Paternal age in rhesus macaques is positively associated with germline mutation accumulation but not with measures of offspring sociability

Richard J. Wang [1], Gregg W.C. Thomas[1,2], Muthuswamy Raveendran[3,4], R. Alan Harris[3,4], Harshavardhan Doddapaneni[3,4], Donna M. Muzny[3,4], John P. Capitanio[5], Predrag Radivojac[2,6], Jeffrey Rogers[3,4] and Matthew W. Hahn[1,2]

[1] Department of Biology, Indiana University, Bloomington, IN
[2] Department of Computer Science, Indiana University, Bloomington, IN
[3] Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX
[4] Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX
[5] California National Primate Research Center, UC Davis, Davis, CA
[6] Khoury College of Computer Sciences, Northeastern University, Boston, MA

Abstract

Mutation is the ultimate source of all genetic novelty and the cause of heritable genetic disorders. Mutational burden has been linked to complex disease, including neurodevelopmental disorders such as schizophrenia and autism. The rate of mutation is a fundamental genomic parameter and direct estimates of this parameter have been enabled by accurate comparisons of whole-genome sequences between parents and offspring. Studies in human have revealed that the paternal age at conception explains nearly all variation in mutation rate: each additional year of paternal age in humans leads to approximately 1.5 additional mutations inherited by the child. Here, we present an estimate of the *de novo* mutation rate in the rhesus macaque (*Macaca mulatta*) using whole-genome sequence data from 32 individuals in four large pedigrees. We estimated an average mutation rate of 0.37×10^{-8} per site per generation (at an average parental age of 7.5 years), much lower than found in direct estimates from other primates (including human, chimpanzee, and gorilla). As in humans, older macaque fathers transmit more mutations to their offspring, approximately 1.3 extra mutations per year in our probands. Mutations at CpG sites accounted for 24% of all observed point mutations and were only weakly associated with paternal age, consistent with a non-replicative origin of mutations at these sites. We found that the rate of mutation accumulation after puberty is lower in macaques than in humans, but that most of the difference in overall mutation rate is due to a smaller number of mutations that accumulate before puberty in macaques. We additionally investigated the role of paternal age on offspring sociability, a proxy for normal neurodevelopment. In 203 male macaques studied in large social groups, we found no relationship between paternal age and multiple measures of social function. Our findings support the hypothesis that the increased risk of neurodevelopmental disorders with paternal age in primates is not primarily due to *de novo* mutations.

Introduction

Paternal age at conception is the single strongest predictor of the number of *de novo* mutations that a human will inherit. Studies show that children will inherit approximately 1.5 additional mutations per year of paternal age, and that the average father contributes new mutations at a rate that is three to four times greater than the mother per year (Kong et al. 2012; Besenbacher et al. 2015; Francioli et al. 2015; Jónsson et al. 2017). This male bias in the number of *de novo* mutations has been attributed to the continuous nature of spermatogenesis, which results in the accumulation of replication errors in the male germline (Crow 2000). As new mutations are responsible for the incidence of spontaneous genetic disorders, this pattern has considerable implications for human health. The male bias in mutation rate has also been observed in other primates (Venn et al. 2014; Thomas et al. 2018), making it an important feature in models of mutation rate evolution (Thomas and Hahn 2014; Amster and Sella 2016; Moorjani et al. 2016; Besenbacher et al. 2019).

Children conceived by older fathers are also at greater risk for numerous adverse developmental outcomes (Crow 2000). These include a greater risk of certain genetic disorders due to new mutations inherited at single genes, as with Apert syndrome, Marfan syndrome, and achondroplasia (Glaser et al. 2003; Green et al. 2010; Goriely et al. 2013). Evidence from epidemiological studies also suggests a relationship between advanced paternal age and complex neurodevelopmental disorders: an increased risk of schizophrenia, autism, and bipolar disorder have all been associated with advanced paternal age (Sipos et al. 2004; Reichenberg et al. 2006; Durkin et al. 2008; Frans et al. 2008; Menezes et al. 2010; Lee et al. 2011). The mechanisms underlying these epidemiological associations with paternal age have not been conclusively determined. A central question lies in whether the risk inherited from older fathers comes from genetic predisposition or *de novo* mutations. In the *de novo* model, neurological disorders are highly polygenic and result from the additive effects of a rising mutational burden (Malaspina et al. 2002; Hultman et al. 2011; Kong et al. 2012; Ronemus et al. 2014). The competing predisposition hypothesis attributes the epidemiological association to underlying pre-existing genetic factors that may actually contribute to delayed reproduction in males (Gratten et al. 2016; Janecka et al. 2017). For example, a genetic correlation between psychiatric disorders and delayed fatherhood could explain the association seen with advanced paternal age.

Primate models provide a powerful experimental system for investigating the relationship between the increasing number of *de novo* mutations and the increased risk of developmental outcomes with paternal age. Direct estimation of the mutation rate by tracking the appearance of new germline mutations across generations has become possible with advances in the availability and economy of next-generation sequencing technologies. While there are some technical concerns in sampling and sequencing of rare *de novo* events (Ségurel et al. 2014; Scally 2016), there are few barriers to estimating mutation rates in any species. Along with estimates of *de novo* mutations, testing this relationship also requires information on neurodevelopmental

progression in non-human systems. Although such data are much harder to collect, several long-term studies of captive primate populations have tracked multiple aspects of neurological and social abilities. In the important model system, rhesus macaque (*Macaca mulatta*), researchers have found high levels of social intelligence (Capitanio 1999; Sclafani et al. 2016). As a consequence, the rhesus monkey has become a model for studying schizophrenia (Gil-da-Costa et al. 2013) and autism spectrum disorder (Bauman and Schumann 2018; Parker et al. 2018).

Here, we performed whole-genome sequencing on 32 rhesus macaque individuals from four multiple-generation pedigrees to identify *de novo* mutations. We find that the mutation rate increases with paternal age in rhesus macaques, at a rate similar to humans, though there are substantially fewer mutations in macaques upon reaching puberty. Nevertheless, our findings are largely consistent with a model in which differences in mutation rates per-generation between species are determined by differences in life-history traits, rather than by changes in the proofreading or replication proteins themselves. By analyzing behavioral data from a large, captive macaque population, we also show that there is no association between paternal age at conception and behavioral metrics of sociability, including those linked to autism. Our data support a model of neurodevelopmental disease risk from inherited genetic factors that do not require *de novo* change.

Results

The number of mutations inherited increases with paternal age in rhesus monkeys

We sequenced 32 individuals from 4 three-generation families of rhesus monkeys (Figure S1) to high coverage using Illumina short-read sequencing. These families contained 14 trios with different offspring, from which we identified 303 *de novo* single nucleotide mutations. After controlling for the observable size of the genome, applying stringent quality filters, and verifying the transmission frequency of mutations to the second generation (Materials and Methods), we estimated an average mutation rate of 0.37×10^{-8} per site per generation for parents with an average age of 7.5 years.

We found a strong association between paternal age and the number of *de novo* mutations inherited by offspring (Figure 1). For each additional year of paternal age at conception, offspring inherited an additional 1.3 *de novo* mutations ($R^2 = 0.75$, $p < 5 \times 10^{-5}$). Due to the structure of the pedigree among sampled individuals, we were able to unambiguously phase 125 of the 303 mutations (Materials and Methods). We found that 66% of these phased mutations were inherited from the father, consistent with a paternal origin for the age effect. In contrast to this relationship, we found no significant association between maternal age and either the number of mutations inherited or the estimated per-generation mutation rate (Figure S2). Early

studies of human trios also found no significant effect from maternal age (Kong et al. 2012; Francioli et al. 2015), though studies with much larger sample sizes have detected a small effect, likely due to the accumulation of environmental damage in the maternal germline (Goldmann et al. 2016; Jónsson et al. 2017).

We found that C→T transitions at CpG sites accounted for 24% of all observed point mutations (similar to the 17% and 24% reported in humans and chimpanzees, respectively; Kong et al. 2012; Venn et al. 2014). CpG sites are more prone to mutation than other sites due to spontaneous deamination (Bird 1980). We found a significant, but weaker association between paternal age and the per-generation CpG mutation rate ($R^2 = 0.27$, $p = 0.03$; Figure 1). This contrast with the overall per-generation rate is consistent with a non-replicative mechanism for the accumulation of mutations at these sites. It also suggests that with a larger sample size we would likely observe a weak association between maternal age and the mutation rate.

The mutation spectrum in rhesus macaques is similar to that found in other primates, except for a slight excess of C→T transitions (Figure 2). C→T transitions accounted for 53% of all observed *de novo* mutations, significantly higher than the 41% found in human (χ^2 test, $p < 5 \times 10^{-3}$; Kong et al. 2012). This excess in C→T transitions results in a higher transition-to-transversion ratio (2.79) than observed in humans (2.10). This excess also contributes to a higher mutation bias towards weak base pairs (A:T) versus strong base pairs (G:C) when compared to humans (macaque: 3.22, human: 2.15).

A lower mutation rate in the rhesus macaque

The average paternal age at conception explains most of the variation among studies in direct estimates of the human mutation rate (Kong et al. 2012; Rahbari et al. 2016; Jónsson et al. 2017). A meaningful comparison of mutation rates between species must therefore also consider variation in the paternal age at conception among sampled individuals. Our overall estimate of the per-generation mutation rate in rhesus monkeys (0.37×10^{-8} per site per generation) is lower than direct estimates from other primates—1.3×10^{-8} in humans (Jónsson et al. 2017), 1.2×10^{-8} in chimpanzees (Venn et al. 2014), 0.81×10^{-8} in owl monkeys (Thomas et al. 2018)—but both the average age of parents and the average age at puberty differs among species. To compare the estimate from rhesus macaques to other species while accounting for different average ages of reproduction, we considered a model of reproductive longevity (Thomas et al. 2018).

The model of reproductive longevity extends the predictive power of paternal age across species by incorporating different ages of puberty. Reproductive longevity is defined here as the amount of time a parent is in a reproductive state prior to offspring conception: here, we use the paternal age at conception minus the age at sexual maturity (i.e. the age at puberty). The main biological assumption of this model is that reproductive longevity between species can be compared directly, as the number of mutations within each species is largely driven by the accumulation of

replication-dependent errors after puberty. The mutation rate can evolve between species if the rate of replication-dependent errors changes, if the rate of germline cell-division post-puberty changes, or if there are different numbers of mutations that accumulate prior to puberty (Thomas and Hahn 2014; Thomas et al. 2018). This model predicts that differences in reproductive timing between species, specifically the beginning of spermatogenesis and average time of conception, should explain much of the observed variation in mutation rate between species—much as paternal age explains variation within species. Though key parameters of spermatogenesis surely differ between species, such simple models make compelling null models for understanding mutation rate variation.

We compared the rate of mutation accumulation with paternal age in the macaque to the rate observed in humans (Jónsson et al. 2017). If we assume both that the number of mutations before puberty and the rate of accumulation of mutations after puberty are the same in humans as in macaques, we overestimate the expected number of mutations per generation (approx. 54 vs. 23, using 7.5 years as the average paternal age in macaques). In order to investigate the source of this difference, we compared a linear regression of mutation accumulation with paternal age from the beginning of sexual maturity in the two species.

Much of the difference in the per-generation mutation rate between human and macaque can be attributed to the number of mutations present in the germline pre-puberty (Figure 3): there are fewer than half as many in macaques as in humans (17.6, 95% CI: [14.1, 21.0] in macaque and 43.2 [41.8, 44.6] in human, based on estimates of the regression intercepts with the respective timing of puberty). The rate at which mutations increase with paternal age after puberty is slightly lower in the macaque (2.2×10^{-10} per bp, per year, 95% CI [1.5, 2.9]) relative to human (3.3×10^{-10} per bp, per year, [3.2, 3.4]; equal slopes test, $p = 0.005$). Though statistically significant, this corresponds to only ~3 fewer mutations over the average lifespan of a macaque.

The evolutionary changes leading to a lower number of mutations before puberty appear to have occurred along the branch leading to macaques, as owl monkeys show a similar number of mutations pre-puberty as humans (Thomas et al. 2018). The smaller number of mutations in macaques may be due to a decreased number of germline cell divisions or a decreased rate of mutation per cell division before puberty. Whatever the case, the data here suggest that the mutation rate per cell division post-puberty has had little effect on the overall difference in the per-generation rate between species.

Sociability in male rhesus monkeys show no connection with sire age

Sociability is a consistent personality dimension in humans that has also been identified in rhesus monkeys (Capitanio 1999). Low social ability in infant rhesus monkeys has been shown to predict poor adult social function (Sclafani et al. 2016), consistent with deficits in childhood social interaction and communication as risk factors for the development of autism spectrum

disorder in humans (Ozonoff et al. 2010; Jones et al. 2014). We examined sociability across a sample of 203 male monkeys studied in adulthood to determine whether paternal age at conception was a significant contributor to low social function. These monkeys came from the same colony as those used for sequencing, but none of the individuals were the same.

In addition to sociability, we measured the frequency of eight behaviors associated with general social functioning, stratified by sex (Supplemental Table 1). We performed principal component analysis on these variables to reduce dimensionality and to extract a useful general score of social functioning from these behaviors. The first two principal components (PC) explain > 94% of the variance in these observations. Offspring social behavior PC1 explains the tendency for behaviors to be directed towards females versus males, while offspring social behavior PC2 explains overall contact and proximity with both sexes. Social behavior PC2 was significantly correlated with observer ratings of the sociability personality measure (Pearson's $r = 0.37$, $p < 5 \times 10^{-9}$).

We found no evidence for a relationship between paternal age and any measure of lowered social function (Figure 4). Rather than a negative effect on sociability, there was a weakly positive correlation suggested between sociability and parental age at conception (sire age: $r = 0.07$, $p = 0.21$; dam age: $r = 0.02$, $p = 0.09$). Because of the high correlation between sire rank and sire age, we also calculated pairwise partial correlations between sire age and all measures of social functioning while attempting to control for sire rank. None of these correlations were significant (Supplemental Table 2).

Discussion

Both the rate and the spectrum of mutations are intimately linked with life history (Walter et al. 2004; Goldmann et al. 2016; Rahbari et al. 2016), complicating comparisons across studies that report point estimates. We discovered a significant paternal age effect on mutation rate in rhesus macaques, consistent with findings from other direct estimates of the mutation rate in primates. This linear increase in the number of mutations with paternal age is consistent with replication-dependent mutagenesis driven by successive rounds of spermatogenesis. The overall per-generation mutation rate in the macaque is substantially lower than has been found in other primates. Our analysis indicates that this is largely due to a lower number of mutations before puberty, with little effect from differences in the rate at which mutations accumulate with paternal age after puberty.

The parental ages at conception from our sample are somewhat clustered around the age of puberty (Fig. 1), a fact that may in part explain the relative preponderance of mutations at CpG sites in our sample compared to other primates. The proportion of all mutations transmitted by fathers that occur at CpG sites is expected to decline with age, as they are not thought to be

primarily the result of replicative error. Both the slow rate of mutation accumulation at CpG sites with paternal age, and the positive (though not statistically significant) accumulation of mutations with maternal age, suggest a limited role for non-replicative mechanisms of mutation accumulation in the macaque. Coupled with the high rate of replicative errors that accumulate with paternal age, the ratio of mutations of replicative versus non-replicative origin is roughly constant, as has been observed in humans (Gao et al. 2019).

With the paternal age effect explaining much of the variation in mutation rate within species, differences in key parameters of spermatogenesis—including timing, length, and efficiency—have become an important part of studies seeking to explain variation in mutation and substitution rates between species (Wilson Sayres et al. 2011; Thomas and Hahn 2014; Amster and Sella 2016; Moorjani et al. 2016; Scally 2016). For instance, it is known that the seminiferous epithelial cycle time is 34% shorter in the macaque than in humans (de Rooij et al. 1986). All things being equal, this suggests that male macaques should accumulate mutations post-puberty at a higher rate than male humans. However, our results reveal little difference between macaques and humans in how mutation accumulation scales with paternal age. While it is possible that other details of spermatogenesis have evolved in a way to exactly cancel out this increase, a more straightforward explanation is that the cycling time of active cells has little effect on how fast mutations accumulate. This hypothesis is also consistent with the disconnect between the male-to-female ratio of germline divisions under the assumption of constant spermatogonial division and the ratio of X-to-autosome substitutions from phylogenetic data (Wilson Sayres and Makova 2011; Ségurel et al. 2014; Gao et al. 2016; Scally 2016). It is more likely that only a fraction of spermatogonial stem cells are actively dividing at any one time, leading to fewer effective cycles per unit time.

The biggest difference we find between species is the number of mutations present at puberty, before active spermatogenesis begins. In humans, the number of mutations that accumulate before puberty is significantly different between the sexes (Goldmann et al. 2016; Jónsson et al. 2017; Gao et al. 2019), suggesting that this trait can evolve. It is not clear, however, what changes have occurred before puberty to lower the mutation rate. Though our data suggest that the mutation rate per cell division post-puberty is the same between species, it is possible that there are differences between species in the error-prone divisions of early embryogenesis (Huang et al. 2014; Rahbari et al. 2016; Ju et al. 2017). Under such a model, the decreased number of mutations before puberty in the macaque may be explained by a lower number of embryonic mutations, a process that is not modeled well by mutation rates during spermatogenesis. In any case, the evolution of life-history appears to have played a large role in shaping the differences in the per-generation mutation rate between human and macaque.

Previous studies have found that both the number of *de novo* mutations and the risk of neurodevelopmental disorders increase with paternal age in humans (Kong et al. 2012). We find no link between paternal age and negative social behavioral outcomes in offspring, despite an

increasing number of mutations with paternal age in the rhesus macaque. We must acknowledge that social behavior is a complex human construct that our assay is unlikely to fully capture. Furthermore, differences in sociability are only a single component in the complex syndromes that constitute neurodevelopmental disorder. Nevertheless, our findings are consistent with the hypothesis that the increased risk of neurodevelopmental disorder with paternal age in humans is not primarily driven by *de novo* mutations, a position with both epidemiological and population genetic support (Hultman et al. 2011; Gratten et al. 2016; Janecka et al. 2017). While this hypothesis does not exclude a role for inherited genetic factors in the development of such disorders, there is no direct role for the elevated *de novo* mutation rate in the higher risk of disorders observed in offspring of older fathers.

Methods

Sequencing

Genomic DNA isolated from blood samples (buffy coats) was used to perform whole genome sequencing. We generated standard PCR-free Illumina paired-end sequencing libraries using KAPA Hyper PCR-free library reagents (KK8505, KAPA Biosystems) in Beckman robotic workstations (Biomek FX and FXp models). We sheared total genomic DNA (500 ng) into fragments of approximately 200-600 bp using the Covaris E220 system (96-well format) followed by purification of the fragmented DNA using AMPure XP beads. We next employed a double size selection step with different ratios of AMPure XP beads to select a narrow size band of sheared DNA molecules for library preparation. DNA end-repair and 3′-adenylation were performed in the same reaction followed by ligation of the barcoded adaptors to create PCR-Free libraries. We ran each library on the Fragment Analyzer (Advanced Analytical Technologies, Ames, Iowa) to assess library size and presence of remaining adaptor dimers. This was followed by qPCR assay using KAPA Library Quantification Kit and their SYBR FAST qPCR Master Mix to estimate the size and quantification. These WGS libraries were sequenced on an Illumina HiSeq-X instrument to generate 150 bp paired-end reads. All flow cell data (BCL files) are converted to barcoded FASTQ files.

Mapping and variant calling

BWA-MEM version 0.7.12-r1039 (Li 2013) was used to align the Illumina sequencing reads to the rhesus macaque reference assembly Mmul_8.0.1 (GenBank accession GCA_000772875.3) and to generate BAM files for each of the 32 individuals. Picard MarkDuplicates version 1.105 (Broad Institute 2019) was used to identify and mark duplicate reads. Single nucleotide variants (SNVs) and small indels (up to 60 bp) were called using GATK version 3.6 (Van der Auwera et al. 2013) following best practices. We used HaplotypeCaller to generate gVCFs for each sample,

and performed joint genotype calling on all samples using GenotypeGVCFs, generating a VCF file. GATK hard filters (SNPs: "QD < 2.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0"; Indels: "QD < 2.0 || FS > 200.0 || ReadPosRankSum < -20.0") were applied and we removed calls that failed. Finally, we used GATK's PhaseByTransmission to identify Mendelian violations that represent possible *de novo* variants.

Filtering candidate mutations

After removing candidates returned by PhaseByTransmission where a genotype was missing in either the mother or father, our initial list of candidates included 21069 Mendelian violations (MVs). In the next steps, we progressively applied filters that reduced both the number of candidates and the total number of callable sites to improve the accuracy of identifying actual *de novo* mutations. To be specific, we:

Removed candidate sites that had fewer than 20 reads or more than 60 reads. Sites with too few reads have a higher chance of being undersampled, while sites with too many reads may represent problematic repetitive regions (Li 2014).

Remaining candidates: 12046

Removed sites that were not homozygous reference in both parents, or appeared as a variant in an unrelated trio. This latter step reduces the chances that a child inherited a segregating variant that was miscalled as homozygous reference in a parent.

Remaining candidates: 2794

Removed heterozygotes that were called with reads that were more than 65% reference or more than 65% alternate. (i.e., heterozygotes with an allelic balance that was not between 0.35 and 0.65; Figure S6b). These represent candidate MVs that are not well-supported by the distribution of allele counts.

Remaining candidates: 303

We found a single cluster of 4 mutations on chromosome 10 in trio 8 (all within 20 bp), making it a candidate multinucleotide mutation (Schrider et al. 2011). To evaluate the sensitivity of our mutation calls to the allelic balance filter, we re-estimated the mutation rate (with correction, see below) at several filter limits (Figure S5). Based on the stability of the mutation rate estimate and the distribution of allelic balance among mendelian violations (Figure S6b), we selected limits of (0.35, 0.65).

Mutation rate estimates

Raw counts of *de novo* mutations were converted into a mutation rate by dividing by the total number of callable sites and adjusting by estimates of the false negative and false positive rate. Callable sites in each proband, *C*, were restricted to those that passed the same filters as used to call candidate mutations. Almost all of the reduction in callable sites came from limiting the read

count (i.e. the first filter listed above). Across the 14 trios, there was a mean of 2.492×10^9 callable sites.

The false positive rate, α, was estimated by considering the transmission frequency of mutations to a third generation. We observed the percentage of transmitted mutations among probands with a child, which should be 50% if no false positives existed, and calculated α as

$$\alpha = 1 - \frac{\text{transmission frequency}}{0.5}$$

We calculated a transmission frequency of 37.4%, leading to an estimate of 25.3% for the false positive rate.

The false negative rate, β, was estimated by considering the number of heterozygous SNPs with an allelic balance that fell outside of our selected limits of (0.35, 0.65). We expect sites that are truly *de novo* mutations to have the same distribution of allelic balance as all other heterozygous SNPs (Figure S6a). As such, our limits truncate true *de novo* mutations at a rate that can be calculated as

$$\beta = 1 - \frac{\text{SNPs with allelic balance} = (0.35, 0.65)}{\text{total number of SNPs}}$$

We counted 167,264,837 SNPs with an allelic balance within the limits of (0.35, 0.65) out of a total of 188,234,430 SNPs, for a false negative rate of 11.1%.

Our per-generation mutation rate, μ_g, for each trio was then calculated based on the number of observed mutations, N_μ, as

$$\mu_g = \frac{N_\mu(1-\alpha)}{2C(1-\beta)}$$

where $2C$ represents twice the callable haploid genome size. Our overall estimate of the mutation rate is a mean of this per-generation rate across the 14 trios.

Phasing mutations

We traced the parent of origin for *de novo* mutations that were transmitted to the third generation. This was accomplished by tracking their inheritance on haplotype blocks that we assembled from phase-informative sites. These informative sites were biallelic and had genotypes that were different between grandparents, heterozygous in the next generation, and not heterozygous in both the third-generation proband and the other parent. These phase-informative sites could be traced unambiguously to one of the grandparents. Sites were assembled into

haplotype blocks under the assumption that multiple recombination events in a single meiosis were unlikely to occur within a 1-Mb interval (Rogers et al. 2006; Huang et al. 2009; Smeds et al. 2016; Xue et al. 2016).

Mutation rates with paternal age

We compared our estimate of the mutation rate in rhesus macaque to a simple model of the mutation rate in human: a linear regression of the per-generation mutation rate with paternal age from a large human dataset (Jónsson et al. 2017). If the same number of mutations accumulate before puberty in rhesus macaque as in human, and the rate of accumulation after puberty is the same, our sample of macaques should have a mutation rate corresponding to an average 17.5-year-old human. This is based on the mean paternal age of 7.5 years among macaques in our dataset and an age of 3.5 years for male puberty in macaques (Plant et al. 2005), resulting in 4 years of post-puberty mutation accumulation. Together with the 13.5 years to reach male puberty in humans, the corresponding age for the human model becomes 17.5 years (=13.5 + 4). Next, we performed a linear regression of the macaque mutation rate with paternal age, and calculated its intercept with the approximate age of male puberty (3.5 years). We tested whether the slope coefficient in this regression was significantly different from the one found in the human dataset with an equal slopes test. This test statistic was equal to the absolute difference between slopes divided by the pooled standard error. To illustrate the difference in per-generation mutation rates between species, we calculated estimates for the number of mutations in several comparisons. In each of these estimates, we multiplied the predicted rates by twice their respective haploid genome size, as approximated by the UCSC golden path length.

Collecting sociability data from captive rhesus monkeys

We observed 203 male rhesus monkeys, in cohorts of 5 to 8 animals (mean age = 6.9 years, range = 4.0 to 19.2 years), unobtrusively in their half-acre outdoor enclosures across four summers. Observations were conducted over the course of eight days within a two-week period, and consisted of two 10-minute sessions per day. Using focal animal sampling (Altmann 1974), behavioral observers recorded the frequencies of the following behaviors directed at other adult animals: approach (locomotion to within arm's reach), proximity (being within arm's reach for at least three seconds), contact (physical, non-aggressive contact between animals), and grooming (picking with fingers and/or licking another animal's hair). Following completion of behavioral observations on each cohort, the observer rated each animal using a 7-point Likert-type scale on three trait adjectives. Previous work (Capitanio and Widaman 2005) had demonstrated these ratings form a scale that reflected personality characteristics as follows: Sociability—affiliative, agreeable, sociable, appears to like the company of others and seeks out social contact with other animals; Warm—seeks or elicits bodily closeness, touching, grooming; and Solitary—subject prefers to spend considerable time alone, avoids or does not often seek contact with other

animals. Observers were trained to show greater than 85% agreement on coding behavior, and Cronbach's alpha (a measure of scale reliability) was greater than 0.92 for all samples.

Acknowledgments

This work was funded by the Precision Health Initiative of Indiana University. Behavioral assessments were made possible by funding from CNPRC base grant P51 OD011107 and NIH grants R37 AG033590 and R24 OD010962. We would like to acknowledge the production staff of the Human Genome Sequencing Center and its Director, Richard Gibbs.

References

Altmann, J. 1974. Observational study of behavior: Sampling methods. Behaviour 49:227–266.

Amster, G., and G. Sella. 2016. Life history effects on the molecular clock of autosomes and sex chromosomes. Proc. Natl. Acad. Sci. 113:1588–1593.

Bauman, M. D., and C. M. Schumann. 2018. Advances in nonhuman primate models of autism: Integrating neuroscience and behavior. Exp. Neurol. 299:252–265.

Besenbacher, S., C. Hvilsom, T. Marques-Bonet, T. Mailund, and M. H. Schierup. 2019. Direct estimation of mutations in great apes reconciles phylogenetic dating. Nat. Ecol. Evol. 3:286–292.

Besenbacher, S., S. Liu, J. M. G. Izarzugaza, J. Grove, K. Belling, J. Bork-Jensen, S. Huang, T. D. Als, S. Li, R. Yadav, A. Rubio-García, F. Lescai, D. Demontis, J. Rao, W. Ye, T. Mailund, R. M. Friborg, C. N. S. Pedersen, R. Xu, J. Sun, H. Liu, O. Wang, X. Cheng, D. Flores, E. Rydza, K. Rapacki, J. Damm Sørensen, P. Chmura, D. Westergaard, P. Dworzynski, T. I. A. Sørensen, O. Lund, T. Hansen, X. Xu, N. Li, L. Bolund, O. Pedersen, H. Eiberg, A. Krogh, A. D. Børglum, S. Brunak, K. Kristiansen, M. H. Schierup, J. Wang, R. Gupta, P. Villesen, and S. Rasmussen. 2015. Novel variation and *de novo* mutation rates in population-wide *de novo* assembled Danish trios. Nat. Commun. 6:5969.

Bird, A. P. 1980. DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res. 8:1499–1504.

Broad Institute. 2019. Picard Toolkit. Broad Institute, Github Repository. http://broadinstitute.github.io/picard/

Capitanio, J. P. 1999. Personality dimensions in adult male rhesus macaques: Prediction of behaviors across time and situation. Am. J. Primatol. 47:299–320.

Capitanio, J. P., and K. F. Widaman. 2005. Confirmatory factor analysis of personality structure in adult male rhesus monkeys (*Macaca mulatta*). Am. J. Primatol. 65:289–294.

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

de Rooij, D. G., M. M. van Alphen, and H. J. van de Kant. 1986. Duration of the cycle of the seminiferous epithelium and its stages in the rhesus monkey (*Macaca mulatta*). Biol. Reprod. 35:587–591.

Durkin, M. S., M. J. Maenner, C. J. Newschaffer, L.-C. Lee, C. M. Cunniff, J. L. Daniels, R. S. Kirby, L. Leavitt, L. Miller, W. Zahorodny, and L. A. Schieve. 2008. Advanced parental age and the risk of autism spectrum disorder. Am. J. Epidemiol. 168:1268–1276.

Francioli, L. C., P. P. Polak, A. Koren, A. Menelaou, S. Chun, I. Renkens, Genome of the Netherlands Consortium, C. M. van Duijn, M. Swertz, C. Wijmenga, G. van Ommen, P. E. Slagboom, D. I. Boomsma, K. Ye, V. Guryev, P. F. Arndt, W. P. Kloosterman, P. I. W. de Bakker, and S. R. Sunyaev. 2015. Genome-wide patterns and properties of *de novo* mutations in humans. Nat. Genet. 47:822–826.

Frans, E. M., S. Sandin, A. Reichenberg, P. Lichtenstein, N. Långström, and C. M. Hultman. 2008. Advancing paternal age and bipolar disorder. Arch. Gen. Psychiatry 65:1034–1040.

Gao, Z., P. Moorjani, T. A. Sasani, B. S. Pedersen, A. R. Quinlan, L. B. Jorde, G. Amster, and M. Przeworski. 2019. Overlooked roles of DNA damage and maternal age in generating human germline mutations. Proc. Natl. Acad. Sci. 116:9491–9500.

Gao, Z., M. J. Wyman, G. Sella, and M. Przeworski. 2016. Interpreting the dependence of mutation rates on age and time. PLoS Biol. 14:e1002355.

Gil-da-Costa, R., G. R. Stoner, R. Fung, and T. D. Albright. 2013. Nonhuman primate model of schizophrenia using a noninvasive EEG method. Proc. Natl. Acad. Sci. 110:15425–15430.

Glaser, R. L., K. W. Broman, R. L. Schulman, B. Eskenazi, A. J. Wyrobek, and E. W. Jabs. 2003. The paternal-age effect in Apert syndrome is due, in part, to the increased frequency of mutations in sperm. Am. J. Hum. Genet. 73:939–947.

Goldmann, J. M., W. S. W. Wong, M. Pinelli, T. Farrah, D. Bodian, A. B. Stittrich, G. Glusman, L. E. L. M. Vissers, A. Hoischen, J. C. Roach, J. G. Vockley, J. A. Veltman, B. D. Solomon, C. Gilissen, and J. E. Niederhuber. 2016. Parent-of-origin-specific signatures of *de novo* mutations. Nat. Genet. 48:935–939.

Goriely, A., J. J. McGrath, C. M. Hultman, A. O. M. Wilkie, and D. Malaspina. 2013. "Selfish spermatogonial selection": A novel mechanism for the association between advanced paternal age and neurodevelopmental disorders. Am. J. Psychiatry 170:599–608.

Gratten, J., N. R. Wray, W. J. Peyrot, J. J. McGrath, P. M. Visscher, and M. E. Goddard. 2016. Risk of psychiatric illness from advanced paternal age is not predominantly from *de novo* mutations. Nat. Genet. 48:718–724.

Green, R. F., O. Devine, K. S. Crider, R. S. Olney, N. Archer, A. F. Olshan, and S. K. Shapira. 2010. Association of paternal age and risk for major congenital anomalies from the National Birth Defects Prevention Study, 1997 to 2004. Ann. Epidemiol. 20:241–249.

Huang, A. Y., X. Xu, A. Y. Ye, Q. Wu, L. Yan, B. Zhao, X. Yang, Y. He, S. Wang, Z. Zhang, B. Gu, H.-Q. Zhao, M. Wang, H. Gao, G. Gao, Z. Zhang, X. Yang, X. Wu, Y. Zhang, and L. Wei. 2014. Postzygotic single-nucleotide mosaicisms in whole-genome sequences of clinically unremarkable individuals. Cell Res. 24:1311–1327.

Huang, X., Q. Feng, Q. Qian, Q. Zhao, L. Wang, A. Wang, J. Guan, D. Fan, Q. Weng, T. Huang, G. Dong, T. Sang, and B. Han. 2009. High-throughput genotyping by whole-genome resequencing. Genome Res. 19:1068–1076.

Hultman, C. M., S. Sandin, S. Z. Levine, P. Lichtenstein, and A. Reichenberg. 2011. Advancing paternal age and risk of autism: new evidence from a population-based study and a meta-analysis of epidemiological studies. Mol. Psychiatry 16:1203–1212.

Janecka, M., J. Mill, M. A. Basson, A. Goriely, H. Spiers, A. Reichenberg, L. Schalkwyk, and C. Fernandes. 2017. Advanced paternal age effects in neurodevelopmental disorders—review of potential underlying mechanisms. Transl. Psychiatry 7:e1019.

Jones, E. J. H., T. Gliga, R. Bedford, T. Charman, and M. H. Johnson. 2014. Developmental pathways to autism: A review of prospective studies of infants at risk. Neurosci. Biobehav. Rev. 39:1–33.

Jónsson, H., P. Sulem, B. Kehr, S. Kristmundsdottir, F. Zink, E. Hjartarson, M. T. Hardarson, K. E. Hjorleifsson, H. P. Eggertsson, S. A. Gudjonsson, L. D. Ward, G. A. Arnadottir, E. A. Helgason, H. Helgason, A. Gylfason, A. Jonasdottir, A. Jonasdottir, T. Rafnar, M. Frigge, S. N. Stacey, O. T. Magnusson, U. Thorsteinsdottir, G. Masson, A. Kong, B. V. Halldorsson, A. Helgason, D. F. Gudbjartsson, and K. Stefansson. 2017. Parental influence on human germline *de novo* mutations in 1,548 trios from Iceland. Nature 549:519–522.

Ju, Y. S., I. Martincorena, M. Gerstung, M. Petljak, L. B. Alexandrov, R. Rahbari, D. C. Wedge, H. R. Davies, M. Ramakrishna, A. Fullam, S. Martin, C. Alder, N. Patel, S. Gamble, S. O'Meara, D. D. Giri, T. Sauer, S. E. Pinder, C. A. Purdie, Å. Borg, H. Stunnenberg, M. van de Vijver, B. K. T. Tan, C. Caldas, A. Tutt, N. T. Ueno, L. J. van 't Veer, J. W. M. Martens, C. Sotiriou, S. Knappskog, P. N. Span, S. R. Lakhani, J. E. Eyfjörd, A.-L. Børresen-Dale, A. Richardson, A. M. Thompson, A. Viari, M. E. Hurles, S. Nik-Zainal, P. J. Campbell, and M. R. Stratton. 2017. Somatic mutations reveal asymmetric cellular dynamics in the early human embryo. Nature 543:714–718.

Kong, A., M. L. Frigge, G. Masson, S. Besenbacher, P. Sulem, G. Magnusson, S. A. Gudjonsson, A. Sigurdsson, A. Jonasdottir, A. Jonasdottir, W. S. W. Wong, G. Sigurdsson, G. B. Walters, S. Steinberg, H. Helgason, G. Thorleifsson, D. F. Gudbjartsson, A. Helgason, O. T. Magnusson, U. Thorsteinsdottir, and K. Stefansson. 2012. Rate of *de novo* mutations and the importance of father's age to disease risk. Nature 488:471–475.

Lee, H., D. Malaspina, H. Ahn, M. Perrin, M. G. Opler, K. Kleinhaus, S. Harlap, R. Goetz, and D. Antonius. 2011. Paternal age related schizophrenia (PARS): Latent subgroups detected by k-means clustering analysis. Schizophr. Res. 128:143–149.

Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Q-Bio ArXiv13033997.

Li, H. 2014. Toward better understanding of artifacts in variant calling from high-coverage samples. Bioinformatics 30:2843–2851.

Malaspina, D., A. Brown, D. Goetz, N. Alia-Klein, J. Harkavy-Friedman, S. Harlap, and S. Fennig. 2002. Schizophrenia risk and paternal age: A potential role for *de novo* mutations in schizophrenia vulnerability genes. CNS Spectr. 7:26–29.

Menezes, P. R., G. Lewis, F. Rasmussen, S. Zammit, A. Sipos, G. L. Harrison, P. Tynelius, and D. Gunnell. 2010. Paternal and maternal ages at conception and risk of bipolar affective disorder in their offspring. Psychol. Med. 40:477–485.

Moorjani, P., Z. Gao, and M. Przeworski. 2016. Human germline mutation and the erratic evolutionary clock. PLoS Biol. 14:e2000744.

Ozonoff, S., A.-M. Iosif, F. Baguio, I. C. Cook, M. M. Hill, T. Hutman, S. J. Rogers, A. Rozga, S. Sangha, M. Sigman, M. B. Steinfeld, and G. S. Young. 2010. A prospective study of the emergence of early behavioral signs of autism. J. Am. Acad. Child Adolesc. Psychiatry 49:256-266.e1–2.

Parker, K. J., J. P. Garner, O. Oztan, E. R. Tarara, J. Li, V. Sclafani, L. A. D. Rosso, K. Chun, S. W. Berquist, M. G. Chez, S. Partap, A. Y. Hardan, E. H. Sherr, and J. P. Capitanio. 2018. Arginine vasopressin in cerebrospinal fluid is a marker of sociality in nonhuman primates. Sci. Transl. Med. 10:eaam9100.

Plant, T. M., S. Ramaswamy, D. Simorangkir, and G. R. Marshall. 2005. Postnatal and pubertal development of the rhesus monkey (*Macaca mulatta*) testis. Ann. N. Y. Acad. Sci. 1061:149–162.

Rahbari, R., A. Wuster, S. J. Lindsay, R. J. Hardwick, L. B. Alexandrov, S. Al Turki, A. Dominiczak, A. Morris, D. Porteous, B. Smith, M. R. Stratton, Uk10k Consortium, and M. E. Hurles. 2016. Timing, rates and spectra of human germline mutation. Nat. Genet. 48:126–133.

Reichenberg, A., R. Gross, M. Weiser, M. Bresnahan, J. Silverman, S. Harlap, J. Rabinowitz, C. Shulman, D. Malaspina, G. Lubin, H. Y. Knobler, M. Davidson, and E. Susser. 2006. Advancing paternal age and autism. Arch. Gen. Psychiatry 63:1026–1032.

Rogers, J., R. Garcia, W. Shelledy, J. Kaplan, A. Arya, Z. Johnson, M. Bergstrom, L. Novakowski, P. Nair, A. Vinson, D. Newman, G. Heckman, and J. Cameron. 2006. An initial genetic linkage map of the rhesus macaque (*Macaca mulatta*) genome using human microsatellite loci. Genomics 87:30–38.

Ronemus, M., I. Iossifov, Dan Levy, and M. Wigler. 2014. The role of *de novo* mutations in the genetics of autism spectrum disorders. Nat. Rev. Genet. 15:133–141.

Scally, A. 2016. Mutation rates and the evolution of germline structure. Philos. Trans. R. Soc. B Biol. Sci. 371:20150137.

Schrider, D. R., J. N. Hourmozdi, and M. W. Hahn. 2011. Pervasive multinucleotide mutational events in eukaryotes. Curr. Biol. 21:1051–1054.

Sclafani, V., L. A. D. Rosso, S. K. Seil, L. A. Calonder, J. E. Madrid, K. J. Bone, E. H. Sherr, J. P. Garner, J. P. Capitanio, and K. J. Parker. 2016. Early predictors of impaired social functioning in male rhesus macaques (*Macaca mulatta*). PLoS ONE 11:e0165401.

Ségurel, L., M. J. Wyman, and M. Przeworski. 2014. Determinants of Mutation Rate Variation in the Human Germline. Annu. Rev. Genomics Hum. Genet. 15:47–70.

Sipos, A., F. Rasmussen, G. Harrison, P. Tynelius, G. Lewis, D. A. Leon, and D. Gunnell. 2004. Paternal age and schizophrenia: A population based cohort study. BMJ 329:1070.

Smeds, L., C. F. Mugal, A. Qvarnström, and H. Ellegren. 2016. High-resolution mapping of crossover and non-crossover recombination events by whole-genome re-sequencing of an avian pedigree. PLoS Genet. 12:e1006044.

Thomas, G. W. C., and M. W. Hahn. 2014. The human mutation rate is increasing, even as it slows. Mol. Biol. Evol. 31:253–257.

Thomas, G. W. C., R. J. Wang, A. Puri, R. A. Harris, M. Raveendran, D. S. T. Hughes, S. C. Murali, L. E. Williams, H. Doddapaneni, D. M. Muzny, R. A. Gibbs, C. R. Abee, M. R. Galinski, K. C. Worley, J. Rogers, P. Radivojac, and M. W. Hahn. 2018. Reproductive longevity predicts mutation rates in primates. Curr. Biol. 28:3193–3197.

Van der Auwera, G. A., M. O. Carneiro, C. Hartl, R. Poplin, G. del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, E. Banks, K. V. Garimella, D. Altshuler, S.

Gabriel, and M. A. DePristo. 2013. From FastQ data to high-confidence variant calls: The Genome Analysis Toolkit best practices pipeline. Curr. Protoc. Bioinforma. 43:11.10.1-11.10.33.

Venn, O., I. Turner, I. Mathieson, N. de Groot, R. Bontrop, and G. McVean. 2014. Strong male bias drives germline mutation in chimpanzees. Science 344:1272–1275.

Walter, C. A., G. W. Intano, C. A. McMahan, K. Kelner, J. R. McCarrey, and R. B. Walter. 2004. Mutation spectral changes in spermatogenic cells obtained from old mice. DNA Repair 3:495–504.

Wilson Sayres, M. A., and K. D. Makova. 2011. Genome analyses substantiate male mutation bias in many species. BioEssays 33:938–945.

Wilson Sayres, M. A., C. Venditti, M. Pagel, and K. D. Makova. 2011. Do variations in substitution sates and male mutation bias correlate with life-history traits? A study of 32 mammalian genomes. Evolution 65:2800–2815.

Xue, C., M. Raveendran, R. A. Harris, G. L. Fawcett, X. Liu, S. White, M. Dahdouli, D. R. Deiros, J. E. Below, W. Salerno, L. Cox, G. Fan, B. Ferguson, J. Horvath, Z. Johnson, S. Kanthaswamy, H. M. Kubisch, D. Liu, M. Platt, D. G. Smith, B. Sun, E. J. Vallender, F. Wang, R. W. Wiseman, R. Chen, D. M. Muzny, R. A. Gibbs, F. Yu, and J. Rogers. 2016. The population genomics of rhesus macaques (*Macaca mulatta*) based on whole-genome sequences. Genome Res. 26:1651–1662.

Figures and Tables

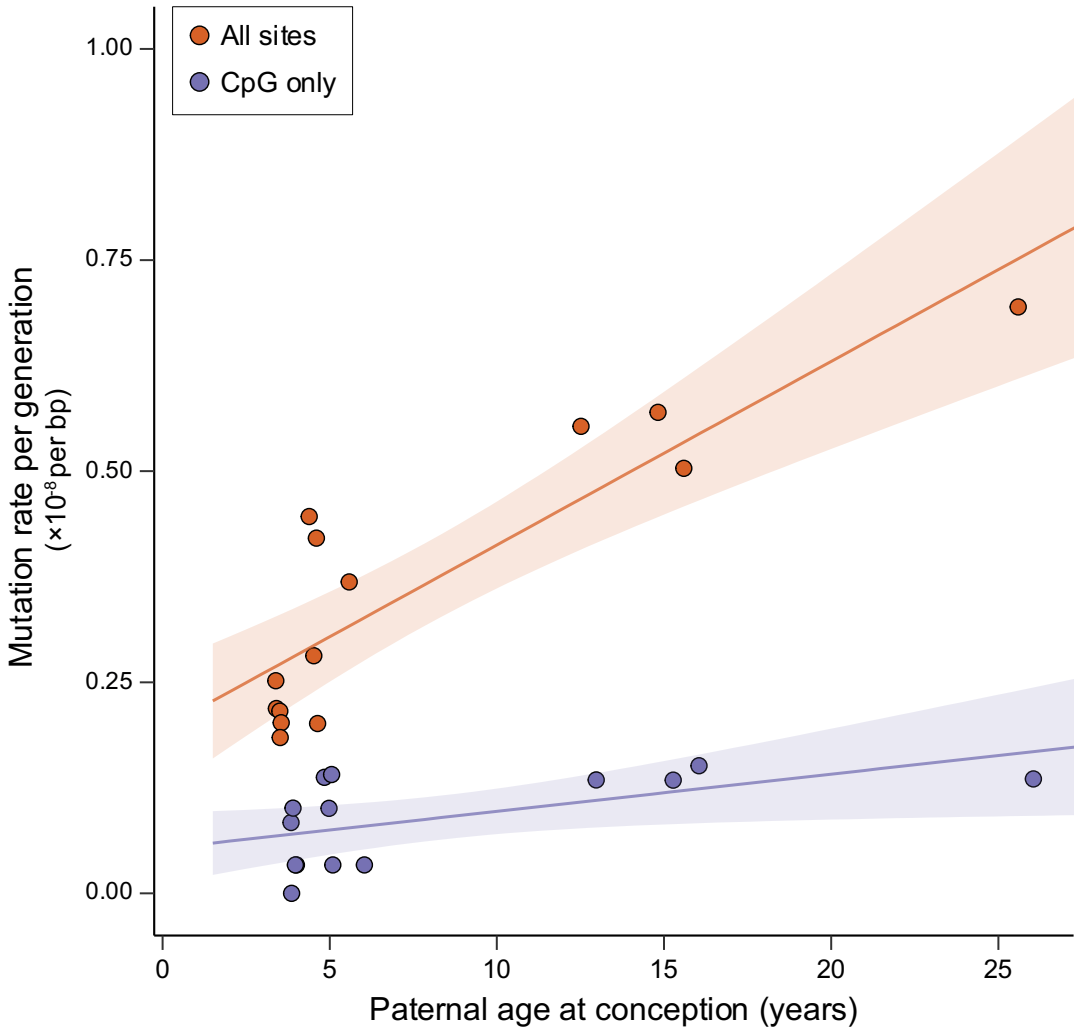

Figure 1. Mutation rates and paternal age in rhesus macaque

Estimates of the overall mutation rate from 14 independent rhesus macaque trios (in orange), and the mutation rate from CpG sites (in purple) in those same trios. There is a strong linear relationship between overall mutation rate and paternal age at conception ($R^2 = 0.77$, $p < 5 \times 10^{-5}$; regression line in orange with shaded area showing 95% CI). This relationship is consistent with the model of accumulating replicative errors in the male germline. The relationship with age is not as pronounced for CpG mutations, which typically accumulate by a non-replicative mechanism ($R^2 = 0.27$, $p = 0.03$; regression in purple with 95% CI shaded).

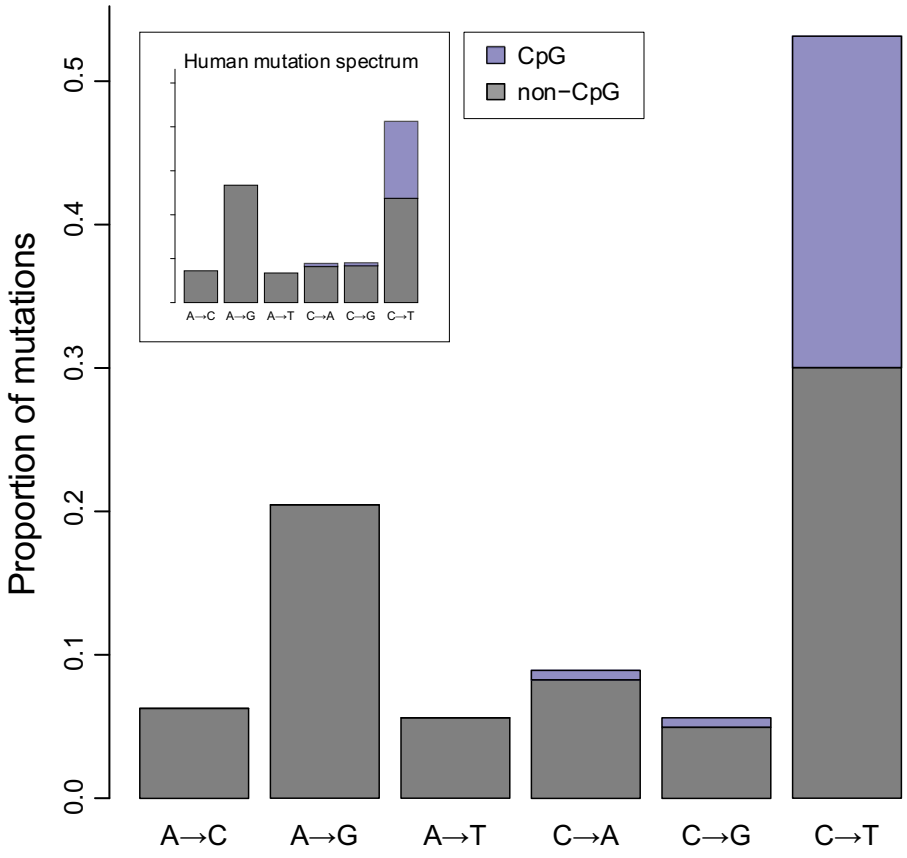

Figure 2. Mutation spectrum in rhesus macaque

The frequency of each type of mutation from among the 303 identified. Mutations at CpG sites accounted for 24% of all mutations and represent 43% of all strong-to-weak transitions. Mutation categories represent their reverse complement as well.

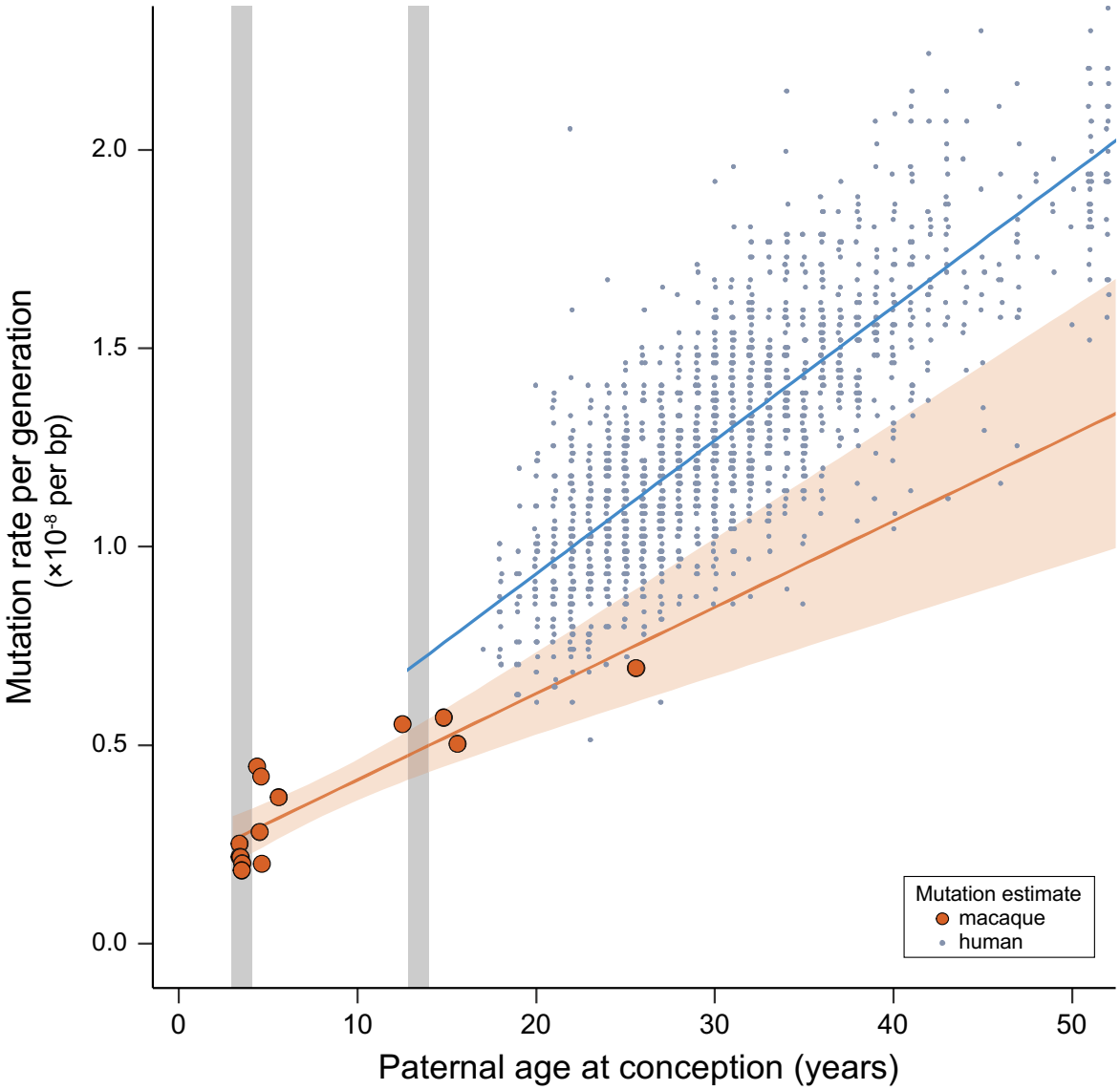

Figure 3. Similar rates of mutation accumulation post-puberty in human and rhesus macaque. Mutation rate accumulation with paternal age estimated from trios in macaques (orange) and humans (blue, from Jónsson et al. (2017)). Approximate age at male puberty in the macaque (3.5 y) and human (13.5 y) are shown in grey. Human trios with paternal age up to 50 are shown here, but the regression line is from the full dataset. The rate at which the mutation rate increases with paternal age is lower in the macaque (2.2×10^{-10} per bp, per year) than in human (3.3×10^{-10} per bp, per year). The intercept with puberty is much lower in macaque (2.7×10^{-9} per bp, per year) than in human (7.0×10^{-9} per bp, per year).

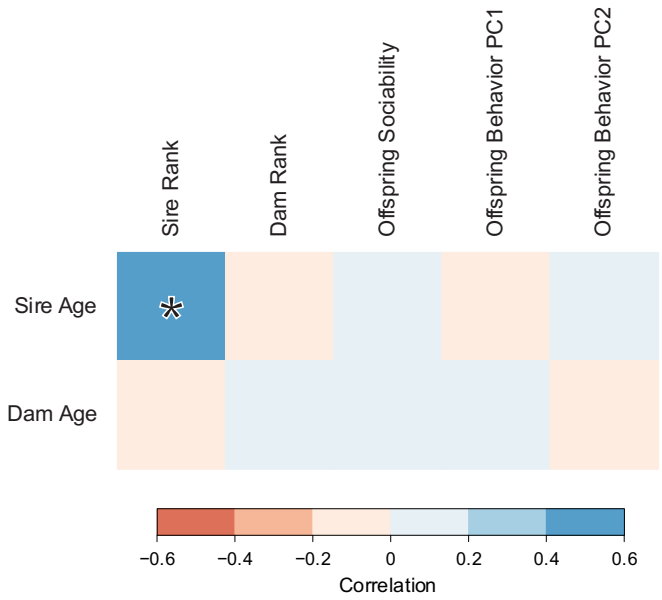

Figure 4. Correlations between parental age and behavioral traits in male rhesus monkeys. Boxes are shaded by the intensity of correlation in pairwise comparisons. Legend shows range of Pearson's correlation coefficient for each color. Significant correlation ($p < 0.05$) highlighted with an asterisk.

Supplemental Table 1. Loadings for first four principal components of behavioral observations

Parameter[a]	PC1	PC2	PC3	PC4
API(M)	-0.02	0.02	-0.01	-0.02
API(F)	0.02	0.01	0.03	-0.02
CON(M)	-0.30	0.49	0.01	0.65
CON(F)	0.44	0.31	-0.81	0.10
PRX(M)	-0.46	0.66	0.01	-0.51
PRX(F)	0.70	0.46	0.54	-0.04
GRI(M)	-0.08	0.12	-0.02	0.22
GRI(F)	0.07	0.04	-0.20	-0.51

[a]Abbreviations,

API approaches initiated to males(M) and females(F)
CON contact with males and females
PRX proximity to males and females
GRI grooming initiated to males and females

Supplemental Table 2. Partial correlation of offspring sociability and sire age while controlling for sire rank

Measure of social function	Pearson's *r*	*p*-value
Sociability	0.06	0.49
Behavioral PC1	-0.08	0.31
Behavioral PC2	0.04	0.64
Behavioral PC3	-0.07	0.37
Behavioral PC4	0.09	0.27

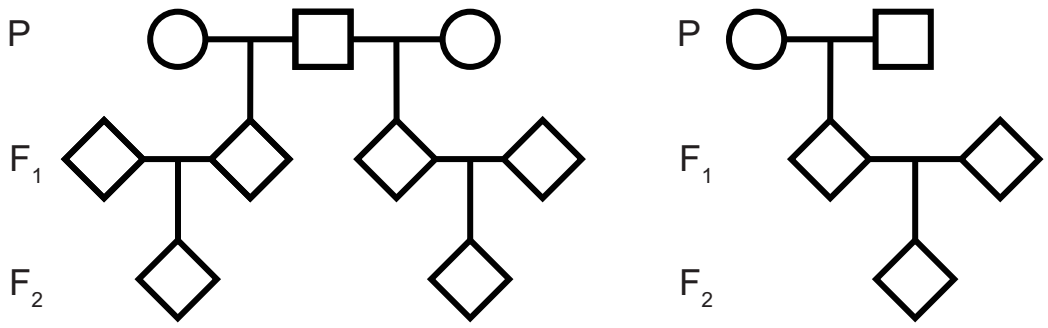

Figure S1. Pedigree structure of sequenced macaques

The 32 individuals sampled in this study were each part of the three-generation family structures depicted above. We sampled individuals that formed three pedigrees of the type shown on the left, and one of the type shown on the right.

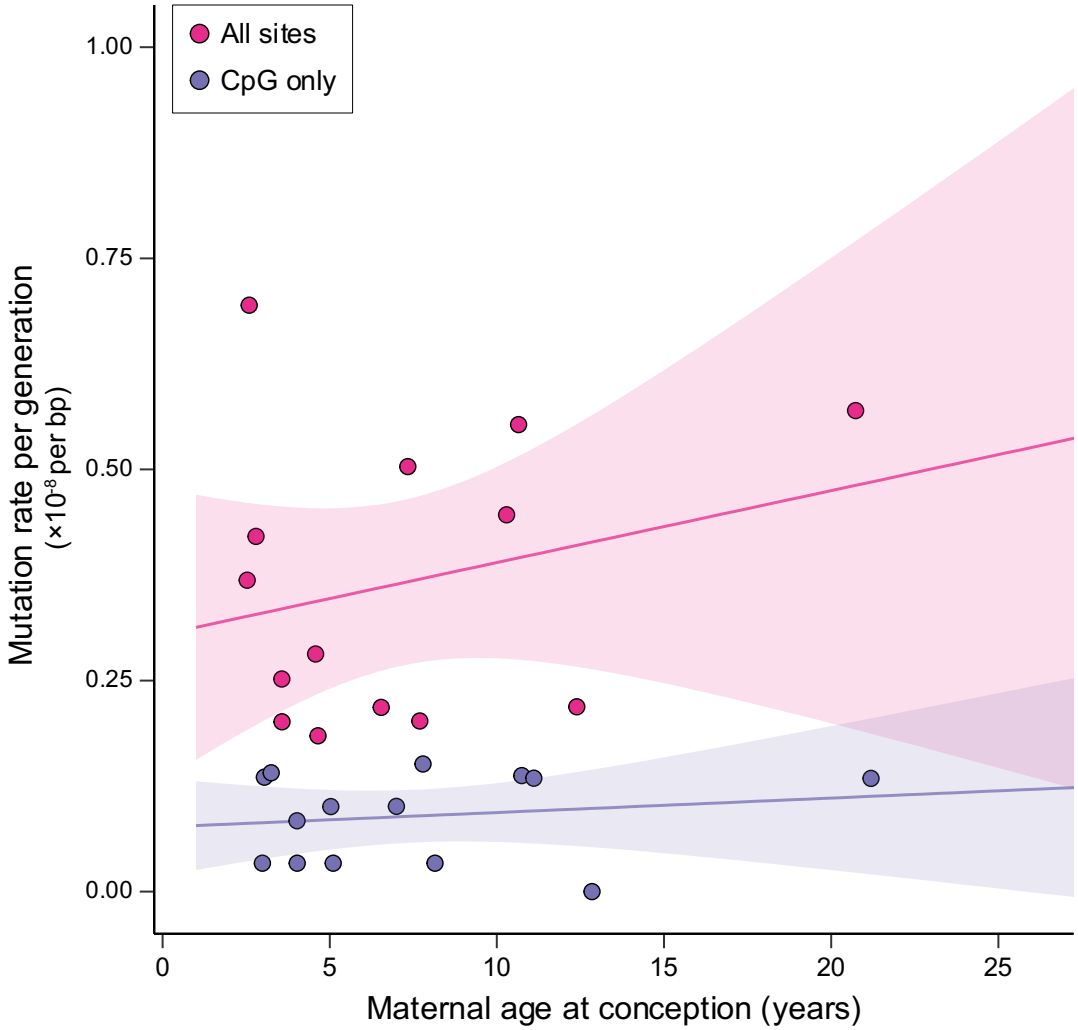

Figure S2. Mutation rates and maternal age in the rhesus macaque

Estimates of the overall mutation rate from 14 independent rhesus macaque trios (in magenta), and the mutation rate from CpG sites (in purple) in those same trios. No significant relationship exists between overall mutation rate and maternal age at conception ($R^2 = 0.07$, $p = 0.37$; regression line in magenta with shaded area showing 95% CI). Similarly, the rate of CpG mutations was not significantly associated with maternal age ($R^2 = 0.07$, $p = 0.57$; regression in purple with 95% CI shaded).

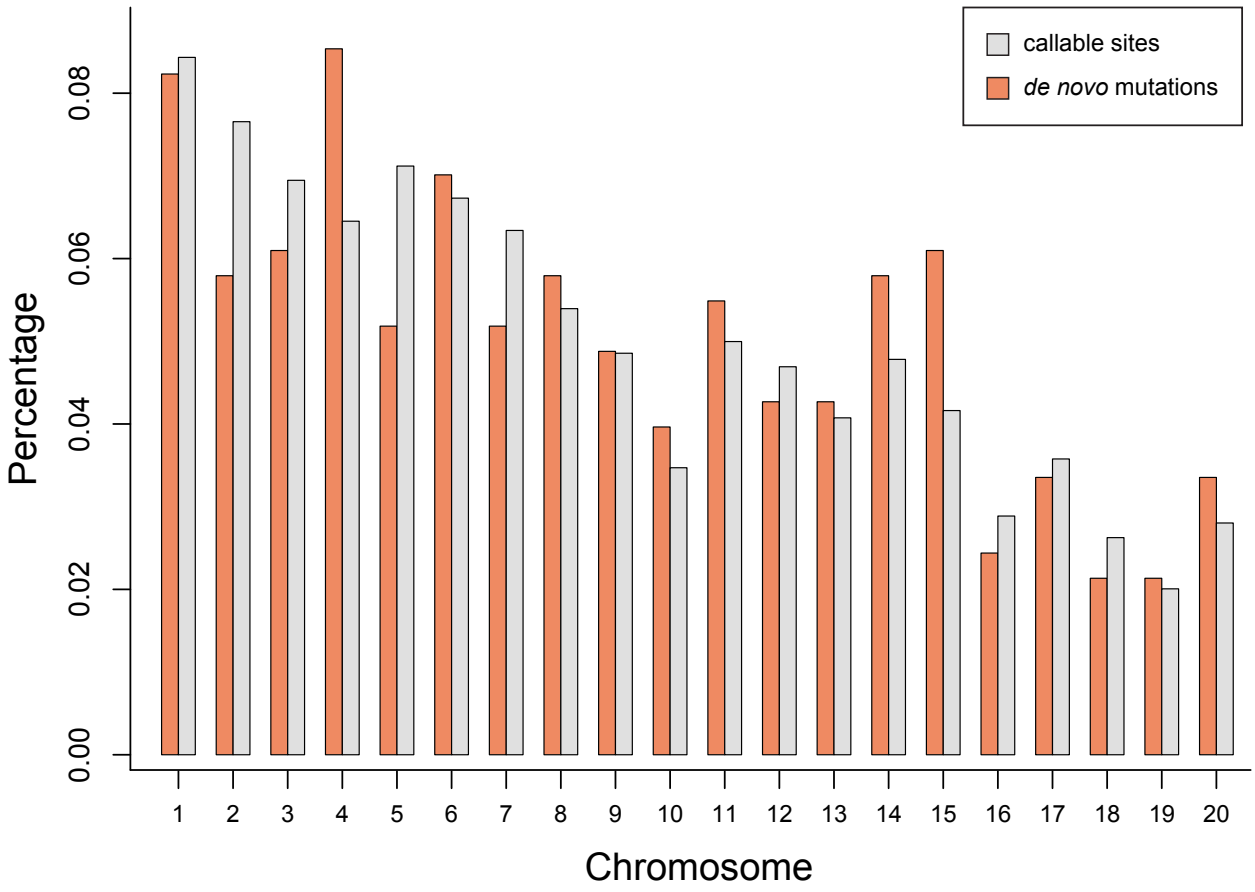

Figure S3. Percentage of *de novo* mutations identified on each autosome

The percentage of *de novo* mutations identified on each autosome, in orange, compared to the percentage of callable sites, in gray. Variation in frequency of mutations across chromosomes does not significantly differ from that expected from chance alone (χ^2 test, $p = 0.28$).

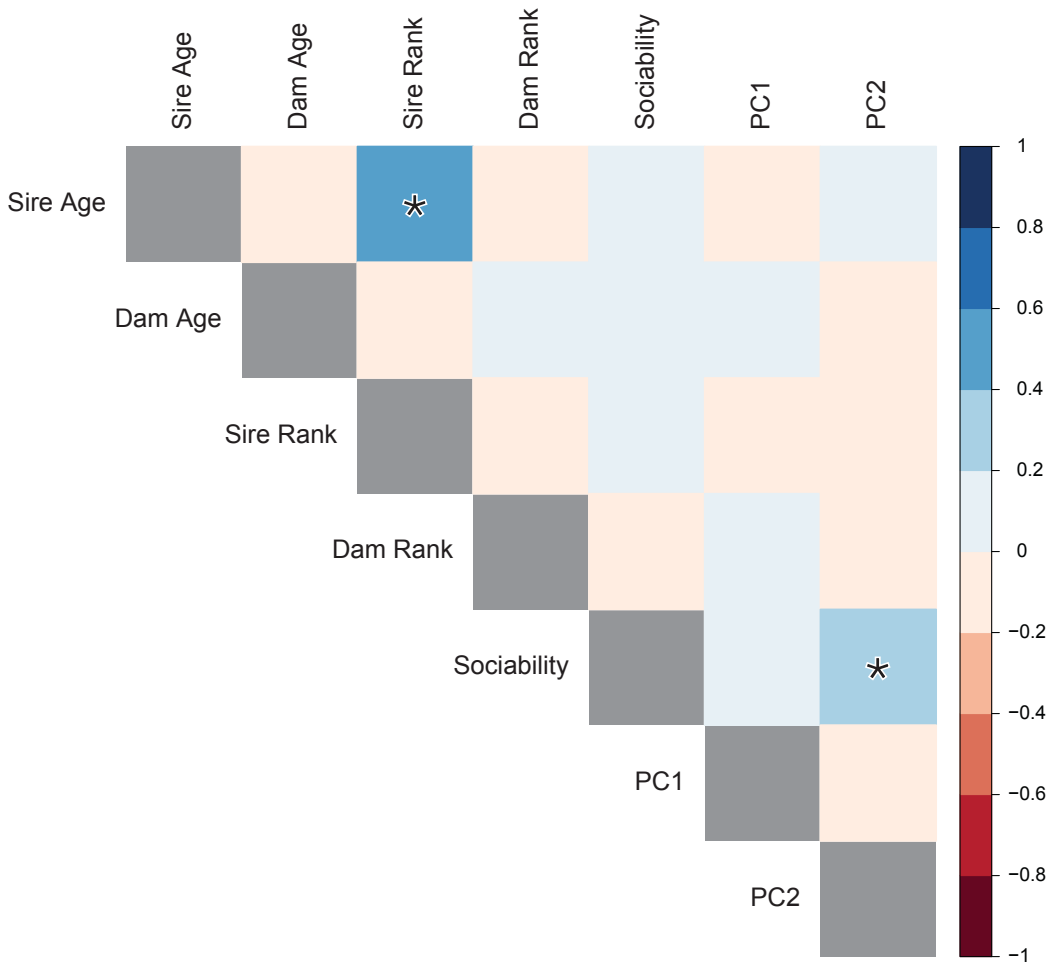

Figure S4. Correlations between parental age and behavioral traits in male rhesus monkeys. Boxes are shaded by the intensity of correlation in pairwise comparisons between each row and column. Diagonal values have been omitted. Legend on the right shows range of Pearson's correlation coefficient for each color. Significant correlations ($p < 0.05$) are highlighted with an asterisk.

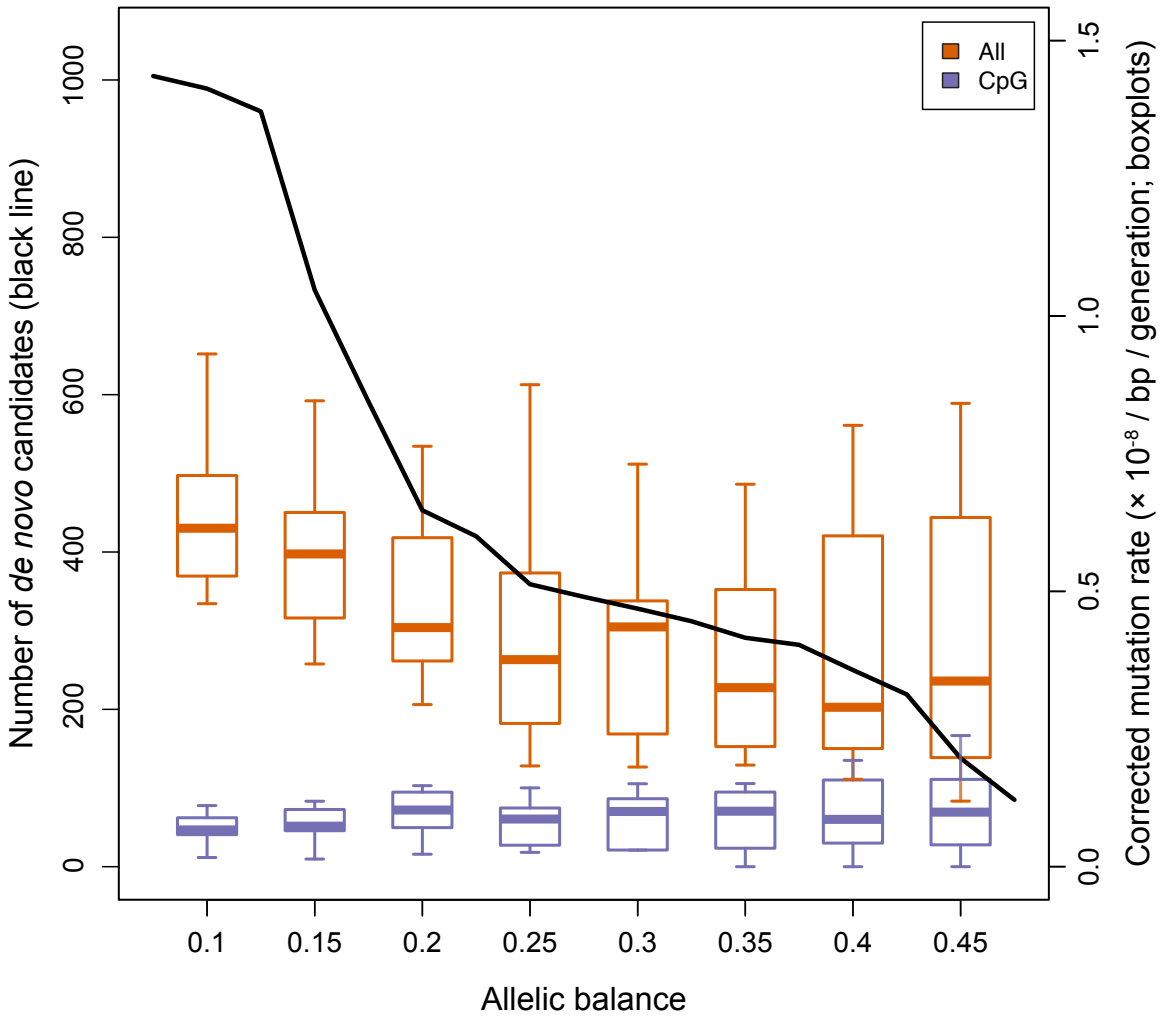

Figure S5. Corrected mutation rates and the number of *de novo* mutation candidates with increasing stringency of allelic balance filter.
Each box-whisker shows the estimated mutation rate after correcting for allelic balance and observed transmission rates across trios. More stringent filtering by allelic balance reduces the total number of *de novo* candidates (black line), but mutation rates are stabilized after correction. The mutation rate when considering only CpG sites (purple) remain unchanged with increasing filter stringency.

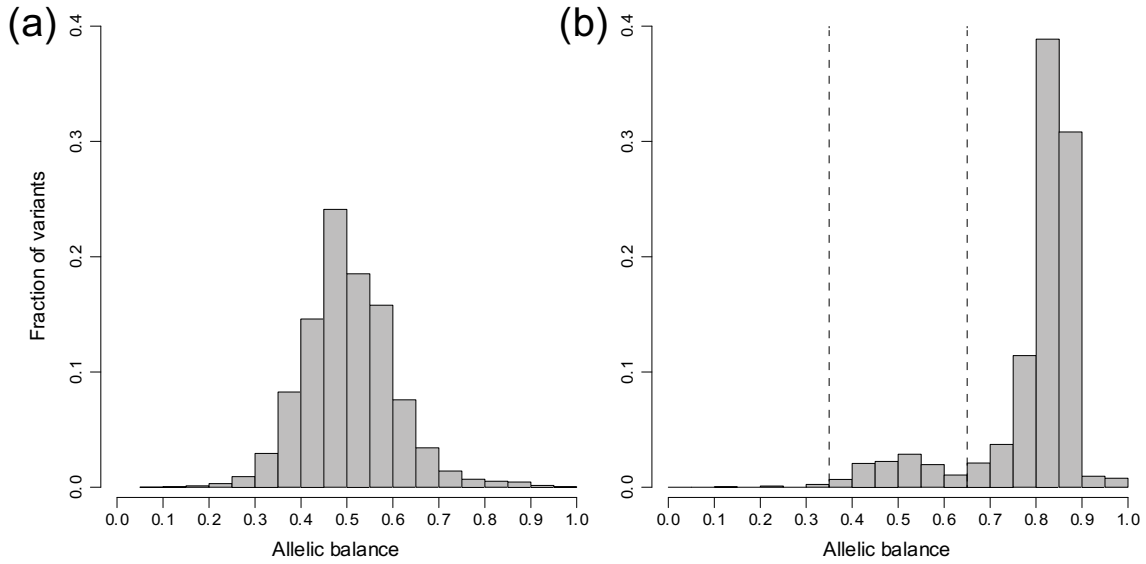

Figure S6. (a) Distribution of allelic balance at all heterozygous SNP sites versus (b) distribution at mendelian violations before applying the allelic balance filter. Dashed lines indicate cutoffs for calling de novo variants (0.35, 0.65).