The Jackson Laboratory
The Mouseion at the JAXlibrary

5-2019

Significant Strain Variation in the Mutation Spectra of Inbred Laboratory Mice.

Beth L Dumont

The Jackson Laboratory, beth.dumont@jax.org

Follow this and additional works at: https://mouseion.jax.org/stfb2019

Part of the Life Sciences Commons, and the Medicine and Health Sciences Commons

Recommended Citation

Significant Strain Variation in the Mutation Spectra of Inbred Laboratory Mice

Beth L. Dumont*,[1]

[1]The Jackson Laboratory, Bar Harbor, ME

*Corresponding author: E-mail: beth.dumont@jax.org.
Associate editor: Anne Yoder

Abstract

Mutation provides the ultimate source of all new alleles in populations, including variants that cause disease and fuel adaptation. Recent whole genome sequencing studies have uncovered variation in the mutation rate among individuals and differences in the relative frequency of specific nucleotide changes (the mutation spectrum) between populations. Although parental age is a major driver of differences in overall mutation rate among individuals, the causes of variation in the mutation spectrum remain less well understood. Here, I use high-quality whole genome sequences from 29 inbred laboratory mouse strains to explore the root causes of strain variation in the mutation spectrum. My analysis leverages the unique, mosaic patterns of genetic relatedness among inbred mouse strains to identify strain private variants residing on haplotypes shared between multiple strains due to their recent descent from a common ancestor. I show that these strain-private alleles are strongly enriched for recent de novo mutations and lack signals of widespread purifying selection, suggesting their faithful recapitulation of the spontaneous mutation landscape in single strains. The spectrum of strain-private variants varies significantly among inbred mouse strains reared under standardized laboratory conditions. This variation is not solely explained by strain differences in age at reproduction, raising the possibility that segregating genetic differences affect the constellation of new mutations that arise in a given strain. Collectively, these findings imply the action of remarkably precise nucleotide-specific genetic mechanisms for tuning the de novo mutation landscape in mammals and underscore the genetic complexity of mutation rate control.

Key words: house mice, germline mutation, mutation spectrum.

Introduction

The de novo mutation rate determines the frequency at which new alleles arise in populations, with the potential for such variants to drive adaptive evolution or cause disease. Knowledge of this fundamental quantity is critical for interpreting levels of neutral diversity in populations (Kimura 1983), dating historical events from genetic data (Scally and Durbin 2012; Moorjani et al. 2016), and forecasting the ultimate fate of a species (Lynch 2016). De novo mutation rates also determine the incidence of many rare Mendelian diseases (Kondrashov 2003) and influence genetic risk for multiple common diseases (Vissers et al. 2010; Girard et al. 2011; O'Roak et al. 2011, 2012; Xu et al. 2011; Iossifov et al. 2012; Neale et al. 2012; Sanders et al. 2012; Bomba et al. 2017).

Despite its critical importance to human health and evolution, the rate of mutation exhibits considerable variability. Mutation rates fluctuate across genomes (Hodgkinson and Eyre-Walker 2011; Ségurel et al. 2014), conditional on aspects of the local chromatin environment (Schuster-Böckler and Lehner 2012; Chen et al. 2017; Carlson et al. 2018), recombination rate (Lercher and Hurst 2002; Hardison et al. 2003; Besenbacher et al. 2016), GC content (Hardison et al. 2003; Schaibley et al. 2013), DNA replication timing (Stamatoyannopoulos et al. 2009; Francioli et al. 2015), transcription (McVicker and Green 2010; Park et al. 2012; Chen et al. 2017), and flanking nucleotide context (Duncan and Miller 1980; Hwang and Green 2004; Carlson et al. 2018). The mutation rate varies by orders of magnitude between species (Uchimura et al. 2015; Smeds et al. 2016; Senra et al. 2018), presumably reflecting species differences in DNA repair mechanisms (Hart and Setlow 1974; MacRae et al. 2015), metabolism (Martin and Palumbi 1993), and life history (Nabholz et al. 2008). There are striking mutation rate differences among individuals, including a marked dimorphism between males and females and a pronounced age effect (Crow 2000; Conrad et al. 2011; Kong et al. 2012; O'Roak et al. 2012; Francioli et al. 2015; Besenbacher et al. 2016; Jonsson et al. 2017). Population genomic analyses even point to differences in the mutation spectrum—the relative fraction of de novo mutations that result in particular types of nucleotide changes—among individuals (Harris 2015; Assaf et al. 2017; Harris and Pritchard 2017; Mathieson and Reich 2017). Thus, a subset of individuals in a population disproportionately contributes to the pool of new mutations that arise each generation, and different individuals are more prone to transmitting particular types of single nucleotide mutations to their offspring.

While differences in parental age at reproduction may account for much of the observed variation in human germline mutation rates (Kong et al. 2012; Jonsson et al. 2017),

environmental and genetic factors likely also play a role. Exposure to ionizing radiation (Dubrova, Bersimbaev, et al. 2002; Dubrova, Grant, et al. 2002), cigarette smoke (Zenzes 2000; Shi et al. 2001; DeMarini 2004), caffeine (Robbins, Vine, et al. 1997), and chemotherapeutic agents (Robbins, Meistrich, et al. 1997; Frias et al. 2003) have been previously associated with increased germline mutation loads in humans. Differences in exposures among individuals could contribute to observed mutation rate variation. In addition, there are >1,000 genes in the mammalian genome with annotated functions in DNA damage surveillance, DNA repair, and the metabolism of genotoxic compounds (Carbon et al. 2009). These loci represent potential reservoirs of functional genetic variation modifying the fidelity and efficiency of DNA damage repair and tuning genomic sensitivity to mutagens (Baer et al. 2007). Consistent with this possibility, mutations in DNA repair genes have been associated with elevated *somatic* mutation rates (Ahn et al. 2016), increased cancer risk (Easton et al. 2007; Dowty et al. 2013), and premature aging (de Boer et al. 2002; Lombard et al. 2005). Despite the considerable functional overlap of the DNA proofreading and repair machinery between the soma and the germline, there have been few efforts to directly link variation at putative modifier loci to variation in germline mutation rates in mammals (Uchimura et al. 2015; Seoighe and Scally 2017).

Comparative genomic analyses of cancer tumors and control tissues have uncovered remarkably precise, sequence-dependent mechanisms of mutation. Defects in specific DNA repair genes and pathways can result in distinct somatic mutation signatures, defined by the relative enrichment of mutation events in specific nucleotide contexts (Nik-Zainal et al. 2012; Alexandrov et al. 2013; Helleday et al. 2014). For example, altered activity of the error prone polymerase Polε is associated with 5′-TCT-3′>5′-TAT-3′ and 5′-TTT-3′>5′-TGT-3′ mutations in human cancers (Cancer Genome Atlas Network 2012; Alexandrov et al. 2013; Cancer Genome Atlas Research Network 2013; Shinbrot et al. 2014). Cancers with altered AID/APOBEC mutational activity are characterized by a preponderance of 5′-TCA-3′>5′-TTA-3′ mutations (Nik-Zainal et al. 2012; Alexandrov et al. 2013). Such observations raise the possibility that segregating polymorphisms in DNA repair genes could also influence *germline* mutation rates with extraordinary sequence context precision. Indeed, there are significant differences in the human germline mutation spectrum inferred from population-private (Harris 2015), rare (Mathieson and Reich 2017), and derived alleles (Harris and Pritchard 2017). However, the challenge of disentangling variable parental age, differential environmental exposures, and genetic differences between human populations makes it difficult to address the underlying causes of observed variation in the mutation spectrum in our own species.

The availability of high-quality genomes from multiple inbred laboratory mouse strains provides a powerful opportunity to overcome this limitation (Keane et al. 2011; Adams et al. 2015). As a consequence of their unique historical origins from a small founder population (Wade and Daly 2005; Yang et al. 2007, 2011), variants that are private to single inbred

mouse strains but that reside on haplotypes that are otherwise shared between strains have likely arisen de novo since their inception as laboratory models. I leverage this recognition to derive the germline mutation spectrum in 29 common inbred mouse strains. I document significant differences in the mutation spectrum between mouse strains reared in standard laboratory environments and show that this variation is not solely accounted for by differences in strain age at breeding. My findings suggest that multiple modifiers of the mutation spectrum segregate among inbred mouse strains, implying that the process of germline mutation is itself a complex genetic trait.

Results and Discussion

Identification of Recent, Spontaneous Mutations in Laboratory Mice

The classical laboratory mouse strains derive from a small, ancestral population of wild founder animals that were selectively bred for traits of interest by mouse fanciers in the late 19th and early 20th centuries. As a consequence of their unique historical origins, the genetic diversity captured in laboratory mice represents an extremely limited sample of the diversity found in wild mouse populations (Salcedo et al. 2007; Yang et al. 2011; Phifer-Rixey and Nachman 2015). Notably, variation among inbred strains at over 97% of the genome can be reconciled into fewer than ten distinct haplotypes (Yang et al. 2011). Thus, the genomes of the laboratory strains can be envisioned as mosaics derived from this small founder pool, with the haplotype of a given genomic region in one strain likely shared between multiple strains (fig. 1).

Although the majority of genetic variants segregating among the inbred laboratory strains trace to variation in this "Fancy Mouse" founder population, a small remaining number of sites can be attributed to recent de novo mutations that occurred following their inception as laboratory models. Many inbred strains carry known spontaneous mutations that drifted to fixation in inbreeding colonies. For example, a de novo mutation on the C57BL/6N background generated a missense mutation in *cytoplasmic FMRP interacting protein 2* (*Cflip2*) that confers a sensitized response to cocaine (Kumar et al. 2013). Similarly, a spontaneous mutation in *Tlr4* in C3H/HeJ mice renders this strain uniquely resistant to endotoxin (Qureshi et al. 1999).

One strategy for systematically identifying these recent, spontaneous mutations is to identify alleles that are private to a single laboratory mouse strain, but that reside on a haplotype that is otherwise identical by descent (IBD) in multiple strains (fig. 1). A conceptually similar approach has been previously used to identify de novo mutations in large, multigenerational human pedigrees (Campbell et al. 2012; Palamara et al. 2015; Narasimhan et al. 2017). To this end, I used publicly available high-quality whole-genome sequences from 29 inbred mouse strains to identify autosomal IBD regions ≥5 Mb shared between at least 2 strains (Keane et al. 2011; Adams et al. 2015; see Materials and Methods). On an average, 46.8% of the genome of a given inbred strain is
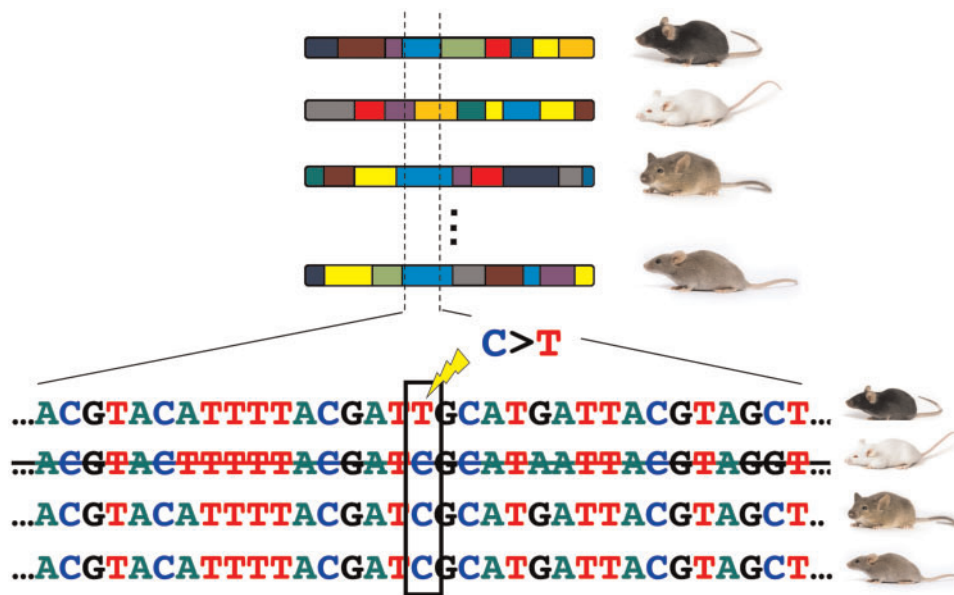
Fig. 1. Schematic of the approach used to identify putative de novo mutations in inbred laboratory mouse strains. The genomes of laboratory mice can be envisaged as mosaics of less than ten haplotypes derived from a small population of founder animals. The ancestral haplotype structure of four strains is illustrated, with unique haplotypes depicted in different colors. Given the small number of founders, multiple strains are likely to share the same haplotype at a given locus by virtue of their descent from a common ancestor. An example of one such region shared by three of the four depicted strains (light blue haplotype) is outlined by dashed lines. Recent mutations that arose in a single focal strain can be detected as strain-private variants resident on haplotypes that are shared identical-by-descent between strains.

present in IBD blocks of this size (range: 6.7% KK/HiJ to 82.7% C57BL/6NJ; supplementary table S3, Supplementary Material online). In contrast, no IBD blocks ≥5 Mb are observed in a sample of 27 wild *Mus musculus domesticus* genomes (supplementary table S2 and fig. S1, Supplementary Material online; Harr et al. 2016). These findings reinforce the close genetic relatedness of the common laboratory strains and confirm their origins from a very small number of founder individuals (Beck et al. 2000; Frazer et al. 2007; Yang et al. 2011).

I next identified 15,311 strain private variants (SPVs) on large (≥5 Mb) IBD blocks (fig. 1; see Materials and Methods). Approximately 91.7% of these SPVs are fixed for the alternative allele, with the remaining 8.3% present in a heterozygous state. On an average, there are 528 SPVs per strain (Range: 103 in KK/HlJ to 1,028 C57BR/cdJ; supplementary tables S3 and S4, Supplementary Material online), with the number of SPVs strongly correlated with the fraction of a given strain's genome captured in IBD regions (Spearman's Rho = 0.824; $P = 4.0 \times 10^{-8}$). Given the absence of IBD blocks ≥5 Mb in wild mouse populations (supplementary table S2 and fig. S1, Supplementary Material online), these SPVs are almost certainly due to recent mutations, as opposed to variants inherited from wild mouse ancestral populations.

Strain Private Variants Approximate Neutral Expectations

The maintenance of inbred strains via small "foundation stocks" of sister-brother mating pairs approximates the experimental design of a mutation accumulation experiment and minimizes the efficacy of natural selection at each

generation (Currer et al. 2009). Consequently, the vast majority of new mutations that arise during strain propagation are expected to be effectively neutral and their ultimate fate will be governed by chance (Eyre-Walker and Keightley 2007). Only a small subset of these neutral mutations will drift to fixation within an inbreeding colony, but those that do fix should be a representative, random sample of all neutral mutations. In contrast, a small number of new mutations are expected to be deleterious, and even fewer will confer an adaptive advantage. Although nonneutral mutations will be subject to selection or go unrealized due to viability defects or infertility, such large-effect variants should comprise a small fraction of all new mutations (Eyre-Walker and Keightley 2007; Uchimura et al. 2015). Based on these considerations, I reasoned that the set of SPVs for a given strain should approximate the strain-specific distribution of germline mutations.

To confirm the interpretation of observed SPVs as recent de novo mutations that have not been strongly biased by natural selection, I took advantage of expected differences in signals of historical selection at young versus old mutations. Mutations that have arisen recently in the laboratory strains have not been segregating for ample time to bear strong signatures of past selection. As a result, new mutations should be approximately randomly distributed within functional genomic regions. In contrast, old ancestral variants have been subject to generations of purifying selection and should be depleted for functional variants. I tested these dueling predictions using two approaches.

First, I determined the fraction of coding SPVs that result in synonymous and nonsynonymous changes. Assuming a

uniform probability of mutation at all amino acid encoding sites, 72.2% and 3.8% of spontaneous mutations should result in missense and nonsense changes, respectively (Assaf et al. 2017). In rough agreement with these null expectations, 67.2% and 3.4% of the coding SPVs identified on IBD haplotypes are missense and nonsense substitutions, respectively (table 1). This represents a significant enrichment for potentially functional protein-coding variation relative to common variants segregating in ≥ 2 mouse strains (G-test, $P = 8.22 \times 10^{-60}$; table 1). The observed deficit of nonsynonymous substitutions relative to the expected 72.2% may be explained by selection against mutations that confer lethality, infertility, or alter stereotyped strain phenotypes. Importantly, these bulk trends are broadly recapitulated on a per strain basis: the frequencies of missense, nonsense, and synonymous SPVs within coding regions do not significantly differ from null expectations for 27 of the 29 inbred strains (G-test of independence, $P > 0.05$; supplementary table S5 and fig. S2, Supplementary Material online). Although power to detect a significant departure from the null expectation is low given the small number of coding variants per strain, these findings provide no reason to suspect that the strength of selection against deleterious mutations differs markedly among strains.

Second, I compared the distribution of sequence conservation scores between common and strain-private variants. Sites that are well conserved across species are typically interpreted as targets of purifying selection to maintain a critical biological function (Siepel et al. 2005). Mutations at these sites are expected to be more deleterious, on an average, than those that arise in poorly conserved (and presumably nonfunctional) regions. As a result, evolutionarily conserved sites should be depleted for older, intermediate frequency variants. On the other hand, recent de novo mutations have not yet been strongly shaped by selection, and their genomic distribution should approximate the genome-wide distribution of sequence conservation scores. Consistent with their hypothesized origins from recent mutations, SPVs are enriched in conserved genomic regions compared with common variants (Kolmogorov–Smirnov test $P < 2.2 \times 10^{-16}$), and closely approximate the cumulative distribution of sequence conservation scores across the mouse reference genome (fig. 2). This overall finding is also preserved on a per-strain basis; for all strains, the distribution of sequence conservation scores at SPVs is skewed toward conserved sites relative to common variants (supplementary fig. S3, Supplementary Material online).

The striking enrichment of SPVs in functional coding regions and conserved sequences relative to common variants is consistent with their recent emergence in the laboratory strains. These findings suggest that few ancestral alleles are masquerading as SPVs in this data set and reveal the absence of pervasive, strong selection against new mutations in laboratory colonies. Taken together, these results indicate that the set of SPVs for a focal inbred laboratory strain approximates the cumulative action of diverse germline mutational processes active in that strain.

Table 1. Frequencies of Coding Variants as a Function of Variant Frequency.

	Common Variants[a] (%; 95% bootstrap CI[b])	Strain Private Variants (%; 95% bootstrap CI[b])
Coding variants	75,509 (NA)	439 (NA)
Synonymous variants	49,964 (66.2; 65.8–66.5)	129 (29.4; 25.1–33.7)
Missense variants	25,303 (33.5; 33.1–33.8)	295 (67.2; 62.4–71.4)
Nonsense variants	242 (0.32; 0.28–0.36)	15 (3.4; 1.9–5.3)

[a]Common variants are defined as variants present in two or more laboratory strains, including the C57BL/6J reference.
[b]95% confidence intervals are derived from 1,000 bootstrap resamples of observed variants.

Cumulative Distribution of PhastCons Scores
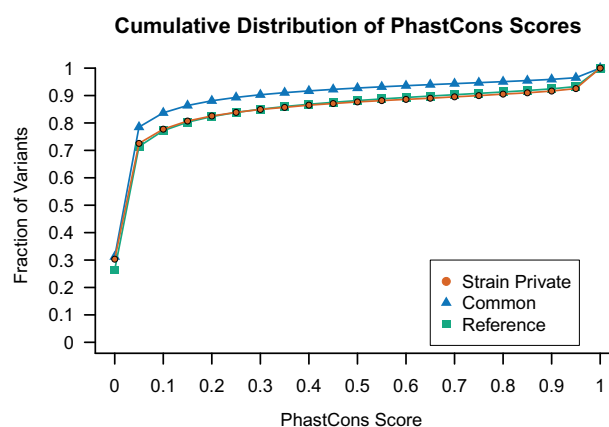
FIG. 2. The cumulative distribution of PhastCons conservation scores for strain private and common variants, as well as all sites on the C57BL/6J reference sequence. Common variants are defined as those with the alternative allele present in two or more classical laboratory strains. Only nonrepeat masked sites are considered. The 95% bootstrap confidence intervals associated with each point are narrower than the plotting characters and are therefore omitted from the figure.

The Spectrum of Strain-Specific Variants in Inbred House Mice

The most parsimonious interpretation for a SPV on an IBD haplotype is that it arose from a single mutational event during the focal strain's breeding history. For example, if all classical laboratory strains have a "G" allele at a particular site, with the exception of BALB/cJ which carries a "T" allele, it can be inferred that a G > T mutation occurred in a recent common ancestor of the BALB/cJ strain. By extending this logic genome-wide, I identified the set of single nucleotide mutations that putatively arose in each classical laboratory strain. Using the information available from other strains, I polarized each mutational event into likely ancestral and derived alleles and then quantified the number of mutations in each strain that are of each possible mutational class. To account for the sequence dependency of mutation rates, mutation counts were standardized by the nucleotide composition of all IBD regions in the focal strain and scaled to sum to one. Owing to ambiguity in the strand of origin of a particular mutation,
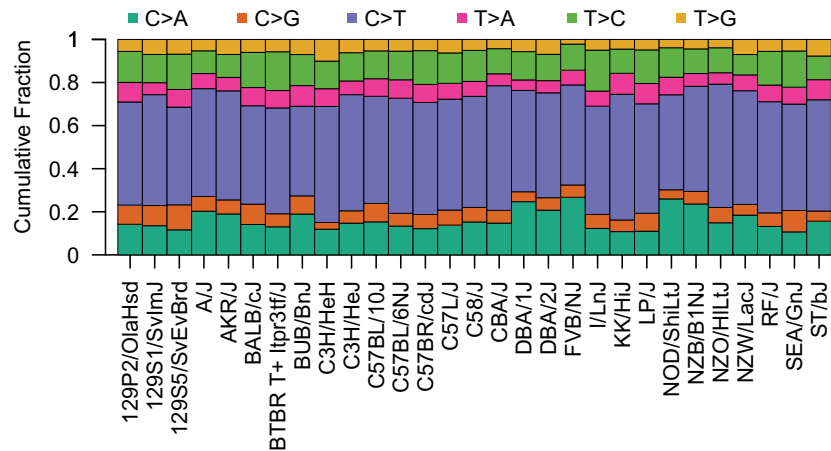
Fig. 3. The folded strain private variant spectrum for 29 common inbred laboratory mouse strains. The relative frequency of each mutation type is displayed as a stacked bar plot for each strain.

complementary mutations were binned to produce the folded SPV spectrum for each strain (fig. 3).

SPV spectra display qualitative similarities among mouse strains, with relative variant frequencies following the same rank order. For all strains, C > G, T > A, and T > G mutations are the rarest mutational classes, with each mutation type accounting for 2.1–11.7% of SPVs. C > T transitions are the most frequent type of SPVs, ranging from 41.8% to 58.6% of SPVs among strains (fig. 3 and supplementary table S3, Supplementary Material online). This C > T fraction is notably higher than the proportion of rare human SNPs (Harris and Pritchard 2017) and de novo mutations (Rahbari et al. 2016; Jonsson et al. 2017) that are C > T transitions (~40%), but is consistent with the elevated relative frequency of de novo C > T mutations observed in mouse pedigrees (Lindsay et al. 2018). Despite an overall excess of C > T SPVs, the laboratory mouse strains actually exhibit a marked deficit of CpG>TpG SPVs compared with both de novo mouse and human mutation spectra (supplementary fig. S4, Supplementary Material online). This discrepancy is likely attributable to differences in variant ascertainment between studies, underscoring the need for caution in comparisons of spectra derived from SPVs and de novo mutations. Despite this caveat, mouse SPVs recapitulate key differences between the de novo mutation spectra of mouse and human, including reduced T > C and increased T > A relative mutation frequencies in mouse (supplementary fig. S4, Supplementary Material online; Lindsay et al. 2018).

Although the ranked relative frequencies of different variant classes exhibit broad conservation among strains, 71.7% of the 406 possible strain pairs possess significantly distinct SPV spectra (G-test, uncorrected $P < 0.05$, d.f. = 5; supplementary table S6, Supplementary Material online). To further explore this variation, I performed a principle component analysis on the nucleotide-adjusted frequencies of each mutational type across the 29 classical inbred strains. Principle component (PC) 1 isolates FVB/NJ, NOD/ShiLtJ, DBA/1J, and NZB/BlNJ from all other strains (fig. 4). These strains are characterized by an elevated relative rate of C > A mutations and exceptionally low proportions of C > G and T > G mutations
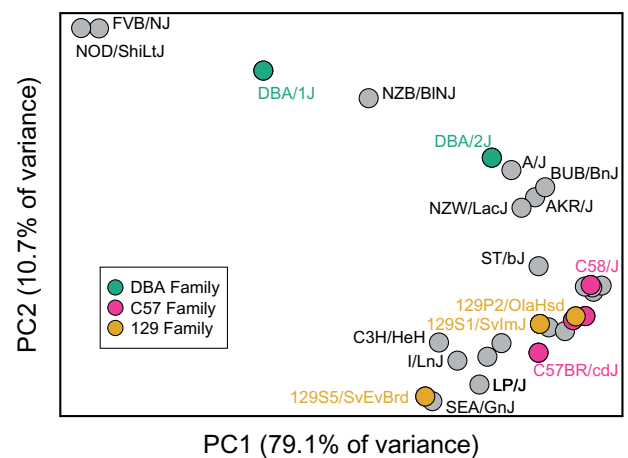
Fig. 4. Principle component analysis of strain variation in the SPV spectrum. Closely related strains in the 129, C57, and DBA families are color-coded.

(supplementary table S3, Supplementary Material online). Thus, the major axis of variance in these data is dominated by multidimensional properties of the mutation spectrum. Strains belonging to the C57 (C57BL/10J, C57BL/6NJ, C57BR/cdJ, C57L/J) and 129 (129P2/OlaHsd, 129S1/SvImJ, 129S5/SvEvBrd) strain families show a loose tendency to cluster (fig. 4), suggesting that more closely related strains have more similar mutation spectra.

Causes of Variation in the Mutation Spectrum among Inbred Mouse Strains

The observed strain differences in the mouse mutation spectrum could be driven by environmental differences, strain variation in parental age at reproduction, or genetic factors segregating among strains. Given that inbred strains are reared under standardized laboratory conditions, environmental contributions to strain variation in the mutation spectrum seem unlikely, although effects from minor differences in animal husbandry cannot be ruled out.

There are established age-related shifts in the human de novo mutation spectrum (Jonsson et al. 2017). Although a

Table 2. Spearman Correlations between Strain Reproductive Traits and the Mutation Spectrum.

	Dam Age at First Litter		Interbirth Interval	
	Spearman's Rho	P value	Spearman's Rho	P value
Principle component 1	0.059	0.831	0.035	0.900
% C>A	−0.026	0.926	−0.421	0.106
% C>G	0.126	0.641	0.647	0.008
% nonCpG>T	0.018	0.952	−0.244	0.361
%CpG>TpG	−0.009	0.978	0.047	0.865
% T>A	0.279	0.294	0.303	0.253
% T>C	0.153	0.571	0.362	0.169
% T>G	−0.203	0.450	0.050	0.857

comparable effect on the mouse mutation spectrum has yet to be shown, differences in reproductive aging between strains could contribute to variation in the mutation spectrum. The inbred strains profiled here significantly differ for two proxy measures of age at reproduction: dam age at first litter (supplementary fig. S5, Supplementary Material online; one-way ANOVA $F_{15,746} = 15.11; P < 2.2 \times 10^{-16}$) and average interbirth interval (supplementary fig. S6, Supplementary Material online; one-way ANOVA $F_{15,2281} = 16.29; P < 2.2 \times 10^{-16}$). However, variation in these life history traits does not fully account for observed differences in the mutation spectrum. PC1, which captures 79% of the variance in the mutation spectrum among inbred laboratory strains (fig. 4), is not correlated with the dam age at first litter or average interbirth interval (table 2). Similarly, with only one exception, strain variation in the proportion of SPVs belonging to each mutational category is not correlated with either life history trait (table 2). The exception is a positive correlation between the frequency of C > G mutations and interbirth interval (table 2). Many C > G mutations are hypothesized to arise from spontaneous double-strand break induced damage in the germline, the frequency of which may increase with longer generation times (Jonsson et al. 2017; Gao Z, Moorjani P, Amster G, Przeworski M, unpublished data; Agarwal I, Przeworski M, unpublished data).

Taken together, these considerations suggest that the spectrum of SPVs in a given strain is at least partially attributable to segregating genetic differences among strains. Strain variation in the fraction of SPVs within each mutational class is continuous (supplementary fig. S7, Supplementary Material online), suggesting that the genetic control of the germline mutation spectrum is potentially both polygenic and complex.

Modifiers of the Mouse Mutation Spectrum

My findings raise the possibility that distinct inbred mouse strains harbor unique suites of mutation modifying loci that collectively exert precise, nucleotide-dependent effects on the spectrum of accumulated de novo mutations. Determining the molecular identity of these mutation spectrum modifiers is an important outstanding research aim, albeit one that falls outside the scope of this paper. Genetic differences in genes involved in DNA repair, replication, genome surveillance, and the metabolism of genotoxic compounds pose strong a priori candidates, particularly given their established effects on mutational signatures extracted from human cancers (Nik-Zainal et al. 2012; Alexandrov et al. 2013). Among the 29 mouse strains examined here, there are 845 segregating SNPs that alter the amino acid sequence of genes with GO terms associated with the maintenance of genome integrity, including 8 SNPs with predicted strongly deleterious effects (supplementary table S7, Supplementary Material online). These latter variants present compelling targets for future investigations.

Although the genetic drivers of mutation spectrum heterogeneity remain unknown, many of the causal variants likely derive from the small, ancestral population of (mostly) *M. m. domesticus* mice that provided the genetic source pool for the laboratory inbred strains (Yang et al. 2007). Evolutionary theory predicts considerable scope for segregating mutation modifiers in natural populations (Lynch 2008, 2010; Sung et al. 2012). Although reduced organismal mutation rates are selectively favored in most scenarios (due to the negative fitness consequences of accumulated deleterious alleles), selection against weak mutation rate modifiers is ineffective in small and modestly sized populations, where the stochastic effects of genetic drift overwhelm the deterministic forces of natural selection (Lynch 2010, 2011; Sung et al. 2012). As a result, even moderate-strength modifiers of the mutation spectrum are potentially long-lived in natural mammalian populations and may rise to intermediate allele frequencies. Consistent with this prediction, a recent analysis of haplotype variation in humans reported the action of multiple historical mutation rate modifiers (Seoighe and Scally 2017). These considerations, combined with the large mutational target size for the accumulation of genetic variance for mutation, suggest the high likelihood of ancestrally derived mutation modifiers segregating in inbred mouse strains. Given that laboratory mice capture a limited subset of wild mouse diversity, the magnitude of mutation rate variation in wild populations is almost certainly far greater than that summarized here.

At the same time, mutation modifiers may have also emerged de novo in laboratory colonies. Consistent with this possibility, there are 219 nonsynonymous and premature stop variants in putative mutation modifier candidates that are private to single strains (supplementary table S7, Supplementary Material online).

Further, the efficacy of natural selection against even large-effect mutation rate modifiers in laboratory colonies is likely quite weak due to small effective population sizes and laboratory housing conditions that potentially minimize the negative fitness consequences of deleterious alleles. Thus, both ancestral and young alleles are likely to shape observed variation in the mouse mutation spectrum, but further investigation is required to determine their relative contributions.

Despite these theoretical arguments, there remains little direct evidence for segregating modifiers of the de novo mutation rate in mammalian populations (see Seoighe and Scally 2017 for a notable exception). Mutation rate variation among

sequenced human trios can be explained almost entirely by variation in parental age (Kong et al. 2012; Goldman et al. 2016; Jonsson et al. 2017). However, the absence of large-effect modifiers of the *mutation rate* in human populations does not preclude the possibility that loci that exert nuanced effects on the *mutation spectrum* are segregating in our species or present in mice (Harris and Pritchard 2017; Seoighe and Scally 2017).

Conclusions

Here, I harnessed the unique history of laboratory mice in conjunction with high-quality whole-genome sequences to define SPV spectra in 29 common inbred mouse strains. I documented significant strain variation in the relative probability of different mutational classes, including a strong mutation dependency on local nucleotide context. I showed that SPVs match neutral variant expectations and approximate multidimensional properties of spontaneous germline mutations in house mice. These considerations support the interpretation of SPVs as recent de novo germline mutations. I show that strain variation in age at reproduction cannot explain observed strain differences in the mutation spectrum, demonstrating that the constellation of new mutations that accumulate at a given generation is at least partially subject to genetic control in house mice.

The finding that genetic background likely influences the mutation *spectrum* raises the related question of whether segregating variation also contributes to differences in the overall de novo mutation *rate* among inbred mouse strains. If so, it is of considerable interest to define the genetic architecture of this cellular phenotype, including identifying germline mutation rate modifying genes. Toward this goal, it may be possible to harness genomic resources from The Collaborative Cross (Srivastava et al. 2017) and other recombinant inbred mouse populations to estimate the pace of mutation accumulation in different genetic backgrounds and map global mutation rate modifiers. Notably, the discovery of mutation rate modifiers in mice could steer the search for modifiers in human populations, where the confounding effects of variable mutagen exposure and parental age are likely to impede direct mapping efforts.

Materials and Methods

SNP Data and Annotation

Publicly available VCF files from the high-quality whole genome sequences of 29 inbred laboratory mouse strains were downloaded from the Sanger Mouse Genomes project FTP site (ftp://ftp-mouse.sanger.ac.uk/current_snps/; last accessed on May 31, 2017). All genomes were sequenced to >12× coverage, with all but three sequenced to >30× coverage (median coverage = 43.76; supplementary table S3, Supplementary Material online). Variants were identified relative to the GRCm38 reference assembly based on the C57BL/6J inbred mouse strain. Variants were subsequently annotated using snpEff (v4.3t; Cingolani et al. 2012) and intersected with the phastCons60wayPlacental, genomicSuperDups, and

RepeatMasker tracks obtained from the UCSC Table Browser (Kent et al. 2002).

The VCF file for 27 wild-caught *M. m. domesticus* mice was obtained from a public repository (http://wwwuser.gwdg.de/~evolbio/evolgen/wildmouse/; last accessed June 2, 2017 Harr et al. 2016). Variants were filtered to include only biallelic SNPs that pass all filters using VCFtools (Danecek et al. 2011).

De novo mutation data sets for cow (Harland C, Charlier C, Karim L, Cambisano N, Deckers M, Mni M, Mullaart E, Coppieters W, Georges M, unpublished data), human (Jonsson et al. 2017), and mouse (Srivastava et al. 2017; Lindsay et al. 2018) were obtained from the supplemental materials of the associated publications.

Identification of Strain-Private Substitutions

Shared haplotypes were identified for each pair of laboratory strains using GERMLINE (v1.5.1; Gusev et al. 2008). Briefly, this program identifies genomic regions shared identical by descent (IBD) over a specified minimum block size with a user-defined tolerance for mismatches. Autosomal biallelic SNPs with a minor allele frequency >0.05 across the 29 inbred laboratory strains were used for the inference of IBD haplotypes. A minimum block size of 200 kb and a cutoff of 0 mismatches were specified.

To relate the IBD block sizes in laboratory strains to those found in wild populations, I identified IBD regions in a set of 27 wild-caught *M. m. domesticus* mice from four populations in the native species range (Harr et al. 2016). The largest IBD track found in wild *M. m. domesticus* populations was 4.78 Mb in length. To ensure a focus on IBD regions in the laboratory strains that are shared by virtue of their descent from a single founder animal, I restrict all analyses of mutations in the laboratory strains to IBD regions spanning ≥5 Mb.

I then imposed a set of stringent filters to identify high-quality, SPVs that reside on IBD haplotypes ≥5 Mb:

(1) The variant is present as either a heterozygous or fixed singleton in 29 inbred laboratory strain genomes sequenced by the Sanger Mouse Genomes Project
(2) The variant passes all filters predefined in the Sanger Mouse Genomes Project VCF file, with the exception of the "Het" filter
(3) Variant is a biallelic SNP
(4) QUAL >50
(5) GQ >60
(6) DP >10 and DP <1.9× average sample coverage
(7) <15% missing data
(8) Site does not overlap segmental duplications annotated in mm10
(9) Site is not repeat-masked
(10) Genotype likelihood difference >20 between the most likely and next most likely genotype calls in the strain harboring the putative SPV
(11) Site is not polymorphic in wild *Mus musculus* or *Mus spretus* genomes (Harr et al. 2016) or present in wild-derived inbred strain genomes (WSB/EiJ, LEWES/EiJ, ZALENDE/EiJ, CAST/EiJ, MOLF/EiJ, PWK/PhJ, and

SPRET/EiJ). This filter is imposed to eliminate ancestral variants, but will also remove sites of frequent recurrent mutation (e.g., CpG dinucleotides).

(12) If the variant is heterozygous, allele balance ratio >0.3.

(13) If the variant is heterozygous, a χ^2 test on the null hypothesis of no allele bias and no strand bias yields $P > 0.05$

SPVs within IBD regions that passed these filters were then polarized into ancestral and derived states under the assumption that the major (i.e., nonprivate) allele is ancestral. The nucleotides flanking either side of each strain-private variant were extracted from the mm10 reference assembly and used to partition sites into their trinucleotide contexts.

Identification of Common Mouse Variants

Common variants were defined as biallelic SNPs in the Sanger Mouse Genomes data that are segregating in at least two laboratory strains. Variants in repetitive regions and annotated segmental duplications were excluded to match the filtering criteria employed for the detection of SPVs. The likely ancestral state at each site was defined using parsimony. Briefly, I identified the subset of common laboratory strain variants that are fixed for a single allele in the wild-derived inbred strains PWK/PhJ (*M. m. musculus*), CAST/EiJ (*M. m. castaneus*), and SPRET/EiJ (*M. spretus*). In these cases, the ancestral state is inferred to be the allele present in the wild-derived samples.

Strain Breeding Characteristics

Measures of strain breeding performance were obtained from the JAX5 database (Currer et al. 2009) downloaded from the Mouse Phenome Database (Bogue et al. 2018). One-way ANOVA tests treating strain as a factor were used to test for significant among-strain variation in measures related to reproductive aging.

Statistical Analyses

All analyses were carried out in the R environment for statistical computing (R Core Team 2016). Mutational spectra were compared using *G*-tests of independence. Principle component analysis was carried out on untransformed relative frequencies of six mutational classes (standardized by the nucleotide content of IBD regions) using the prcomp function in R. Data were zero-centered and scaled to have unit variance prior to analysis.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

References

Adams DJ, Doran AG, Lilue J, Keane TM. 2015. The Mouse Genomes Project: a repository of inbred laboratory mouse strain genomes. *Mamm Genome.* 26(9–10):403–412.

Agarwal I, Przeworski M. unpublished data. Available from: www.biorxiv.org/content/10.1101/519421v1, last accessed January 31, 2019.

Ahn S-M, Ansari AA, Kim J, Kim D, Chun S-M, Kim J, Kim TW, Park I, Yu CS, Jang SJ. 2016. The somatic POLE P286R mutation defines a unique subclass of colorectal cancer featuring hypermutation, representing a potential genomic biomarker for immunotherapy. *Oncotarget* 7(42):68638–68649.

Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Borresen-Dale A-L. 2013. Signatures of mutational processes in human cancer. *Nature* 500(7463):415–421.

Assaf ZJ, Tilk S, Park J, Siegal ML, Petrov DA. 2017. Deep sequencing of natural and experimental populations of *Drosophila melanogaster* reveals biases in the spectrum of new mutations. *Genome Res.* 27(12):1988–2000.

Baer CF, Miyamoto MM, Denver DR. 2007. Mutation rate variation in multicellular eukaryotes: causes and consequences. *Nat Rev Genet.* 8(8):619–631.

Beck JA, Lloyd S, Hafezparast M, Lennon-Pierce M, Eppig JT, Festing MF, Fisher EM. 2000. Genealogies of mouse inbred strains. *Nat Genet.* 24(1):23–25.

Besenbacher S, Sulem P, Helgason A, Helgason H, Kristjansson H, Jonasdottir A, Jonasdottir A, Magnusson OT, Thorsteinsdottir U, Masson G. 2016. Multi-nucleotide de novo mutations in humans. *PLoS Genet.* 12(11):e1006315.

Bogue MA, Grubb SC, Walton DO, Philip VM, Kolishovski G, Stearns T, Dunn MH, Skelly DA, Kadakkuzha B, TeHennepe G, et al. 2018. Mouse Phenome Database: an integrative database and analysis suite for curated empirical phenotype data from laboratory mice. *Nucleic Acids Res.* 4:D843–D850.

Bomba L, Walter K, Soranzo N. 2017. The impact of rare and low-frequency genetic variants in common disease. *Genome Biol.* 18(1):77.

Campbell CD, Chong JX, Malig M, Ko A, Dumont BL, Han L, Vives L, O'Roak BJ, Sudmant PH, Shendure J, et al. 2012. Estimating the human mutation rate using autozygosity in a founder population. *Nat Genet.* 44(11):1277–1281.

Cancer Genome Atlas Network. 2012. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 487:330–337.

Cancer Genome Atlas Research Network. 2013. Integrated genomic characterization of endometrial carcinoma. *Nature* 497:67–73.

Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S. 2009. AmiGO: online access to ontology and annotation data. *Bioinformatics* 25(2):288–289.

Carlson J, Locke AE, Flickinger M, Zawistowski M, Levy S, Myers RM, Boehnke M, Kang HM, Scott LJ, Li JZ, Zöllner S. 2018. Extremely rare variants reveal patterns of germline mutation rate heterogeneity in humans. *Nat Commun.* 9(1):3753.

Chen C, Qi H, Shen Y, Pickrell J, Przeworski M. 2017. Contrasting determinants of mutation rates in germline and soma. *Genetics* 207(1):255–267.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* 6(2):80–92.

Conrad DF, Keebler JEM, DePristo MA, Lindsay SJ, Zhang Y, Casals F, Idaghdour Y, Hartl CL, Torroja C, Garimella KV, et al. 2011. Variation in genome-wide mutation rates within and between human families. *Nat Genet.* 43(7):712–715.

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

Currer JM, Vonder Haar R, Corrow D, Flurkey K. 2009. Genetic quality control – preventing genetic contamination and minimizing genetic drift. In: Flurkey, K editor. The Jackson Laboratory Handbook on genetically standardized mice. Bar Harbor (ME): The Jackson Laboratory. p. 191–200.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The Variant Call Format and VCFtools. *Bioinformatics* 27(15):2156–2158.

de Boer J, Andressoo JO, de Wit J, Huijmans J, Beems RB, van Steeg H, Weeda G, van der Horst GT, van Leeuwen W, Themmen AP, et al. 2002. Premature aging in mice deficient in DNA repair and transcription. *Science* 296(5571):1276–1279.

DeMarini DM. 2004. Genotoxicity of tobacco smoke and tobacco smoke condensate: a review. *Mutat Res.* 567(2–3):447–474.

Dowty JG, Win AK, Buchanan DD, Lindor NM, Macrae FA, Clendenning M, Antill YC, Thibodeau SN, Casey G, Gallinger S, et al. 2013. Cancer risks for MLH1 and MSH2 mutation carriers. *Hum Mutat.* 34(3):490–497.

Dubrova YE, Bersimbaev RI, Djansugurova LB, Tankimanova MK, Mamyrbaeva ZZ, Mustonen R, Lindholm C, Hulten M, Salomaa S. 2002. Nuclear weapons tests and human germline mutation rate. *Science* 295(5557):1037.

Dubrova YE, Grant G, Chumak AA, Stezhka VA, Karakasian AN. 2002. Elevated minisatellite mutation rate in the post-Chernobyl families from Ukraine. *Am J Hum Genet.* 71(4):801–809.

Duncan BK, Miller JH. 1980. Mutagenic deamination of cytosine residues in DNA. *Nature* 287(5782):560–561.

Easton DF, Deffenbaugh AM, Pruss D, Frye C, Wenstrup RJ, Allen-Brady K, Tavtigian SV, Monteiro AN, Iversen ES, Couch FJ, et al. 2007. A systematic genetic assessment of 1,433 sequence variants of unknown clinical significance in the BRCA1 and BRCA2 breast cancer – predisposition genes. *Am J Hum Genet.* 81(5):873–883.

Eyre-Walker A, Keightley PD. 2007. The distribution of fitness effects of new mutations. *Nat Rev Genet.* 8(8):610–618.

Francioli LC, Polak PP, Koren A, Menelaou A, Chun S, Renkens I, van Duijn CM, Swertz M, Wijmenga C, van Ommen G, et al. 2015. Genome-wide patterns and properties of *de novo* mutations in humans. *Nat Genet.* 47(7):822–826.

Frazer KA, Eskin E, Kang HM, Bogue MA, Hinds DA, Beilharz EJ, Gupta RV, Montgomery J, Morenzoni MM, Nilsen GB, et al. 2007. A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. *Nature* 448(7157):1050–1053.

Frias S, Van Hummelen P, Meistrich ML, Lowe XR, Hagemeister FB, Shelby MD, Bishop JB, Wyrobek AJ. 2003. NOVP chemotherapy for Hodgkin's Disease transiently induces sperm aneuploidies associated with the major clinical aneuploidy syndromes involving chromosomes X, Y, 18, and 21. *Cancer Res.* 63:44–51.

Gao Z, Moorjani P, Amster G, Przeworski M. unpublished data. Available from: https://www.biorxiv.org/content/10.1101/327098v3, last accessed January 31, 2019.

Girard SL, Gauthier J, Noreau A, Xiong L, Zhou S, Jouan L, Dionne-Laport A, Spiegelman D, Henrion E, Diallo O, et al. 2011. Increased exonic *de novo* mutation rate in individuals with schizophrenia. *Nat Genet.* 43(9):860–863.

Goldman JM, Wong WSW, Pinelli M, Farrah T, Bodian D, Stittrich AB, Glusman G, Vissers LE, Hoischen A, Roach JC, et al. 2016. Parent-of-origin-specific signatures of *de novo* mutations. *Nat Genet.* 48:935–939.

Gusev A, Lowe JK, Stoffel M, Daly MJ, Altshuler D, Breslow JL, Friedman JM, Pe'er I. 2008. Whole population, genome-wide mapping of hidden relatedness. *Genome Res.* 19(2):318–326.

Hardison RC, Roskin KM, Yang S, Diekhans M, Kent WJ, Weber R, Elnitski L, Li J, O'Connor M, Kolbe D. 2003. Covariation in frequencies of substitution, deletion, transposition, and recombination during eutherian evolution. *Genome Res.* 13(1):13–26.

Harland C, Charlier C, Karim L, Cambisano N, Deckers M, Mni M, Mullaart E, Coppieters W, Georges M. unpublished data. Available from: https://www.biorxiv.org/content/10.1101/079863v2, last accessed December 3, 2018.

Harr B, Karakoc E, Neme R, Teschke M, Pfeifle C, Pezer Z, Babiker H, Linnenbrink M, Montero I, Scavetta R, et al. 2016. Genomic resources for wild populations of the house mouse, *Mus musculus* and its close relative *Mus spretus*. *Sci Data.* 3:160075.

Harris K. 2015. Evidence for recent, population-specific evolution of the human mutation rate. *Proc Natl Acad Sci U S A.* 112(11):3439–3444.

Harris K, Pritchard JK. 2017. Rapid evolution of the human mutation spectrum. *eLife* 6:e24284.

Hart RW, Setlow RB. 1974. Correlation between deoxyribonucleic acid excision-repair and life-span in a number of mammalian species. *Proc Natl Acad Sci U S A.* 71(6):2169–2173.

Helleday T, Eshtad S, Nik-Zainal S. 2014. Mechanisms underlying mutational signatures in human cancers. *Nat Rev Genet.* 15(9):585–598.

Hodgkinson A, Eyre-Walker A. 2011. Variation in the mutation rate across mammalian genomes. *Nat Rev Genet.* 12(11):756.

Hwang DG, Green P. 2004. Bayesian Markov chain Monte Carlo sequence analysis reveals varying neutral substitution patterns in mammalian evolution. *Proc Natl Acad Sci U S A.* 101(39):13994–14001.

Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, Yamrom B, Lee YH, Narzisi G, Leotta A, et al. 2012. De novo gene disruptions in children on the autistic spectrum. *Neuron* 74(2):285–299.

Jonsson H, Sulem P, Kehr B, Kristmundsdottir S, Zink F, Hjartarson E, Hardarson MT, Hjorleifsson KE, Eggertsson HP, Gudjonsson SA, et al. 2017. Parental influence on human germline de novo mutations in 1,548 trios from Iceland. *Nat Genet.* 549(7673):519–522.

Keane TM, Goodstadt L, Danecek P, White MA, Wong K, Yalcin B, Heger A, Agam A, Slater G, Goodson M, et al. 2011. Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* 477(7364):289–294.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The Human Genome Browser at UCSC. *Genome Res.* 12(6):996–1006.

Kimura M. 1983. The neutral theory of molecular evolution. Cambridge: Cambridge University Press.

Kondrashov AS. 2003. Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. *Hum Mutat.* 21(1):12–27.

Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, Gudjonsson SA, Sigurdsson A, Jonasdottir A, Jonasdottir A, et al. 2012. Rate of *de novo* mutations and the importance of father's age to disease risk. *Nature* 488(7412):471–475.

Kumar V, Kim K, Joseph C, Kourrich S, Yoo S-H, Huang HC, Vitaterna MH, de Villena FP, Churchill GA, Bonci A, et al. 2013. C57BL/6N mutation in *Cytoplasmic FMRP interacting protein 2* regulates cocaine response. *Science* 342(6165):1508–1512.

Lercher MJ, Hurst LD. 2002. Human SNP variability and mutation rate are higher in regions of high recombination. *Trends Genet.* 18(7):337–340.

Lindsay SJ, Rahbari R, Kaplanis J, Keane T, Hurles M. unpublished data. Available from: https://www.biorxiv.org/content/10.1101/082297v2, last accessed January 31, 2019.

Lombard DB, Chua KF, Mostoslavsky R, Franco S, Gostissa M, Alt FW. 2005. DNA repair, genome stability, and aging. *Cell* 120(4):497–512.

Lynch M. 2008. The cellular, developmental and population-genetic determinants of mutation-rate evolution. *Genetics* 180(2):933–943.

Lynch M. 2010. Evolution of the mutation rate. *Trends Genet.* 26(8):345–352.

Lynch M. 2011. The lower bound to the evolution of mutation rates. *Genome Biol Evol.* 3:1107–1118.

Lynch M. 2016. Mutation and human exceptionalism: our future genetic load. *Genetics* 202(3):869–875.

MacRae SL, Croken MM, Calder RB, Aliper A, Milholland B, White RR, Zhavoronkov A, Gladyshev VN, Seluanov A, Gorbunova V, et al. 2015. DNA repair in species with extreme lifespan differences. *Aging* 7(12):1171–1182.

Martin AP, Palumbi SR. 1993. Body size, metabolic rate, generation time, and the molecular clock. *Proc Natl Acad Sci U S A.* 90(9):4087–4091.

Mathieson I, Reich D. 2017. Differences in the rare variant spectrum among human populations. *PLoS Genet.* 13(2):e1006581.

McVicker G, Green P. 2010. Genomic signatures of germline gene expression. *Genome Res.* 20(11):1503–1511.

Moorjani P, Amorim CEG, Arndt PF, Przeworski M. 2016. Variation in the molecular clock of primates. *Proc Natl Acad Sci U S A.* 113(38):10607.

Nabholz B, Glémin S, Galtier N. 2008. Strong variations of mitochondrial mutation rate across mammals—the longevity hypothesis. *Mol Biol Evol.* 25(1):120–130.

Narasimhan VM, Rahbari R, Scally A, Wuster A, Mason D, Xue Y, Wright J, Trembath RC, Maher ER, van Heel DA, et al. 2017. Estimating the human mutation rate from autozygous segments reveals population differences in human mutational processes. *Nat Commun.* 8(1):303.

Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, Lin C-F, Stevens C, Wang L-S, Makarov V, et al. 2012. Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* 485(7397):242–245.

Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, Jones D, Hinton J, Marshall J, Stebbings LA, et al. 2012. Mutational processes molding the genomes of 21 breast cancers. *Cell* 149(5):979–993.

O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, Levy R, Ko A, Lee C, Smith JD, et al. 2012. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* 485:246–250.

O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, Karakoc E, MacKenzie AP, Ng SB, Baker C, et al. 2011. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat Genet.* 43(6):585–589.

Palamara PF, Francioli LC, Wilton PR, Genovese G, Gusev A, Finucane HK, Sankararaman S, Sunyaev SR, de Bakker PIW, Wakeley J, et al. 2015. Leveraging distant relatedness to quantify human mutation and gene-conversion rates. *Am J Hum Genet.* 97(6):775–789.

Park C, Qian W, Zhang J. 2012. Genomic evidence for elevated mutation rates in highly expressed genes. *EMBO Rep.* 13(12):1123–1129.

Phifer-Rixey M, Nachman MW. 2015. Insights into mammalian biology from the wild house mouse *Mus musculus.* eLife 4:e05959.

Qureshi ST, Larivière L, Leveque G, Clermont S, Moore KJ, Gros P, Malo D. 1999. Endotoxin-tolerant mice have mutations in Toll-like receptor 4 (Tlr4). *J Exp Med.* 189(4):615–625.

R Core Team. 2016. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing.

Rahbari R, Wuster A, Lindsay SJ, Hardwick RJ, Alexandrov LB, Turki SA, Dominiczak A, Morris A, Porteous D, Smith B, et al. 2016. Timing, rates and spectra of human germline mutation. *Nat Genet.* 48(2):126–133.

Robbins WA, Meistrich ML, Moore D, Hagemeister FB, Weier H-U, Cassel MJ, Wilson G, Eskenazi B, Wyrobek AJ. 1997. Chemotherapy induces transient sex chromosomal and autosomal aneuploidy in human sperm. *Nat Genet.* 16(1):74–78.

Robbins WA, Vine MF, Truong KY, Everson RB. 1997. Use of fluorescence in situ hybridization (FISH) to assess effects of smoking, caffeine, and alcohol on aneuploidy load in sperm of healthy men. *Environ Mol Mutagen.* 30(2):175–183.

Salcedo T, Geraldes A, Nachman MW. 2007. Nucleotide variation in wild and inbred mice. *Genetics* 177(4):2277–2291.

Sanders SJ, Murtha M, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, Ercan-Sencicek AG, DiLullo NM, Parikshak NN, Stein JL, et al. 2012. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 485(7397):237–241.

Scally A, Durbin R. 2012. Revising the human mutation rate: implications for understanding human evolution. *Nat Rev Genet.* 13(10):745–753.

Schaibley VM, Zawistowski M, Wegmann D, Ehm MG, Nelson MR, St Jean PL, Abecasis GR, Novembre J, Zollner S, Li JZ. 2013. The influence of genomic context on mutation patterns in the human genome inferred from rare variants. *Genome Res.* 23(12):1974–1984.

Schuster-Böckler B, Lehner B. 2012. Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature* 488(7412):504–507.

Ségurel L, Wyman MJ, Przeworski M. 2014. Determinants of mutation rate variation in the human germline. *Annu Rev Genomics Hum Genet.* 15:47–70.

Senra MVX, Sung W, Ackerman M, Miller SF, Lynch M, Soares CAG. 2018. An unbiased genome-wide view of the mutation rate and spectrum of the endosymbiotic bacterium *Teredinibacter turnerae.* *Genome Biol Evol.* 10(3):723–730.

Seoighe C, Scally A. 2017. Inference of candidate germline mutator loci in humans from genome-wide haplotype data. *PLoS Genet.* 13(1):e1006549.

Shi Q, Ko E, Barclay L, Hoang T, Rademaker A, Martin R. 2001. Cigarette smoking and aneuploidy in human sperm. *Mol Reprod Dev.* 59(4):417–421.

Shinbrot E, Henninger EE, Weinhold N, Covington KR, Göksenin AY, Schultz N, Chao H, Doddapaneni H, Muzny DM, Gibbs RA, et al. 2014. Exonuclease mutations in DNA polymerase epsilon reveal replication strand specific mutation patterns and human origins of replication. *Genome Res.* 24(11):1740–1750.

Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, et al. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* 15(8):1034–1050.

Smeds L, Qvarnström A, Ellegren H. 2016. Direct estimate of the rate of germline mutation in a bird. *Genome Res.* 26(9):1211–1218.

Srivastava A, Morgan AP, Najarian ML, Sarsani VK, Sigmon JS, Shorter JR, Kashfeen A, McMullan RC, Williams LC, Giusti RP, et al. 2017. Genomes of the mouse collaborative cross. *Genetics* 206(2):537–556.

Stamatoyannopoulos JA, Adzhubei I, Thurman RE, Kryukov GV, Mirkin SM, Sunyaev SR. 2009. Human mutation rate associated with DNA replication timing. *Nat Genet.* 41(4):393–395.

Sung W, Ackerman MS, Miller SF, Doak TG, Lynch M. 2012. Drift-barrier hypothesis and mutation-rate evolution. *Proc Natl Acad Sci U S A.* 109(45):18488–18492.

Uchimura A, Higuchi M, Minakuchi Y, Ohno M, Toyoda A, Fujiyama A, Miura I, Wakana S, Nishino J, Yagi T. 2015. Germline mutation rates and the long-term phenotypic effects of mutation accumulation in wild-type laboratory mice and mutator mice. *Genome Res.* 25(8):1125–1134.

Vissers LE, de Ligt J, Gilissen C, Janssen I, Steehouwer M, de Vries P, van Lier B, Arts P, Wieskamp N, del Rosario M, et al. 2010. A de novo paradigm for mental retardation. *Nat Genet.* 42(12):1109–1112.

Wade CM, Daly MJ. 2005. Genetic variation in laboratory mice. *Nat Genet.* 37(11):1175–1180.

Xu B, Roos JL, Dexheimer P, Boone B, Plummer B, Levy S, Gogos JA, Karayiorgou M. 2011. Exome sequencing supports a de novo mutational paradigm for schizophrenia. *Nat Genet.* 43(9):864–868.

Yang H, Bell TA, Churchill GA, Pardo-Manuel de Villena F. 2007. On the subspecific origin of the laboratory mouse. *Nat Genet.* 39(9):1100–1107.

Yang H, Wang JR, Didion JP, Buus RJ, Bell TA, Welsh CE, Bonhomme F, Yu AH-T, Nachman MW, Pialek J, et al. 2011. Subspecific origin and haplotype diversity in the laboratory mouse. *Nat Genet.* 43(7):648–655.

Zenzes MT. 2000. Smoking and reproduction:gene damage to human gametes and embryos. *Hum Reprod.* 6(2):122–131.