SUPPLEMENTAL INFORMATION

Estimation of the frequency of false negatives in SNV calls. To evaluate false-negative rates, we counted the correctly called number of variants in our pipeline within the tentatively generated synthetic mutations, as described previously (Keightley et al. 2014). In this analysis, we first randomly picked up 2,000 base points over the entire mouse genome, including the sex chromosomes, and generated synthetic substitution mutations at each point. To determine the false-negative rate in homozygous variants, we changed all of the major base reads at each selected point to a different, randomly selected base. For heterozygous variants, we altered some fraction of the major base reads, such that the number of alterations in the sequencing data was determined by the binomial distribution of the number of major base reads, at each point. To maintain the effects of sequencing errors, we did not change any non-major base reads. Then, the complete data set of reads including the synthetic mutations was remapped to the reference genomic sequence, and candidate *de novo* variants were called using an identical pipeline to that used in the present study. We performed this procedure, using the same 2,000 selected points, independently in conA, conB, mutC, and mutD. Of the 2,000 points, 1,126 were in the EWC region. Thus, a total of 4,504 (=1,126×4) homozygous and 4,504 heterozygous synthetic mutations were tested here. Only two heterozygous synthetic mutations were missed in our pipeline; one missing variant in mutC had a 23.7% altered allele frequency, which was removed by our filter setting the lower bound of the alternate allele frequency, and the other variant was in mutD, located near a repeat sequence that was presumably removed by misalignment (no missing variants were detected in conA and conB). Although we used a binomial distribution in

generating heterozygous synthetic mutations, which presumably underestimated the false negative rate because the actual sequencing read data was likely to be more variable, the results suggested that the false negative rate was not very high.

We also looked for missing calls that would drop out with our custom filters for credible variant calls: (i) the lower bound of the alternate allele frequency was set to 25% for a heterozygous variant call, and (ii) alternative allele identification required both forward and reverse strand coverage. For filter-(i), histograms of candidate variant frequencies showed the presence of a few low-frequency variants (<25%), which were excluded from the heterozygous variant number in this analysis. Presumably, many of these low-frequency variants would not be genuine germline mutations, indicating the possibility of a few false negatives (at most 3%) (Supplemental Fig. S2). Filter-(ii) removed fewer raw variants—seven heterozygous variants, one in mutC, four in mutD, and two in conB. These results also suggested that the conditions we used to call variants did not yield a high rate of false negatives.

Highly accurate detection of de novo variants in our EWC regions. Our detection of *de novo* variants with few sequencing errors and few missing variants was highly accurate compared to previous reports (Keane et al 2011, Simon et al 2013). We propose three possible reasons for this high accuracy: (1) General analysis conditions. Our high-throughput sequencing conditions (100-bp or 150-bp paired end sequencing, high coverage sequencing, and the use of an updated mm10 database [a reference for the C57BL/6 genome]) should provide a better analysis than the previous conditions described in Simon et al, 2013. (2) Our EWC regions and variant call conditions. When we prepared the EWC regions, we adjusted many of the settings, including

the variant call conditions, and checked some *de novo* candidate variants by manual inspection and Sanger sequencing. In this study, we had a relatively large number of called *de novo* variants (derived from mutator lines) and the sequencing data of the original mice "Adam/Eve" to use as a reference. The high number of candidates yielded a well-shaped frequency distribution of called variants (Supplemental Fig. S2), which served as a useful landmark for helping us find good conditions. The sequencing data of "Adam/Eve" enabled us to efficiently remove artificial errors derived from the high-throughput sequencing. To evaluate our conditions for detecting *de novo* variants, we used another software, the DeNovoGear software package (Ramu et al. 2013). The lower limit of read coverage was set to 0, without a contradiction filter. This setting is frequently used in trio analysis of *de novo* variants. The DeNovoGear results were similar to those obtained with SAMtools/BCFtools, in which there were at most 3% called homozygous or heterozygous variants. These approaches differed mostly in their treatment of low-frequency variant calls and highly repetitive sites. We concluded that the SAMtools/BCFtools were more suitable for this analysis because of the higher credibility of variant calls, in both the heterozygous and homozygous states. In our experience, the selection of EWC regions and the filtering out of sequencing errors were crucial for the highly accurate detection of *de novo* mutations. (3) Damaged genomic DNA sample. This is just a possibility. As shown in Supplemental Table S1, the whole-genome sequencing results for the "Eve" genome were not good, presumably due to the poor condition of the preserved genomic DNA (it is possible that accidental freezing and thawing damaged the sample). As a result of this problem with the "Eve" genome, our EWC regions became more restricted. We cannot rule out the possibility that damage-sensitive genomic regions would have

been difficult regions in which to detect *de novo* variants.

In our preliminary data (not included in our presented data), we tested the more suspicious candidate *de novo* variants (heterozygous SNVs on chromosome X in males, for which we used analysis settings similar to those used for the EWC region) by Sanger sequencing. All of the tested candidates (a total of 5 SNVs: 2 SNVs from mutC, 3 SNVs from mutD) were confirmed to be true *de novo* variants (presumably, ~50% mosaic mutations), present in the sequenced individual but not in its parents. The 3 SNVs from mutD were confirmed to be inherited by some females of the next generation. (The 2 SNVs from mutC could not be analyzed due to a lack of daughters from the mutC sequenced mouse.) These data supported the accuracy of our detection of *de novo* variants and indicated that the high accuracy of our sequencing analyses might have increased our ability to detect somatic mosaic mutations in C57BL/6 mice.

Frequency of recurrent mutations in mutator mice. Abundant recurrent mutations would interfere with our estimation of per-generation mutation rates. To evaluate the extent of recurrent mutations in the mutator mice, we compared all of the homozygous and heterozygous *de novo* variants in the mutC or mutD lines with variants called from the whole-genome sequencing data of an independently derived mutator line, mutE, at the 12th generation (shown in Fig. 1). Only two identical SNVs were found; one was shared by the mutE and mutC lines, and the other by the mutE and mutD lines. Both shared variants occurred at a homopolymer site, and the variants were confirmed to be *de novo* mutations in both lines by Sanger sequencing. This result indicated that there was some bias in mutation occurrence, but that the frequency of recurrent mutations was low in the mutator breeding lines, and the effect of recurrent mutations

on our mutation-rate estimates was negligible.

Effect of initial variants on mutator-line phenotypes. To produce *Pold1*$^{exo/exo}$ mice, we first used C57BL/6J ES cells to generate *Pold1* gene-targeting mice, which contained a floxed polyadenylation signal and a Neo-resistance cassette. We then crossed the gene-targeted mice with Tg mice expressing *Cre* in the germline (C57BL/6), leading to the constitutive expression of 3'-5' exonuclease-deficient Pold1 (D400A), and the generation of *Pold1*$^{exo/+}$ mice. To produce the two types of mutator (*Pold1*$^{exo/exo}$) mice used in our present breeding experiment, we intercrossed *Pold1*$^{exo/+}$ mice before or after backcrossing to C57BL/6J mice at least eight times. The backcross was performed to simultaneously prevent mutation accumulation and maintain a C57BL/6J genetic background. The mutator lines mutA, mutB, mutC, mutD, mutE, mutF, and mutP were generated by crossing heterozygotes before the backcross, while the mutator lines mutK, mutL, mutM, mutN, and mutR were generated by crossing heterozygotes after the backcross. Notably, whole-genome sequencing called some candidate genetic polymorphisms between Adam and Eve in the mutC and mutD lines. However, in the absence of filtration, it was difficult to determine whether these called variants were sequencing artifacts or actual polymorphisms. To evaluate the possible effects of variants in the base population, we compared the numbers of called polymorphism sites (initial) and no call sites (*de novo*) in Adam and Eve, both of which were observed as homozygous sites in the mutC or mutD lines (Supplemental Table S14). Although we do not know the false-positive rate of these initial variants, there were ~800 candidate variants in each line, about half of which were known polymorphisms in C57BL/6N (Mouse Genomes Project SNP and Indel Release v3 of the

Sanger Center). The C57BL/6N contamination was presumably introduced with the *Cre*-expressing Tg mice. Although the number of initial SNVs was smaller than the number of accumulated *de novo* SNVs in sequenced individuals of the mutC and mutD lines, these initial variants could certainly affect the phenotypes in mutC and mutD to some extent. We observed that all of the phenotypic results, except for the tendency of the male body weight to decrease, had similar tendencies in the backcrossed mutator lines (mutK, mutL, mutM, mutN, and mutR). These lines would be expected to contain far fewer initial genetic variations, indicating that the phenotypic changes observed in the mutator mouse lines resulted from many accumulated *de novo* mutations rather than from the initial variants.

SUPPLEMENTAL METHODS

Phenotypic screening for anomalies. Mice were screened for hydrocephaly and other apparent anomalies at 4 weeks of age, and were assessed at 8 weeks of age for coat color and eye appearance, for behavioral abnormalities in the home-cage environment, and for morphological anomalies in the digits, skull, tail, and elsewhere. Tumorigenic phenotypes related to the *Pold1*$^{exo/exo}$ genotype were excluded from the anomaly analysis. Phenotypic frequencies of the breeding lines (Supplemental Table S8) were analyzed for statistical significance by the χ^2-test (and Fisher's exact test for a rough indication) with GraphPad PRISM 6.0 software.

Body-weight analysis. Body weights were typically measured in the afternoon (14:00-18:00). To compare specific traits across breeding lines, we used at least 50 mice from each breeding line (or mice from sub-lines that were within six relatives of the ancestral pair) (Fig. 1); the number (male, female) of each line was C(98,84), D_1(66,74), D_2(30,27), A_1(33,26), A_2(46,45), A_3(29,31), A_4(61,72), P(63,50), F(132,132), B_1(136,140), B_2(86,85), B_3(38,34), E_1(146,149), E_2(63,82), E_3(60,61), E_4(103,82), E_5(21,31), K(105,105), M(63,68), N(88,99), R(62,46), conA_1(47,45), conA_2(31,23), conE(30,30), conD(49,48), conB(67,69), conF(39,36), and conC(54,50). We divided the breeding lines as follows: four early-generation (six or fewer generations) mutator lines, 17 late-generation (more than six generations) mutator lines, and seven control lines. We grouped the late-generation control and mutator lines according to their ancestral pair (e.g., Adam/Eve in Fig. 1), and analyzed differences in the breeding-line groups (represented by colors in Fig. 3A and Supplemental Fig. S6A) by one-way ANOVA and Tukey's multiple-comparison post-test using GraphPad PRISM 6.0 software. Body weights

were measured at 7−10 weeks of age prior to February 2010, and at 8 weeks of age thereafter. Only body-weight data from 8-week-old mice were included in the present analysis.

Analysis of changes in reproductive ability. All mice were examined at 8 weeks of age or later, and all capable females were assessed for reproductive ability. Virginal females were mated with an arbitrarily selected male littermate over a 1-week period. We recorded the number of births per mating trial (birth rate) and the number of pups born (the P0 litter size) (see Supplemental Tables S9, S10). We defined the survival rate as the percentage of pups reaching 8 weeks of age. To exclude the effect of the *Pold1*$^{exo/exo}$ tumorigenesis genotype on reproductive ability, we excluded data from matings if either parent died within 9 weeks of the start of mating, regardless of the cause of death. This excluded most of the data from hydrocephalic mice as well, since these mice often died at 1−3 months of age. Data relating to reproductive ability were recorded for individual breeding lines beginning in February 2010. To determine statistical significance, we used GraphPad PRISM 6.0 software (Supplemental Table S9, S10).

Regression analysis on phenotypic analyses. We used regression analysis to investigate whether the three kinds of phenotypes—body-weight, reproduction (represented by the offspring number per a mating), and visible abnormalities—showed trends over generations of breeding. Except in an additional analysis of the offspring number using the recessive lethal mutation model, the means of the responses were modeled as simple linear functions of the generation number: $E(y) = \beta_0 + \beta_1 \times$ generation number, where the β_i's are parameters to be

estimated, and y is body weight, the presence or absence of a phenotypic abnormality (1 or 0), or the offspring number. The distributions of y conditional on the generation number and β_i's were assumed to be as follows: for body weight, a normal distribution with an unknown variance parameter δ^2; for offspring number, a negative binomial with an unknown overdispersion parameter k; and for phenotypic abnormality (binary data), binomial. The negative binomial model was used because we found that in a given generation, the variances in offspring number were much larger than the means (on average, 2.78-fold higher for control mice, and 3.12-fold higher for mutator mice). All data points in each analysis were assumed to be independent. Body weight varied by sex and was analyzed for each sex separately.

Body-weight changes relative to the generation number were analyzed by simple linear regression; the CI and P values for the slope (β_1) were obtained by standard methods. For the other two phenotypes, the P value for β_1 was obtained using the fact that the log-likelihood ratio is asymptotically distributed as χ^2 with one degree of freedom. The CIs for β_1 and the mean responses (shown only in the analysis of offspring numbers) were obtained by the Wald method (Wald CI). For the above regression analyses, we used the lm and glm functions in the stat package and the glm.nb function in the MASS package in the R software version 3.0.3 (http://www.r-project.org) to analyze body weight, the presence or absence of a phenotypic abnormality, and the offspring number, respectively.

For the offspring number, we also fitted the relatively simple "recessive lethal mutation model," presented previously (Lyon 1959) (see APPENDIX I, “Frequency of intercrosses for a lethal”), for full-sib matings. Consider a full-sib mating line. Assume that a is a recessive lethal mutation, and A is its counterpart wild-type allele in any locus. Consider that one *de novo*

recessive lethal mutation, *a,* occurs in a locus in the initial generation. Let X_n, Y_n, and Z_n denote the occurrence frequencies in matings of individuals with the genotypes *AA*×*AA*, *AA*×*Aa*, and *Aa*×*Aa,* respectively, in the n^{th} generation. Then $X_0 = Z_0 = 0$ and $Y_0 = 1$, and for $n > 0$

$$\begin{cases} X_{n+1} = X_n + \frac{1}{4}Y_n + \frac{1}{12}Z_n \\ Y_{n+1} = \frac{1}{2}Y_n + \frac{1}{3}Z_n \\ Z_{n+1} = \frac{1}{4}Y_n + \frac{1}{3}Z_n \end{cases}$$

Next, consider k_n, which denotes the total number of *Aa*×*Aa* matings for the whole genome in the n^{th} generation. All loci are assumed to evolve independently. Accounting for the successive mutation effect from the 0 to $(n-1)^{th}$ generation, the expectation of k_n is given by

$$\mathrm{E}[k_n] = 2U \sum_{r=0}^{n-1} Z_{n-r},$$

where U is the lethal mutation rate (the *de novo* mutation number per generation per diploid), and $2U$ is the per-sib mutation rate. The reproductive ability (offspring number) in the n^{th} generation relative to the 0^{th} generation is given by $(3/4)^{k_n}$, since *Aa*×*Aa* mating leaves as many 3/4 mice as the other two mating types. Thus, as a first-order approximation (the delta method), the expectation of the offspring number in the n^{th} generation is given by

$$\mathrm{E}[y_n] = \beta_0 \mathrm{E}\left[\frac{3}{4}^{k_n}\right] \approx \beta_0 \frac{3}{4}^{\mathrm{E}[k_n]}$$

where β_0 is the reproductive ability in the 0^{th} generation. The values of $\mathrm{E}[y_n]$ given U and β_0 in the n^{th} generation are calculated numerically.

The distribution of y conditional on the generation number and the parameters U and β_0 was still assumed to be a negative binomial with an unknown overdispersion parameter k, and all the obtained values of y were assumed to be independent of each other. This complex

regression analysis cannot be conducted with the standard default settings on most statistical packages. To maximize likelihood, we used the R function nlm(), which can minimize arbitrary nonlinear functions. Similar to the analyses just described, the *P* value was obtained by the log-likelihood ratio, and the CI was calculated by the Wald method with the Hessian matrix, obtained by nlm(), at the maximum likelihood estimates. We used the Akaike information criterion (Akaike 1974) (AIC) to compare two models: the recessive lethal mutation model, and the linear function model of generation-dependent trends we just described. Supplemental Table S15 summarizes the regression analysis.

Variant calls and *de novo* structural variants. For all samples, structural variants (SVs) were detected with BreakDancerMax v1.1.2 (Chen et al. 2009). We used bam2cfg.pl, which is included in the BreakDancer package, to generate a BreakDancerMax parameter file from the BAM-format mapping data. We used whole mapping data without duplicates as input for bam2cfg.pl and BreakDancerMax, since improperly mapped read pairs are useful for detecting SVs. Although BreakDancerMax detects several types of SV candidates, we focused on large deletions, because other SVs are difficult to detect with high confidence. To identify *de novo* large deletions, we extracted large deletions detected in mutC, mutD, conA, and conB that were not in the original population (Adam/Eve). We subjected these deletions to the following custom filters: (i) the lower bound of the alternate allele frequency was set to 25%, and (ii) the minimum number of read pairs supporting deletions was set to 20. In addition, we used GBrowser (Stein 2013) to visually inspect the short-read sequencing data of all 166 unique sites, and found 5 candidate *de novo* variant sites. PCR followed by Sanger sequencing of genomic

DNA samples from the sequenced individuals and their ancestors confirmed the presence of two true *de novo* deletions (Supplemental Table S13) and three variants derived from the base population.

Cytogenetic analysis. Splenocytes were cultured in RPMI1640 medium supplemented with 10% FBS, 15 μ g/ml LPS, and 5 μ g/ml Con A. After 46 hours, colcemid was added at 1 μ g/ml for 20 min prior to harvesting. A hypotonic solution (0.075 M KCl) was added to the cell pellet, and the samples were incubated for 20 min at 37 °C. The cells were then fixed with Carnoy's fixative (3:1 methanol to glacial acetic acid). Metaphase spreads were prepared and air-dried. A solid Giemsa stain was used to analyze the chromosome number and structure.

Confidence Intervals (CIs) for μ and combined estimates. If all *de novo* mutations are inherited independently, the numbers of mutant heterozygote or homozygote sites at a given generation are distributed as Poisson random variables with the parameter $\mathrm{M} = \mu \times \mathrm{G} \times L$. The CIs for μ can then be calculated under the Poisson assumption. In fact, however, mutations in the same chromosome are inherited perfectly or partially together, and the variance in the number of *de novo* mutations (and thus the CIs for μ) may be larger than under the Poisson assumption. We conducted computer simulations to check the Poisson assumption and to determine the 95% CIs for μ when the assumption is violated. The lower (L) and upper (U) limits were experimentally chosen so that the P value $\mathrm{P_L}(\hat{\mu} \geq \text{observed } \hat{\mu}) = 2.5\%$ and $\mathrm{P_U}(\hat{\mu} \leq \text{observed } \hat{\mu}) = 2.5\%$, respectively. The point estimator of mutation rates given by $\hat{\mu} = \mathrm{M}/(\mathrm{G} \times L)$ was also confirmed to be unbiased, as will be shown later.

We simulated the real inheritance process of the mutations observed in our experiments. The 19 autosomal chromosomes were given relative lengths according to the actual lengths in a mouse, although the results were only slightly different if the lengths were the same. Prior to fertilization, mutations occurred randomly in EWC sites (and the EWC sites themselves were random within the genome) at a rate of μ per site per generation, or at the Poisson rate of 2×μ × (the size of the EWC region) per diploid EWC region. The size of the EWC was 1,516,416,340 for SNVs, and 961,909,845 for indels, assuming no recurrent mutations. Next, genetic recombinations occurred randomly in the genome at a rate of r cM/Mb, or the Poisson rate of r×2,462,745,373 (the genome size) $\times 10^{-8}$ in the total genome. After mutations and recombinations, the 19 autosomal chromosomes were independently inherited by the offspring. The strains were maintained by sibling mating. Repeating the above steps at a given generation (16^{th}, 17^{th}, 18^{th}, and 21^{st} generations in the mutC, mutD, conA, and conB lines, respectively), we obtained the number of mutant heterozygous and homozygous sites (M) in one sibling, and $\hat{\mu} = \mathrm{M}/(\mathrm{G}\times L)$ was calculated based on each M. There were 1,001 simulations per given mutation rate μ.

Using the simulated data for various mutation-rate levels, we checked the overdispersion of the number of mutant heterozygous or homozygous sites compared with Poisson variance (Supplemental Table S16). Here, overdispersion was defined as the squared ratio of the variance to the mean of the number of heterozygous or homozygous sites. The overdispersion value shows the CI widths relative to those obtained by Poisson. The recombination rates (cM/Mb) were set as r=0.5, 0.6, and 0.7, which are likely values for the mouse. As expected, high mutation rates resulted in high overdispersion, because recombination

does not overwhelm mutation. If clear overdispersion was seen, we used a simulation-based method to calculate the CIs, as we will explain (See Supplemental Table S16, the right-most column). The overdispersion values did not vary widely at r = 0.5~0.7. We calculated CIs using r = 0.6, a value derived from the mean among 19 chromosomes, as shown in the revised Shifman map in Table 1 of a previous report (Cox et al. 2009).

Following the above simulations, we calculated the experimental CIs by obtaining the 2.5^{th} and 97.5^{th} percentile of $\hat{\mu}$ (the 26^{th} lowest and highest values of 1001 simulations) per given mutation-rate value (μ); the 2.5^{th} and 97.5^{th} percentiles were regressed on the true μ. The μ values of the intersection point between the regression lines and the observed $\hat{\mu}$ were determined with 95% confidence limits.

Supplementary Fig. S10 shows examples of CIs calculated by the above method. The 2.5^{th} and 97.5^{th} percentiles are on or almost on the regression lines. Thus, the 2.5^{th} and 97.5^{th} regression lines are good representatives of the true 2.5^{th} and 97.5^{th} percentiles, and CIs obtained using these lines should be reliable. This analysis also shows that μ can be estimated by $\hat{\mu}$ with no bias. (The mean values of $\hat{\mu}$ are on the diagonal lines.)

For the mutator lines (mutC and mutD), *de novo* mutations (M) should be detected with few false-positive SNV or indel mutations. For the control lines (conA and conB), candidate *de novo* mutations included both SNV and indel initial variants (Supplemental Table S3). In the conA line, for example, we found one initial variant of 30 (=31-1) to be a candidate heterozygous mutation. In these cases, the "observed" *de novo* mutations replaced the "expected" *de novo* mutation numbers. In the conA line, the number of "expected" mutations would be calculated as $75\times29/30 + 29 = 101.5$ mutations.

The overall mutation rates were estimated using the formula $\hat{\mu}_{\mathrm{tot}} = \mathrm{M}_{\mathrm{tot}}/(\mathrm{G}\times L_{\mathrm{tot}})$, similar to the mutation rate estimates for each line's SNVs and indels. Here $\mathrm{M}_{\mathrm{tot}}$ is the total number of mutations used for the overall estimates: $\mathrm{M}_{\mathrm{tot}} = 101.5 + 92.7 = 194.2$ for SNVs in the control lines, $\mathrm{M}_{\mathrm{tot}} = 1{,}304 + 1{,}944 + 1{,}472 + 1{,}633 = 6{,}353$ for SNVs in the mutant lines, $\mathrm{M}_{\mathrm{tot}} = 4 + 3 = 7$ for indels in the control lines, and $\mathrm{M}_{\mathrm{tot}} = 28 + 28 + 21 + 37 = 114$ for indels in the mutant lines. L_{tot} is the sum of *L*s used for the mutation rate estimation for each line's SNVs and indels. Note that the homozygous mutations in control lines were excluded due to the low ability to discriminate between true mutations and the initial variants. The CIs for the overall mutation rates for indels were calculated simply based on a Poisson assumption, since the mutation rate for indels was low enough to result in low overdispersion (Supplemental Table S16). For the overall mutation rates for SNVs, the CIs were calculated by performing 1,001 simulations of $\hat{\mu}_{\mathrm{tot}}$ as described above.

Our CIs for μ based on the expected number of *de novo* mutations in conA and conB might be anti-conservative, because this calculation ignored the possibility of random error in estimating the expected false-positive rate from the randomly selected Sanger sequencing (except for heterozygous *de novo* indels in conA and conB, which were all confirmed). To determine how much of an effect such random error would have, we calculated conservative CIs for the mutation rates in conA and conB as follows. As an example, consider heterozygous SNVs in the conA line. We calculated the 95% range of the true mutation frequency (= 1 – "false positive rate"), (0.83, 1.00) in conA, estimated from the binomial sampling of one initial variant out of 30 tested variants (Supplemental Table S3), and then calculated the 95% range of the number of *de novo* variants: 29 + 75 (not tested variants) × 95% CI (0.83, 1.00) = (91.1,

104.0). The numbers 91.1 and 104.0 give 95% CIs for μ of (3.9-6.6, $\times 10^{-9}$) and (4.5-7.5, $\times 10^{-9}$), respectively. We then used the two extreme values (3.9-7.5, $\times 10^{-9}$) as the 95% conservative CI for μ in conA heterozygous SNVs. This method is ad hoc, but it seems to give reasonable CIs for μ; conA: (0.5-8.6, $\times 10^{-9}$), conB: (1.5-9.5, $\times 10^{-9}$) for homozygous SNVs; conA: (3.9-7.5, $\times 10^{-9}$), conB: (3.5-7.1, $\times 10^{-9}$) for heterozygous SNVs; conA: (0.04-1.46, $\times 10^{-9}$), conB: (0.01-1.19, $\times 10^{-9}$) for homozygous indels; overall rate in controls: (4.2-6.7, $\times 10^{-9}$) for SNVs. These results show that the estimated mutation rates for homozygous SNVs and indels have much wider ranges than the previously calculated 95% CIs (shown in Table 1), indicating that the estimates from the homozygous variants in conA and conB are less reliable than the others.

References in the Supplemental Material

Akaike H. 1974. New Look at Statistical-Model Identification. *Ieee T Automat Contr* **19**: 716-723.

Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. 2012. Predicting the functional effect of amino acid substitutions and indels. *PloS ONE* **7**: e46688.

Chen K, Wallis JW, McLellan MD, Larson DE, Kalicki JM, Pohl CS, McGrath SD, Wendl MC, Zhang Q, Locke DP et al. 2009. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat Methods* **6**: 677-681.

Cox A, Ackert-Bicknell CL, Dumont BL, Ding Y, Bell JT, Brockmann GA, Wergedal JE, Bult C, Paigen B, Flint J et al. 2009. A new standard genetic map for the laboratory mouse. *Genetics* **182**: 1335-1344.

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.

Keightley PD, Ness RW, Halligan DL, Haddrill PR. 2014. Estimation of the spontaneous mutation rate per nucleotide site in a Drosophila melanogaster full-sib family. *Genetics* **196**: 313-320.

Lyon FM. 1959. Some evidence concerning the "Mutational load" in inbred strains of mice. *Heredity* **13**: 341-352.

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

Simon MM, Greenaway S, White JK, Fuchs H, Gailus-Durner V, Wells S, Sorg T, Wong K, Bedu E, Cartwright EJ et al. 2013. A comparative phenotypic and genomic analysis of C57BL/6J and C57BL/6N mouse strains. *Genome Biol* **14**: R82.

Stein LD. 2013. Using GBrowse 2.0 to visualize and share next-generation sequence data. *Brief Bioinform* **14**: 162-171.