haplotype size are not specified in the manuscript. I mention this above but reiterate here as this was a big point of confusion for me when reading the paper. Haplotype length is important consideration for overall power and for proper extension of this method to other systems/populations.

We may not have been clear enough in our description of our method, and as suggested by Reviewer 1, the name "inter-haplotype distance" may also have been a source of confusion. At a given marker, we compute the aggregate mutation spectrum in BXDs with either B or D genotypes using all genome-wide de novo mutations observed in those BXDs. Since the BXDs were inbred for many generations, we expect that almost all de novo germline mutations observed in an RIL are in near-perfect linkage with the informative genotypes used for distance scans. Thus, the "haplotypes" used in the inter-haplotype distance scans are essentially the lengths of entire genomes.

1. Results, first paragraph, final sentence. I found the language here confusing. I don't understand how one can compute the cosine distance at single markers, as stated. I'm assuming cosine distance is computed from variants residing on haplotypes delimited by some defined window surrounding the focal marker?

As discussed above, we aggregate all genome-wide de novo mutations in each group of BXDs at a given marker, rather than only considering DNMs within a particular window surrounding the marker. The approach is discussed in greater detail in the caption of Figure 1.

1. Nominating candidates for the chr6 locus, Table 1. It would be worth confirming that the three prioritized candidates (Setmar, Ogg1, and Mbd4) all show germline expression.

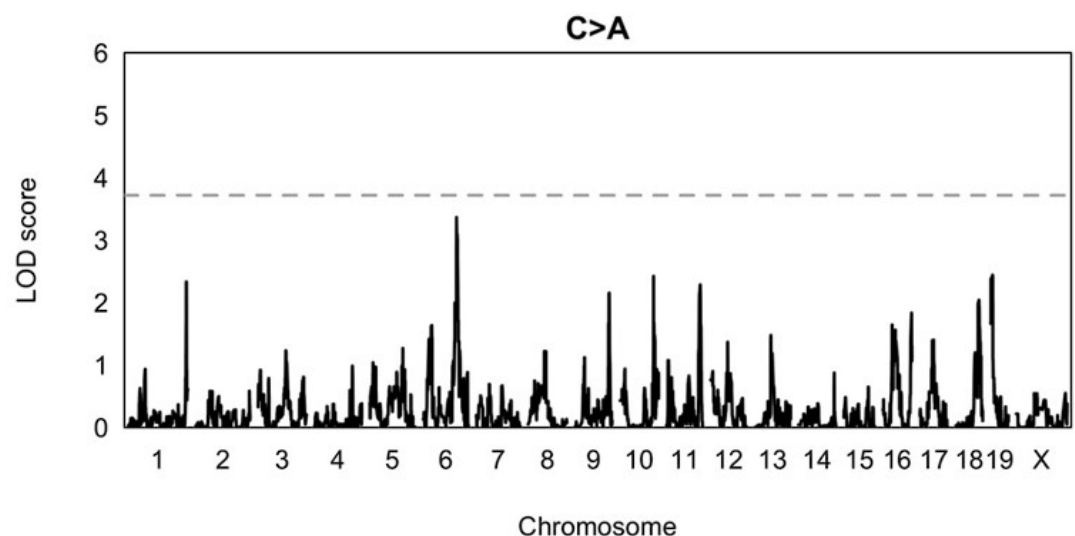
Using the Mouse Genome Informatics online resource, we confirmed that all prioritized candidate genes (now including Setmar and Ogg1, but not Mbd4) are expressed in the male and female gonads, and mention this in the Results at lines 228 and 233-234.

1. Does the chr6 peak on the C>A LOD plot (Figure 2- figure supplement 1) overlap the same peak identified in the IHD scan? And, does this peak rise to significance when using $\alpha = 0.05$? Given that the goal of these QTL scans is to identify loci that interact with the C>A mutator on chr4, it is reasonable to hypothesize that the mutation impact of epistatic loci will also be restricted to C>A mutations. Therefore, I am not fully convinced that the conservative $\alpha = 0.05/7$ threshold is necessary.

The chromosome 6 peak in Figure 2-figure supplement 1 does, in fact, overlap the peak marker we identified on chromosome 6 using IHD. One reason we decided to use a more conservative α of $(0.05 / 7)$ is that we wanted these results to be analogous to the ones we performed in a previous paper (Sasani et al. 2022, Nature), in which we first identified the mutator locus on chromosome 4. However, the C>A peak does not rise to genome-wide significance if we use a less conservative α value of 0.05 (see Author response image 5). As discussed in our response to Reviewer 1, we find that QTL mapping is not as powerful as IHD when haplotypes have accumulated variable numbers of germline mutations (as in the BXDs), which likely explains the fact that the peak on chromosome 6 is not genome-wide significant using QTL mapping.

Author response image 5.

QTL scan for the fraction of C>A mutations in BXDs harboring D alleles at the locus near Myth QTL scan was performed at a genome-wide significance α of 0.05, rather than 0.05/7.



1. Is there significant LD between the IHD peaks on chr6 and chr4 across the BXD? If so, it could suggest that the signal is driven by cryptic population structure that is not fully accounted for in the author's regression based approach. If not, this point may merit an explicit mention in the text as an additional validation for the authenticity of the chr6 mutator finding.

This is a good question. We used the scikit-allel Python package to calculate linkage disequilibrium (LD) between all pairs of genotyped markers in the BXD cohort, and found that the two peak loci (on chromosomes 4 and 6) exhibit weak LD ($r^2 = 4e-5$). We have added a mention of this to the main text of the Results at lines 212-213. That being said, we do not think the chromosome 6 mutator association (or the apparent epistasis between the alleles on chromosomes 4 and 6) could be driven by cryptic population structure. Unlike in human

GWAS and other association studies in natural populations, there is no heterogeneity in the environmental exposures experienced by different BXD subpopulations. In humans, population structure can create spurious associations (e.g., between height and variants that are in LD and are most common in Northern Europe), but this requires the existence of a phenotypic gradient caused by genetic or environmental heterogeneity that is not likely to exist in the context of inbred laboratory mice that are all the progeny of the same two founder strains.

1. Discussion, last sentence of the "Possible causal alleles..." section: I don't understand how the absence of the Mariner-family domain leads the authors to this conclusion. Setmar is involved in NHEJ, which to my knowledge is not a repair process that is expected to have a specific C>A mutation bias. I think this is grounds enough for ruling out its potential contributions, in favor of focusing on other candidates, (e.g., Mbd4 and Ogg1).

The reviewer raises a good point. Our main reason for mentioning the absence of the Marinerfamily domain is that even if NHEJ were responsible for the C>A mutator phenotype, it likely wouldn't be possible for Setmar to participate in NHEJ without the domain. However, the reviewer is correct that NHEJ is not expected to cause a C>A mutation bias, and we have added a mention of this to the text as well at lines 379-382.

1. Discussion, second to last paragraph of section "Mbd4 may buffer...": The authors speculate that reduced activity of Mbd4 could modulate rates of apoptosis in response to DNA damage. This leads to the prediction that mice with mutator alleles at both Mutyh and Mbd4 should exhibit higher overall mutation rates compared to mice with other genotypes. This possibility could be tested with the authors' data.

The reviewer raises a good question. As mentioned above, however, we implemented a new approach to calculate confidence intervals surrounding distance peaks and found that this empirical approach (rather than the ad hoc 10-Mbp window approach we used previously) excluded Mbd4 from the credible interval. Although we still mention Mbd4 as a possible candidate (since it still resides within the 10 Mbp window), we have refactored the Discussion section to focus primarily on the evidence for Ogg1 as a candidate gene on chromosome 6.

In any case, we do not observe that mice with mutator alleles at both the chromosome 4 and chromosome 6 loci have higher overall mutation rates compared to mice with other genotype combinations. This may not be terribly surprising, however, since C>A mutations only comprise about 10% of all possible mutations. Thus, given the variance in other 1-mer mutation counts, even a substantial increase in the C>A mutation rate might not have a detectable effect on the overall mutation rate. Indeed, in our original paper describing the Mutyh mutator allele (Sasani et al. 2022, Nature), we did not identify any QTL for the overall mutation rate in the BXDs and found that mice with the chromosome 4 mutator allele only exhibited a 1.11X increase in their overall mutation rates relative to mice without the mutator allele.

1. Methods, "Accounting for BXD population structure": An "epoch-aware" permutation strategy is described here, but it is not clear when (and whether) this strategy is used to determine significance of IHD P-values.

We have added a more explicit mention of this to the Methods section at lines 670-671, as we do, in fact, use the epoch-aware permutation strategy when calculating empirical distance thresholds.

1. The simulation scheme employed for power calculations is highly specific to the BXD population. This is not a weakness, and perfectly appropriate to the study population used here. However, it does limit the transferability of the power analyses presented in this manuscript to other populations. This limitation may merit an explicit cautionary mention to readers who may aspire to port the IHD method over to their study system.

This is true. Our simulation strategy is relatively simple and makes a number of assumptions about the simulated population of haplotypes (allele frequencies normally distributed around 0.5, expected rates of each mutation type, etc.). In response to concerns from Reviewer 1, we performed an updated series of simulations in which we varied some of these parameters (mutator allele frequencies, mean numbers of mutations on haplotypes, etc.). However, we have added a mention of the simulation approach's limitations and specificity to the BXDs to the text at lines 545-550.