Overlooked roles of DNA damage and maternal age in generating human germline mutations

Ziyue Gao[1,*], Priya Moorjani[2,3], Guy Amster[4,†] and Molly Przeworski[4,5, †,*]

[1] Howard Hughes Medical Institute & Department of Genetics, Stanford University.

[2] Department of Molecular and Cell Biology, University of California, Berkeley.

[3] Center for Computational Biology, University of California, Berkeley.

[4] Department of Biological Sciences, Columbia University.

[5] Department of Systems Biology, Columbia University.

† Contributed equally

[*] Correspondence to: ziyuegao@stanford.edu and mp3284@columbia.edu

Abstract

Human germline mutations are primarily paternal in origin and their total number increases linearly with the age of the father, observations that are thought to support the textbook view that germline point mutations stem primarily from DNA replication errors. Analyzing large germline mutation datasets for humans, we uncovered two lines of evidence that call this understanding into question. First, despite the drastic increase in the ratio of male to female germ cell divisions after the onset of spermatogenesis, the ratio of paternal to maternal mutations is already 3:1 by puberty and barely increases with the age of the parents, pointing to a substantial contribution of damage-induced mutations. Supporting this hypothesis, C to G transversions and CpG transitions, which together constitute about a 1/3 of point mutations, show sex-specific age dependencies indicative of double-strand break repair and methylation-associated damage, respectively. Second, the age of a mother influences not only the number of mutations that her child inherits on the maternal genome, but also the number of mutations on the paternal genome, as expected if children of older mothers accumulate more mutations in embryogenesis. Together, these findings reveal that, rather than arising predominantly from pre-zygotic replication errors, the parental age effects on germline mutations reflect a combination of higher damage rates in males and a maternal age effect on early development.

One Sentence Summary: Analyses of human de novo mutation datasets reveal overlooked roles of DNA damage and maternal age in generating human germline mutations.

Despite the fundamental importance of germline mutations as the sources of heritable diseases and drivers of evolution, their genesis remains poorly understood (Huttley, Jakobsen, Wilson, & Easteal, 2000; Kumar & Subramanian, 2002). In principle, *de novo* point mutations could arise either directly from errors made while copying an intact DNA template (i.e., be "replication driven") or from damage of the template or free nucleotides that occurred before DNA replication (be "damage-induced"). Characterizing the relative importance of these two sources is of inherent interest and carries many implications, including for understanding the erratic behavior of the molecular clock used to date evolutionary events (Hwang & Green, 2004; Sayres, Venditti, Pagel, & Makova, 2011; Wu & Li, 1985); for the nature of selection pressures on replication and repair machinery (Lynch, 2010; Lynch et al., 2016); and in humans, for predicting recurrence risks of Mendelian diseases and disease burdens (Acuna-Hidalgo, Veltman, & Hoischen, 2016; Crow, 1997).

Since germline mutation is extremely difficult to study directly, our understanding of the process is based on experiments in unicellular organisms and cell cultures or, in humans and other mammals, on how patterns of mutation in children depend on parental ages (Harland et al., 2017; Lindsay, Rahbari, Kaplanis, Keane, & Hurles, 2016; Lynch et al., 2016; Ségurel, Wyman, & Przeworski, 2014). Notably, the textbook view that replication errors are the primary source of human mutations (Crow, 2000; Drost & Lee, 1995; Li, Ellsworth, Krushkal, Chang, & Hewett-Emmett, 1996; Strachan & Read, 2010) often invokes the increase in the number of germline mutations with paternal age (Crow, 2000; Drost & Lee, 1995). In fact, however, this increase can arise from DNA replication, which occurs concurrently with cell divisions in spermatogonial stem cells, but also from other metabolic activities associated with cell division or from unrepaired damage that accrues with the passage of time (Müller, 1954). A further complication is that the rate at which DNA lesions are converted into mutations is affected by DNA replication and thus can depend on the cell division rate (Gao, Wyman, Sella, & Przeworski, 2016; Seplyarskiy, Sunyaev, Andrianova, Nikolaev, & Bazykin, 2017).

Insight about the sources of mutations can be gained, however, by contrasting male and female mutation patterns, which reflect distinct development trajectories and epigenetic dynamics. To apply this approach, we re-analyzed *de novo* mutation (DNM) data from a recent study of over 1500 parent-offspring trios (Jónsson et al., 2017) and contrasted the properties of paternal and maternal mutations. We tested the hypothesis that germline mutations are primarily replicative in origin by asking: How well do male and female mutations track the number of germ cell divisions? Do the dependencies of the mutation rate on the parent's sex and age differ by mutation types and if so, why? Is the higher male contribution to *de novo* mutations explained by replication errors in dividing male germ cells?

The ratio of paternal to maternal mutations is already high at puberty and is stable with parental age

If mutations are replication-driven, the ratio of male to female germline mutations (also known as the male mutational bias, α) should reflect the ratio of the number of germ cell divisions in the two sexes. While male and female germlines are thought to undergo similar numbers of cell divisions by the onset of puberty, thereafter the ratio of male-to-female cell divisions increases rapidly with age, because of frequent divisions of spermatogonial stem cells (an estimated 23 divisions per year) and the absence of mitotic division of female germ cells over the same period (Drost & Lee, 1995;

Vogel & Rathenberg, 1975). Thus, α is expected to increase with parental age at conception of the child, even though it need not be strictly proportional to the cell division ratio (depending on per cell division mutation rates during different stages of development) (Gao et al., 2016).

To test this prediction, we analyzed autosomal DNM data from 1548 Icelandic trios (Jónsson et al., 2017), initially focusing on the subset of mutations for which the parental origin of the DNM had been determined (i.e., phased mutations). Given that different numbers of mutations were phased in each child, we considered the fraction of paternal mutations in all phased DNMs and compared this fraction against the father's age (see SOM). Strikingly, for trios with similar ages of the father, G_P, and mother, G_M (where $0.9<G_P/G_M<1.1$), as is the case for most families in the dataset, the paternal contribution to mutations is stable with paternal age (Fig. 1). The correlation between the fraction of paternal DNMs and father's age is not significant at 5% level (by a Spearman correlation test, p=0.09), and the average contribution of paternal mutation lies around 75-80%, i.e., α = 3-4 across paternal ages. This result remains after excluding C>G mutations, which were previously reported to increase disproportionally rapidly with maternal age (Jónsson et al., 2017) (Fig. S1). Moreover, the same result is seen in an independent DNM dataset containing 816 trios (Goldmann et al., 2016; Wong et al., 2016), which are also mostly of European ancestry (henceforth the "Goldmann dataset"; Fig. S2). The finding of a stable α with parental ages calls into question the widespread belief that spermatogenesis drives the male bias in germline mutations (Crow, 1997; Makova & Li, 2002; Ségurel et al., 2014).

A stable α implies that maternal mutations increase with the mother's age at the same rate as paternal mutations increase with father's age. To obtain more precise estimates of parental age effects, we modeled paternal and maternal age effects jointly, leveraging information from both phased and unphased mutations in the Icelandic dataset. Briefly, we modeled the expected number of mutations in a parent as a linear function of her (his) age at conception of the child, and assumed that the observed number of maternal (paternal) mutations follows a Poisson distribution with this expectation. We further modeled the number of maternal (paternal) mutations that were successfully phased as a binomial sample of DNMs (see SOM). We then estimated the sex-specific yearly increases with parental ages by maximum likelihood (Table S1). Uncertainty in the estimates was evaluated by bootstrap resampling of trios. We confirmed earlier reports that the mutation counts on the paternal and maternal sets of chromosomes increased with father and mother's ages, respectively (Fig. 2A) (Goldmann et al., 2016; Jónsson et al., 2017; Wong et al., 2016).

In addition to a linear model, we considered exponential age effects of either or both sexes. We observed a significant improvement in fit of an exponential maternal age effect over a linear one (ΔAIC = -29.9), consistent with a previous analysis of the Goldmann dataset that indicated a more rapid increase in the maternal mutation rate at older ages (Wong et al., 2016). To verify our finding, we divided the 1,548 trios into two groups with maternal age at conception over or under the median age of 27 years and fit a linear model to the two separately. As expected from an accelerating increase in the number of maternal mutations with age, the estimate of the maternal age effect is greater for older mothers than for younger mothers (0.56 vs 0.24, 95% CI: [0.45,0.66] vs [0.12,0.38]; Table S4), whereas the estimates of paternal age effect are similar for the two groups (1.41 vs 1.40, 95% CI: [1.31, 1.51] vs [1.29, 1.53]). We further found that the exponential maternal age effect no longer provides a significantly better fit when excluding the 72 trios with maternal age over 40 (Table S5). As a sanity check on our estimates, we predicted the paternal mutation fraction for individuals with divergent paternal and maternal ages (G_P/G_M=0.9, 1.2 or

1.4); our predictions provide a good fit to the observed patterns for the subset of phased mutations (Fig. S3).

Focusing on the linear model fitted to trios with maternal ages below 40 years at conception (Table S5), the male-to-female mutation ratio is already ~3 (95% CI: [2.8, 3.5]) at the onset of puberty (assumed to be 13 years of age for both sexes; Fig. 2B,C)(Nielsen et al., 1986), consistent with the observation of stable fraction of paternal mutations with paternal age (Fig. 1), and indicating that the male germline has accumulated a substantially greater number of DNMs than the female germline by puberty. The same is seen in our reanalysis of the smaller Goldmann data set (Fig. S4). At face value, this finding is puzzling: male and female germ cells are thought to experience comparable numbers of divisions by then (an estimated 38 vs 31, respectively) and approximately half of these divisions predate sexual differentiation (SD) (Drost & Lee, 1995; Vogel & Rathenberg, 1975), so we would expect males and females to harbor similar numbers of replicative mutations. Moreover, differences in the mutation spectrum between males and females are subtle (Jónsson et al., 2017; Rahbari et al., 2016; Agarwal & Przeworski, unpublished results), suggesting that the sources of most mutations may be shared between the two sexes.

How then to explain that α is already high by puberty and persists at roughly the same value throughout adulthood? Three possible resolutions are that (1) the number of male germ cell divisions from SD to puberty has been vastly underestimated; (2) after SD, germ cell divisions are much more mutagenic in males than in females; or (3) damage-induced mutations contribute a substantial fraction to male germline mutations by puberty. The first two possibilities require specific, implausible conditions to hold on the numbers of cell divisions and the mutation rates per cell division over developmental stages in both sexes, such that male-to-female ratio of replication errors before puberty is coincidentally similar to the ratio of increases in mutation rates in the two sexes after puberty (see SOM)(Gao et al., 2016). Instead, we hypothesize that most germline mutations in both sexes are damage-induced: under this scenario, the elevated α by puberty and the stable α after puberty would reflect damage rates that are roughly constant per unit time in both sexes and higher in males (Ségurel et al., 2014).

Specific sources of DNA damage-induced mutations

To explore possible mutation mechanisms, we classified each DNM into six disjoint and complementary mutation classes based on parental and derived alleles: T>A, T>C, T>G, C>A, C>G, and C>T (each type also includes the corresponding variant on the reverse complement strand), following previous studies (Alexandrov et al., 2013; Harris, 2015; Rahbari et al., 2016). Given the well-characterized hypermutability of methylated CpG sites, we further divided the C>T transitions into sub-types in non-CpG and CpG contexts (excluding CpG islands for the latter; see SOM). Confirming the original analysis of these data (Jónsson et al., 2017), we detected significant paternal and maternal age effects for all seven mutation types, of varying magnitudes (Table S1; Fig. S5). While α is stable with parental age for most mutation types (Fig. S5), C>G transversions and C>T transitions at CpG sites stand out from the general pattern (Jónsson et al., 2017)(Fig. 3). In particular, C>G mutations show a decreasing α with age and CpG>TpG mutations an increasing α, evident in both the direct analysis of phased mutations and in our modeling of all mutations (Fig. 3A, B). We further found that whereas the linear model is a good fit to the other six mutation types individually, for C>G mutations, a model with a linear paternal age effect and an exponential maternal age effect provides a significant better fit (ΔAIC = -18.3; Table S3). An exponential maternal age effect also provides a significantly better fit for all point mutations other than C>G

(ΔAIC<-9; Table S5), and again the effect disappears when the 72 trios with maternal age over 40 are excluded, suggesting that mutation types other than C>G are also increasing at higher rates in mothers of older ages.

Based on their spatial distribution in the genome and their increase with maternal age, maternal C>G mutations were hypothesized to be associated with double-strand breaks in aging oocytes (Titus et al., 2013): this mutation type often appears in clusters with strong strand-concordance and near *de novo* copy number variant breakpoints and is enriched in genomic regions with elevated rates of non-crossover gene conversion--an enrichment that increases rapidly with maternal age (Goldmann et al., 2018; Jónsson et al., 2017). C>G mutations are also over-represented on human pseudo-autosomal 1 region, which experiences an obligate crossover in males, compared to X or autosomes (Agarwal & Przeworski, unpublished results). We additionally found that C>G transversions are significantly more likely than other mutation types to occur on the same chromosome as a *de novo* deletion (≥5bp) in the same individual, and conditional on co-occurrence, that the distance to the closest deletion tends to be shorter (Fig. S6). The same association is not seen for short deletions or insertions, however (Table S4), which are more likely to arise from replication slippage (Kloosterman et al., 2015; Montgomery et al., 2013). Together, these observations support imperfect repair of double-strand breaks (DSB) as an important source of C>G transversions in both sexes (Goldmann et al., 2018; Jónsson et al., 2017; Agarwal & Przeworski, unpublished results).

In turn, biochemical evidence and phylogenetic studies suggest that the high rates of C>T transitions at CpG sites are due to the spontaneous deamination of methylcytosine (Fryxell & Zuckerkandl, 2000; Lindahl & Nyberg, 1974). Since epigenetic profiles differ dramatically between mammalian male and female germlines, in ways that are relatively well characterized, this source of mutation leads to clear predictions about when sex differences in C>T transitions might arise in development. During embryogenesis, there are several rounds of global DNA demethylation and remethylation to enable the erasure and re-establishment of the epigenetic memory (Reik, Dean, & Walter, 2001). Because methylation dynamics are thought to be shared until sex determination of the embryo, during early development, male and female germlines should share most methylation-related mutations. Consistent with this prediction, we estimated a lower α for CpG transitions than for other mutation types at early reproductive ages (e.g., at age 17, α=2.6 [2.2, 3.0] vs 3.4 [3.2, 3.7] for mutations other than CpG>TpG) (Fig. 3B). After sex determination (around 7 weeks post fertilization in humans), the methylation profiles of male and female germ cells diverge: re-methylation takes place early in males, before differentiation of spermatogonia, but seemingly very late in females, shortly before ovulation (Reik et al., 2001). Therefore, the male germline is markedly more methylated compared to the female germline for the long period from sex determination of the parents to shortly before conception of the child. Accordingly, after puberty, the estimated yearly increase in CpG>TpG mutations is 6.5-fold higher in fathers than in mothers, roughly double what is seen for other mutation types, resulting in a marked increase in α with parental age at CpG>TpG (Fig. 3C). In summary, the sex and age dependencies of CpG transitions accord with the sex-specific methylation profiles of the mammalian germline, supporting the notion that deamination of methylated cytosines are the major sources of CpG transitions, and validating our inferences for the one case in which we have independent information about what to expect. Together with C>G mutations, CpG transitions represent approximately 1/3 of germline mutations accumulated in a parent of age 30 at conception; both appear to be generated by DNA lesion-induced mutational mechanisms.

De novo mutations on both maternal and paternal genomes increase with maternal age

In mammals, primary oocytes are formed and arrested in prophase of meiosis I, before the birth of the future mother, with no further DNA replication occurring until fertilization. Therefore, the maternal age effect detected by recent, large DNM studies (Goldmann et al., 2016; Jónsson et al., 2017; Wong et al., 2016) and confirmed here (Fig. 2A) has been interpreted as reflecting the accumulation of DNA lesions or damage-induced mutations in (primary) oocytes during the lengthy meiotic arrest phase (Gao et al., 2016; Goriely, 2016; Jónsson et al., 2017; Ségurel et al., 2014). However, other explanations for a maternal age effect are possible (Wong et al., 2016). For example, such an effect could also arise if oocytes ovulated later in life undergo more mitoses (Fulton, Silva, Bayne, & Anderson, 2005; Polani & Crolla, 1991). In this scenario, the substantial increase in maternal DNMs from age 17 to age 40 (Fig. 2A) would require oocytes ovulated later in life to go through almost double the number of cell divisions compared to oocytes ovulated early (more, if the per cell division mutation rate is higher in early cell divisions; see below). Moreover, this scenario does not provide an explanation for the stability of the male-to-female ratio with parental ages. Thus, while this phenomenon could hypothetically contribute to the maternal age effect, in practice, it is likely to be a minor effect (see SOM for a more detailed discussion).

A more plausible explanation is a maternal age effect on post-zygotic mutations. Although DNMs are usually interpreted as mutations that occur in the parents, in fact what are identified as DNMs in trio studies are the genomic differences between the offspring and the parents in the somatic tissues sampled (here, blood). These differences can arise in the parents but also during early development of the child (Fig. 4A)(Moorjani, Amorim, Arndt, & Przeworski, 2016). Notably, the first few cell divisions of embryogenesis have been found to be relatively mutagenic, leading to somatic and germline mosaic mutations in a study of four generations of cattle (Harland et al., 2017) and to a lesser extent in humans (Acuna-Hidalgo et al., 2015; Huang et al., 2014; Ju et al., 2017), as well as to mutations at appreciable frequency that are discordant between monozygotic twins (Dal et al., 2014). Increased numbers of mutations in the first few cell divisions may be expected, as two key components of base excision repair are missing in spermatozoa, leading lesions accumulated in the last steps of spermatogenesis to be repaired only in the zygote (Smith et al., 2013). More generally, mammalian zygotes are almost entirely reliant on the protein and transcript reservoirs of the oocyte until the 4-cell stage (Braude, Bolton, & Moore, 1988; Dobson et al., 2004; Zhang et al., 2009). Thus, deterioration of the replication or repair machinery in oocytes from older mothers may lead to higher mutation rates in the first few cell divisions of the embryo (Fig. 4B) (Titus et al., 2013; Wei et al., 2015). This scenario predicts that the mother's age would influence not only the number of mutations on the chromosomes inherited from the mother but also from the father (which would be assigned to "paternal mutations") (Fig. 4B).

Any such effect is challenging to detect, because of the noise induced by incomplete phasing of mutations and the large sampling variance in mutation counts, as well as the high correlation between maternal and paternal ages. Nonetheless, when we focused on the 199 probands in which almost all DNMs are phased (>95% DNM phased), a Poisson regression of the count of paternal mutations on both parental ages revealed a significant effect of maternal age ($p=0.035$) and a slight but non-significant improvement in the fit compared to a model with paternal age only ($\Delta AIC=-2.4$; Table S7). We verified that such an effect should not arise artifactually, from the correlation between maternal and paternal ages and the assignment of parental ages to 1-year bins ($p=0.007$ see SOM for details).

Motivated by this surprising finding, we carried out further analyses conditional on paternal age. First, using the 199 trios, we compared all pairs with the same paternal age but different maternal ages. Among 619 such pairs, the child born to the older mother carries more paternal mutations than does the child with the younger mother in 319 cases, fewer in 280 cases, and the same number in 20 cases (Fig. 4C), and greater differences in the number of paternal mutations are associated with greater differences in maternal ages (Kendall's rank test τ=0.09, p=0.024 by a permutation test; see SOM for details and an alternative approach; Fig. S7). In contrast, there is no effect of paternal age on the maternal number of mutations, matching for maternal age (p>0.31; Fig. 4D; Table S7). Adding a maternal age effect on paternal mutations to the model, the fit slightly improves (ΔAIC=-1.9), and the MLE for the paternal age effect on paternal mutations decreases by about 15%: the slope is 1.20 (95% CI [0.89, 1.45]) instead of the 1.39 (95% CI [1.24, 1.52]) obtained without a maternal age effect (Fig. 4C). We note that a similar effect is not seen when considering all Icelandic trios, including those without a third generation, whose mutation properties appear to differ significantly from those of families with a third generation (see SOM for details). However, focusing on the subset of families with three generations, in which error rates are likely to be lower, there is a clear maternal age effect on paternal mutations, consistent with a small post-zygotic effect of maternal age having been soaked up in previous estimates of the paternal age effect (Tables S7, S10).

The estimated effect of maternal age on maternal mutations is 0.34 mutations per year (s.e.=0.04) by Poisson regression (p=3.4e-13). The estimated maternal age effect on paternal mutations is similar but highly uncertain (0.30, s.e.=0.14). Naively, one might expect the maternal age effect on maternal mutations to be stronger, as it includes both pre-zygotic effects (e.g., damage in the oocyte) and post-zygotic effects, whereas the effect on paternal mutations can only be post-zygotic. This expectation is implicitly based on the assumption of same post-zygotic effects of maternal age on maternal and paternal genomes, but they need not be, both because before fertilization, sperm and oocytes may harbor different levels of DNA damage (e.g., oxidative stress may be higher in male germ cells) (De Iuliis et al., 2009; Lim & Luderer, 2011) and because after fertilization but before the first cleavage, the two parental genomes experience distinct epigenetic remodeling and are replicated separately, in their own pronuclei (Ferreira & Carmo-Fonseca, 1997; Mayer, Niveleau, Walter, Fundele, & Haaf, 2000). Thus, even if the precise effects of maternal age on the zygote were known, the relative contributions of pre-zygotic and post-zygotic effects of maternal age on the maternal genome are not distinguishable without additional data. Regardless, the positive association between maternal age and the number of DNMs on paternal chromosomes reveals that a mother's age at conception affects the post-zygotic mutation rate in the developing embryo.

In cattle and humans, the high frequency mosaic mutations that are likely to have arisen in early embryonic development are enriched for C>A transversions (Harland et al., 2017; Huang et al., 2014; Ju et al., 2017), potentially reflecting the accumulation of the oxidative DNA damage 8-hydroxyguanine in oocytes and the last stages of spermatogenesis (De Iuliis et al., 2009; Lim & Luderer, 2011) that remains uncorrected in spermatozoa (Smith et al., 2013). Hypothesizing that a maternal age effect may be particular pronounced for these mutations, we focused on the C>A mutations in the 199 probands with high phasing rates. Although this subset represents only ~8% of mutations, there is a significant effect of maternal age on paternal mutations (p=0.02 by Poisson regression), and the point estimate of the maternal age effect on paternal genome (0.095, s.e.=0.041) is even stronger than that of paternal age (0.057, s.e.=0.033), as well as stronger than the maternal age effect on maternal genome (0.024, s.e.=0.0094; see SOM for details). Such results are rarely

obtained in a random subset of mutations of the same size (p=0.045; see SOM), suggesting paternal C>A mutations are indeed more strongly affected by maternal age than are other DNMs. Thus, a mutation type associated with damage in sperm and known to be enriched in early embryogenesis shows a heightened signal of a maternal age effect on paternal mutations.

Overlooked roles of DNA damage and maternal age in generating human germline mutations

We found that the age of the mother influences the mutation rate in the early development of the child, most plausibly in the early cleavage stage when the zygote is dependent on maternal transcripts, but possibly also because of in utero effects. This finding implies that part of the parental age effects that were previously interpreted as reflecting pre-zygotic mutations are in fact due to mutations that arise in the zygote, and at higher rates in older mothers. What other consequences maternal age may have on cellular processes in the early embryo remain to be investigated (Janny & Menezo, 1996).

Importantly, the effect on mutation rates could arise from mutagens or degradation of the repair machinery or from decreased replication fidelity. For instance, there is a maternal age effect on paternal C>A mutations, which is a signature of oxidative damage and other mutagens (Kuchino et al., 1987). Thus, the existence of a maternal age effect on the early embryo does not distinguish between replicative and damage-induced sources of mutations. Our other findings, however, call into question the textbook view that germline mutations are predominantly replicative in origin. First, multiple lines of evidence suggest that CpG transitions and C>G mutation often arise from methylation-associated damage and double-strand break repair, respectively. Second, even excluding both of these mutation types, roughly three-fold more paternal mutations than maternal mutations have occurred by puberty of the future parent, despite similar numbers of estimated germ cell divisions by that age (Fig. 2B, C). For these remaining mutation types, the male-to-female mutational ratio is remarkably stable with parental age, even as the ratio of male-to-female cell divisions increases rapidly (Fig. 2B, C). That α is already 3 by puberty could be explained by a vast underestimation of the number of germ cell divisions in males between birth and puberty, but not its stability with parental ages. Lastly, despite highly variable cell division rates over development, germline mutations accumulate in rough proportion to absolute time in each sex (Fig. 2A; Fig. S5). Together, these findings point to a substantial role of DNA damage-induced mutations, raising questions about the relative importance of endogenous versus exogenous mutagens, as well as about why male and female germ cells differ in the balance of DNA damage and repair.

These findings also carry implications for the differences in mutation rates across populations and species (e.g., Harris & Pritchard, 2017; Hwang & Green, 2004). In mammals, older ages of reproduction are associated with decreased ratios of X to autosome divergence (interpreted as higher male-to-female mutation ratios; but see Amster & Sella, 2016) and lower substitutions rates, with a weaker relationship seen for CpG transitions (Kim, Elango, Warden, Vigoda, & Yi, 2006; Li & Tanimura, 1987; Makova & Li, 2002; Sayres et al., 2011). These observations that have been widely interpreted as supporting a replicative origin of most non-CpG transitions (Kim et al., 2006; Li et al., 1996; Makova & Li, 2002; Moorjani et al., 2016; Sayres et al., 2011; Ségurel et al., 2014). Instead, our analyses of DNMs show generally weak effects of reproductive age on the ratio of male-to-female mutation as well as on yearly mutation rates (Fig. S6C), an important role for non-

replicative mutations beyond CpG transitions, and a coupling of maternal and paternal age effects. One possibility is that inter-species differences in the male-to-female mutation ratio and in substitution rates reflect changes in the ratio of paternal and maternal ages at reproduction (Fig. 2B; Amster & Sella, 2016) and in rates of DNA damage (e.g., metabolic rates) that covary with life history traits (Martin & Palumbi, 1993; Sayres et al., 2011).

Acknowledgments: We thank M. Eisen and Z. Williams, as well as I. Agarwal, A. Harpak, J. Pritchard and other members of the Przeworski and Pritchard labs for helpful discussions. This work was supported by NIH GM122975 to M. Przeworski, NIH HG009431 to J. Pritchard and a Burroughs Wellcome Fund CASI award to P. Moorjani.

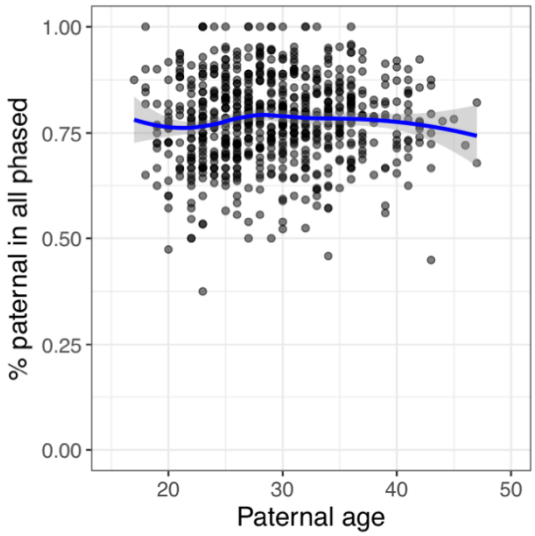

Figure 1. The fraction of paternal mutations among phased mutations, as a function of paternal age at conception. Each point represents the data for one child (proband) with at least three phased mutations and similar parental ages (paternal-to-maternal age ratio between 0.9 to 1.1; 719 trios total). The blue line is the LOESS curve fitted to the scatterplot, with the shaded area representing the 95% confidence interval (calculated with the geom_smooth function in R with default parameters).

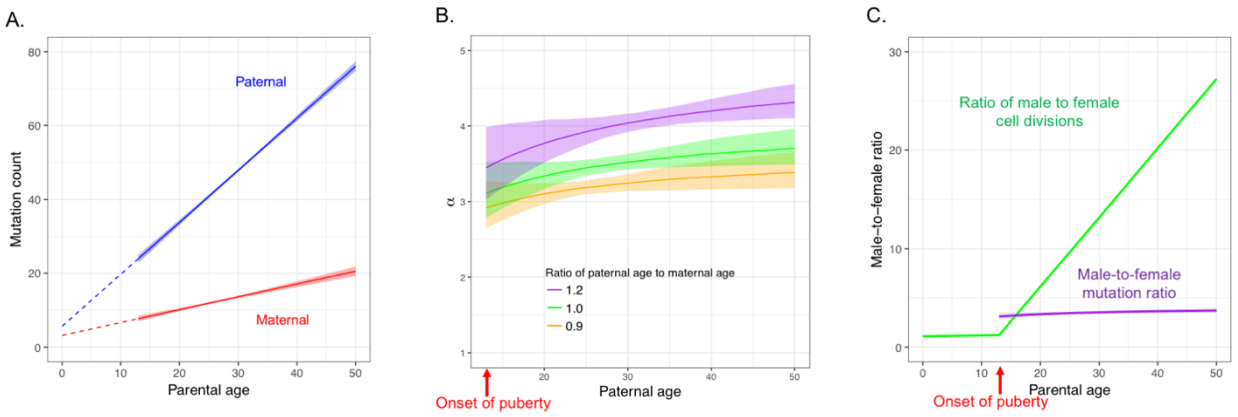

Figure 2. Inferred sex and age dependencies of germline mutations (based on linear model with trios with maternal age no greater than 40y). In all panels, shaded areas and bars represent 95% confidence intervals of the corresponding quantities obtained from bootstrapping. (A) Inferred sex-specific mutation rates as a function of parental ages. Parental ages are measured since birth, i.e., birth corresponds to age 0 (same throughout the manuscript). The extrapolated intercepts at age 0 are small but significantly positive for both sexes, implying a weak but significant effect of reproductive age on yearly mutation rates (Gao et al., 2016). (B) Predicted male-to-female mutation ratio (α) as a function of the ratio of paternal to maternal ages. For reference, the ratio of parental ages is centered around 1.10 in the Icelandic DNM data set (s.d.=0.20). (C) Contrast between male-to-female mutation ratio (purple) and the ratio of male-to-female cell divisions (green), assuming the same paternal and maternal ages. Estimates of the cell division numbers for the two sexes in humans are from Drost & Lee (1995).

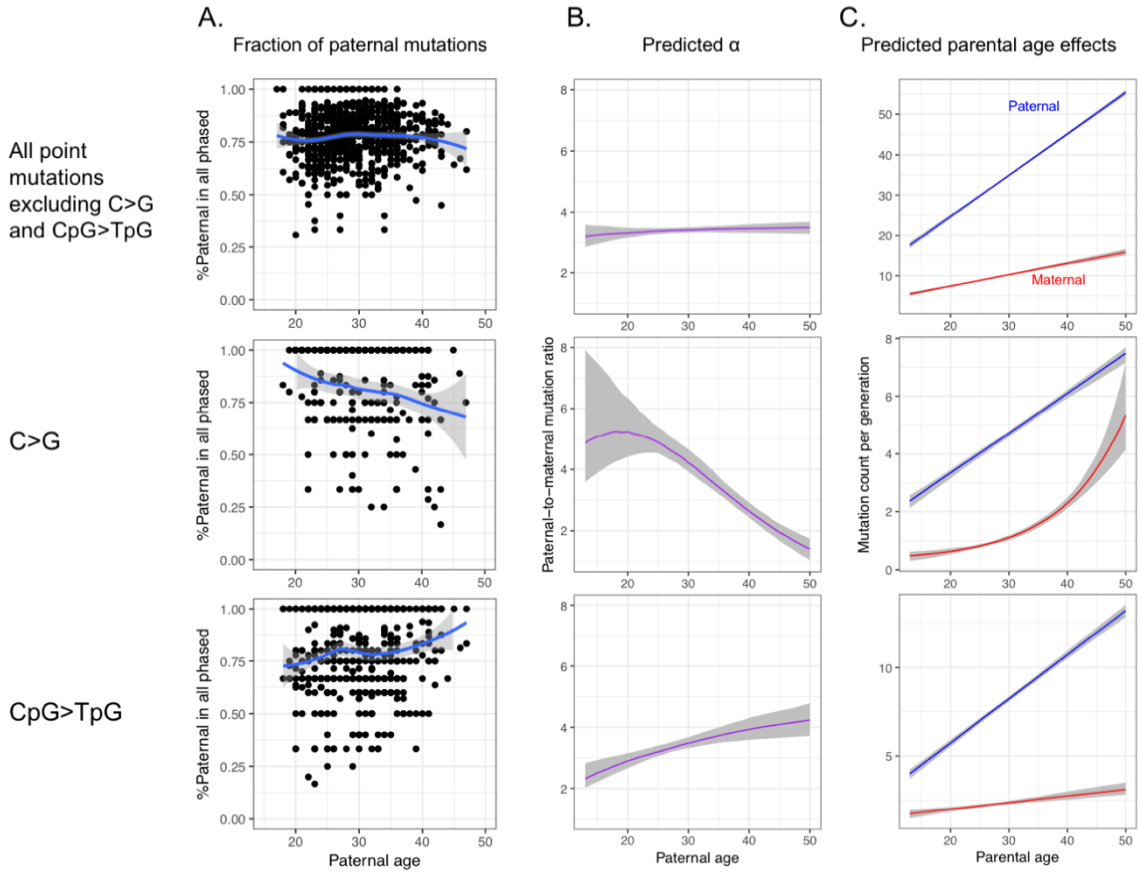

Figure 3. Distinctive sex and age dependencies for C>G and CpG>TpG DNMs. The shaded areas in all panels represent 95% confidence intervals. See Fig S5 for similar plots for other mutation types. The male-to-female mutation ratio at age 17 is significantly lower for CpG>TpG than for other mutation types (see Main text). (A) Fraction of paternal mutations in phased DNMs; (B) Predicted male-to-female mutation ratio (α); (C) Predicted parental age effects.

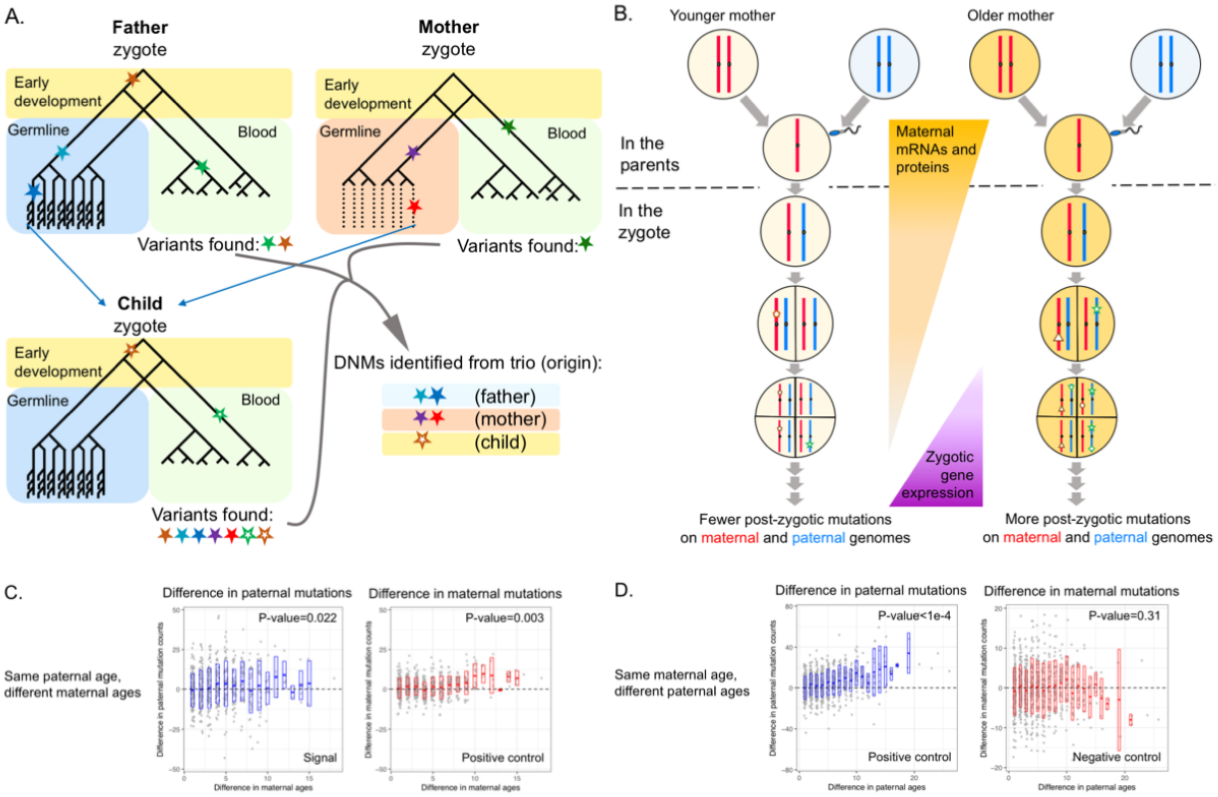

Figure 4. Maternal age effect on mutations that occur on paternally inherited chromosomes. (A) An illustration of mutations occurring during development and gametogenesis. In this cartoon, we assume that the most recent common ancestor of all cells in an individual is the fertilized egg. Filled stars represent mutations that arise in the parents and hollow stars in the child. The standard trio approach requires allelic balance in the child and no or few reads carrying the alternative allele in the parent, leading to inclusion of some early post-zygotic mutations in the child (brown hollow) and exclusion of a fraction of early mutations in the parents (brown filled). The two effects partially cancel out when estimating the per generation mutation rate, but potentially lead to underestimate of the fraction of mutations that are early embryonic (Harland et al., 2017; Moorjani et al., 2016). Damage-induced mutations in the oocyte (red filled) and post-zygotic mutations (brown hollow) in the child are both child-specific, but their mosaic status can be evaluated by examining transmissions to a third generation. (B) An illustration of a maternal age effect on the number of post-zygotic mutations. (C) Pairwise comparison conditional on the same paternal age. Each point represents a pair of trios, with x-axis showing the difference in maternal ages and y-axis the difference in paternal mutation counts (left; older mother – younger mother) or maternal mutation counts (right; older mother – younger mother); position is slightly shifted to show overlapping points. P-values are evaluated by 10,000 permutations, using Kendall's rank correlation test statistic (see SOM). (D) Pairwise comparison conditional on the same maternal age, similar to (C). The ranges of y-axis differ for the plots on left and right.

References and Notes:

Acuna-Hidalgo, R., Bo, T., Kwint, M. P., Van De Vorst, M., Pinelli, M., Veltman, J. A., … Gilissen, C. (2015). Post-zygotic Point Mutations Are an Underrecognized Source of de Novo Genomic Variation. *American Journal of Human Genetics*, *97*(1), 67–74. https://doi.org/10.1016/j.ajhg.2015.05.008

Acuna-Hidalgo, R., Veltman, J. A., & Hoischen, A. (2016). New insights into the generation and role of de novo mutations in health and disease. *Genome Biology*, *17*(1), 1–19. https://doi.org/10.1186/s13059-016-1110-1

Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A. J. R., Behjati, S., Biankin, A. V., … Stratton, M. R. (2013). Signatures of mutational processes in human cancer. *Nature*, *500*(7463), 415–421. https://doi.org/10.1038/nature12477

Amster, G., & Sella, G. (2016). Life history effects on the molecular clock of autosomes and sex chromosomes. *Proceedings of the National Academy of Sciences*, *113*(6), 1588–1593. https://doi.org/10.1073/pnas.1515798113

Braude, P., Bolton, V., & Moore, S. (1988). Human gene expression first occurs between the four- and eight-cell stages of preimplantation development. *Nature*, *332(6163)*, 459–461.

Crow, J. F. (1997). The high spontaneous mutation rate: is it a health risk? *Proceedings of the National Academy of Sciences of the United States of America*, *94*(16), 8380–8386. https://doi.org/10.1073/pnas.94.16.8380

Crow, J. F. (2000). The origins, patterns and implications of human spontaneous mutation. *Nature Reviews Genetics*, *1*(1), 40–47. https://doi.org/10.1038/35049558

Dal, G. M., Ergüner, B., Sağıroğlu, M. S., Yüksel, B., Onat, O. E., Alkan, C., & Özçelik, T. (2014). Early postzygotic mutations contribute to de novo variation in a healthy monozygotic twin pair. *Journal of Medical Genetics*, *51*(7), 455–459. https://doi.org/10.1136/jmedgenet-2013-102197

De Iuliis, G. N., Thomson, L. K., Mitchell, L. A., Finnie, J. M., Koppers, A. J., Hedges, A., … Aitken, R. J. (2009). DNA Damage in Human Spermatozoa Is Highly Correlated with the Efficiency of Chromatin Remodeling and the Formation of 8-Hydroxy-2′-Deoxyguanosine, a Marker of Oxidative Stress1. *Biology of Reproduction*, *81*(3), 517–524. https://doi.org/10.1095/biolreprod.109.076836

Dobson, A. T., Raja, R., Abeyta, M. J., Taylor, T., Shen, S., Haqq, C., & Pera, R. A. R. (2004). The unique transcriptome through day 3 of human preimplantation development. *Human Molecular Genetics*, *13*(14), 7–9. https://doi.org/10.1093/hmg/ddh157

Drost, J. B., & Lee, W. R. (1995). Biological Basis of Germline Mutation : Comparisons of Spontaneous Germline Mutation Rates. *Environmental and Molecular Mutagenesis*, *64*(1 995), 48–64.

Ferreira, J., & Carmo-Fonseca, M. (1997). Genome replication in early mouse embryos follows a defined temporal and spatial order. *Journal of Cell Science*, *110 (Pt 7*, 889–897.

Fryxell, K. J., & Zuckerkandl, E. (2000). Cytosine deamination plays a primary role in the evolution of mammalian isochores. *Molecular Biology and Evolution*, *17*(9), 1371–1383. https://doi.org/10.1093/oxfordjournals.molbev.a026420

Fulton, N., Silva, S. J. M., Bayne, R. A. L., & Anderson, R. A. (2005). Germ Cell Proliferation and Apoptosis in the Developing, *90*(May), 4664–4670. https://doi.org/10.1210/jc.2005-0219

Gao, Z., Wyman, M. J., Sella, G., & Przeworski, M. (2016). Interpreting the Dependence of Mutation Rates on Age and Time. *PLoS Biology*, *14*(1), 1–16. https://doi.org/10.1371/journal.pbio.1002355

Goldmann, J. M., Seplyarskiy, V. B., Wong, W. S. W., Vilboux, T., Neerincx, P. B., Bodian, D. L., … Niederhuber, J. E. (2018). Germline de novo mutation clusters arise during oocyte aging in genomic regions with high double-strand-break incidence. *Nature Genetics*, 1–6. https://doi.org/10.1038/s41588-018-0071-6

Goldmann, J. M., Wong, W. S. W., Pinelli, M., Farrah, T., Bodian, D., Stittrich, A. B., … Niederhuber, J. E. (2016). Parent-of-origin-specific signatures of de novo mutations. *Nature Genetics*, *48*(8), 935–939. https://doi.org/10.1038/ng.3597

Goriely, A. (2016). Decoding germline de novo point mutations. *Nature Genetics*, *48*(8), 823–824. https://doi.org/10.1038/ng.3629

Harland, C., Charlier, C., Karim, L., Cambisano, N., Deckers, M., Mni, M., … Georges, M. (2017). Frequency of mosaicism points towards mutation-prone early cleavage cell divisions. *BioRxiv*.

Harris, K. (2015). Evidence for recent, population-specific evolution of the human mutation rate. *Proceedings of the National Academy of Sciences*, *112*(11), 3439–3444. https://doi.org/10.1073/pnas.1418652112

Harris, K., & Pritchard, J. K. (2017). Rapid evolution of the human mutation spectrum. *ELife*, *6*, 1–17. https://doi.org/10.7554/eLife.24284

Huang, A. Y., Xu, X., Ye, A. Y., Wu, Q., Yan, L., Zhao, B., … Wei, L. (2014). Postzygotic single-nucleotide mosaicisms in whole-genome sequences of clinically unremarkable individuals. *Cell Research*, *24*(11), 1311–1327. https://doi.org/10.1038/cr.2014.131

Huttley, G. A., Jakobsen, I. B., Wilson, S. R., & Easteal, S. (2000). How important is DNA replication for mutagenesis? *Molecular Biology and Evolution*, *17*(6), 929–937. https://doi.org/10.1093/oxfordjournals.molbev.a026373

Hwang, D. G., & Green, P. (2004). Bayesian Markov chain Monte Carlo sequence analysis reveals varying neutral substitution patterns in mammalian evolution. *Proceedings of the National Academy of Sciences*, *101*(39), 13994–14001. https://doi.org/10.1073/pnas.0404142101

Janny, L., & Menezo, Y. J. (1996). Maternal age effect on early human embryonic development and blastocyst formation. *Molecular Reproduction and Development*, *45*(1), 31–37. https://doi.org/10.1002/(SICI)1098-2795(199609)45:1<31::AID-MRD4>3.0.CO;2-T

Jónsson, H., Sulem, P., Kehr, B., Kristmundsdottir, S., Zink, F., Hjartarson, E., … Stefansson, K. (2017). Parental influence on human germline de novo mutations in 1,548 trios from Iceland. *Nature*, *549*(7673), 519–522. https://doi.org/10.1038/nature24018

Ju, Y. S., Martincorena, I., Gerstung, M., Petljak, M., Alexandrov, L. B., Rahbari, R., … Stratton, M. R. (2017). Somatic mutations reveal asymmetric cellular dynamics in the early

human embryo. *Nature*, *543*(7647), 714–718. https://doi.org/10.1038/nature21703

Kim, S. H., Elango, N., Warden, C., Vigoda, E., & Yi, S. V. (2006). Heterogeneous genomic molecular clocks in primates. *PLoS Genetics*, *2*(10), 1527–1534.

Kloosterman, W. P., Francioli, L. C., Hormozdiari, F., Marschall, T., Hehir-kwa, J. Y., Abdellaoui, A., … Guryev, V. (2015). Characteristics of de novo structural changes in the human genome. *Genome Res.*, *25*, 792–801. https://doi.org/10.1101/gr.185041.114.19

Kuchino, Y., Mori, F., Kasai, H., Inoue, H., Iwai, S., Miura, K., … Nishimura, S. (1987). Misreading of DNA templates containing 8-hydroxydeoxyguanosine at the modified base and at adjacent residues. *Nature*, *327*(6117), 77–79. https://doi.org/10.1038/327077a0

Kumar, S., & Subramanian, S. (2002). Mutation rates in mammalian genomes. *Proceedings of the National Academy of Sciences*, *99*(2), 803–808. https://doi.org/10.1073/pnas.022629899

Li, W.-H., Ellsworth, D. L., Krushkal, J., Chang, B. H. J., & Hewett-Emmett, D. (1996). Rates of nucleotide substitution in primates and rodents and the generation-time effect hypothesis. *Molecular Phylogenetics and Evolution*, *5*(1), 182–187. https://doi.org/10.1006/mpev.1996.0012

Li, W.-H., & Tanimura, M. (1987). The molecular clock runs more slowly in man than in apes and monkeys. *Nature*, *326*(6108), 93–96. https://doi.org/10.1038/326093a0

Lim, J., & Luderer, U. (2011). Oxidative Damage Increases and Antioxidant Gene Expression Decreases with Aging in the Mouse Ovary. *Biology of Reproduction*, *84*(4), 775–782. https://doi.org/10.1095/biolreprod.110.088583

Lindahl, T., & Nyberg, B. (1974). Heat-induced deamination of cytosine residues in deoxyribonucleic acid. *Biochemistry*, *13*(16), 3405–3410. https://doi.org/10.1021/bi00713a035

Lindsay, S. J., Rahbari, R., Kaplanis, J., Keane, T. M., & Hurles, M. E. (2016). Striking differences in patterns of germline mutation between mice and humans. *BioRxiv*.

Lynch, M. (2010). Evolution of the mutation rate. *Trends in Genetics*, *26*(8), 345–352. https://doi.org/10.1016/j.tig.2010.05.003

Lynch, M., Ackerman, M. S., Gout, J.-F., Long, H., Sung, W., Thomas, W. K., & Foster, P. L. (2016). Genetic drift, selection and the evolution of the mutation rate. *Nature Reviews Genetics*, *17*(11), 704–714. https://doi.org/10.1038/nrg.2016.104

Makova, K. D., & Li, W. H. (2002). Strong male-driven evolution of DNA sequences in humans and apes. *Nature*, *416*(6881), 624–626. https://doi.org/10.1038/416624a

Martin, A. P., & Palumbi, S. R. (1993). Body size, metabolic rate, generation time, and the molecular clock. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(9), 4087–4091. https://doi.org/10.1073/pnas.90.9.4087

Mayer, W., Niveleau, A., Walter, J., Fundele, R., & Haaf, T. (2000). Demethylation of the zygotic paternal genome. *Nature*, *403*(6769), 501–502. https://doi.org/10.1038/35000656

Montgomery, S. B., Goode, D. L., Kvikstad, E., Albers, C. A., Zhang, Z. D., Mu, X. J., … Lunter, G. (2013). The origin , evolution , and functional impact of short insertion – deletion variants identified in 179 human genomes, 749–761.

https://doi.org/10.1101/gr.148718.112.Freely

Moorjani, P., Amorim, C. E. G., Arndt, P. F., & Przeworski, M. (2016). Variation in the molecular clock of primates. *Proceedings of the National Academy of Sciences*, *113*(38), 10607–10612. https://doi.org/10.1073/pnas.1600374113

Müller, H. . (1954). The nature of genetic effects produced by irradiation. *Radiation Biology*, (7), 351–473.

Nielsen, C. T., Skakkæk, N. E., Richardson, D. W., Darling, J. A. B., Hunter, W. M., Jørgensen, M., … Müller, J. (1986). Onset of the release of spermatozia (supermarche) in boys in relation to age, testicular growth, pubic hair, and height. *Journal of Clinical Endocrinology and Metabolism*, *62*(3), 532–535. https://doi.org/10.1210/jcem-62-3-532

Polani, P. E., & Crolla, J. A. (1991). A test of the production line hypothesis of mammalian oogenesis, 64–70.

Rahbari, R., Wuster, A., Lindsay, S. J., Hardwick, R. J., Alexandrov, L. B., Al Turki, S., … Hurles, M. E. (2016). Timing, rates and spectra of human germline mutation. *Nature Genetics*, *48*(2), 126–133. https://doi.org/10.1038/ng.3469

Reik, W., Dean, W., & Walter, J. (2001). Epigenetic reprogramming in mammalian development. *Science*, *293*(5532), 1089–1093.

Sayres, M. A. W., Venditti, C., Pagel, M., & Makova, K. D. (2011). Do variations in substitution rates and male mutation bias correlate with life-history traits? a study of 32 mammalian genomes. *Evolution*, *65*(10), 2800–2815. https://doi.org/10.1111/j.1558-5646.2011.01337.x

Ségurel, L., Wyman, M. J., & Przeworski, M. (2014). Determinants of Mutation Rate Variation in the Human Germline. *Annual Review of Genomics and Human Genetics*, *15*(1), 47–70. https://doi.org/10.1146/annurev-genom-031714-125740

Seplyarskiy, V. B., Sunyaev, S. R., Andrianova, M. A., Nikolaev, S. I., & Bazykin, G. A. (2017). Error-prone bypass of DNA lesions during lagging strand replication is a common source of germline and cancer mutations. *BioRxiv*.

Smith, T. B., Dun, M. D., Smith, N. D., Curry, B. J., Connaughton, H. S., & Aitken, R. J. (2013). The presence of a truncated base excision repair pathway in human spermatozoa that is mediated by OGG1. *Journal of Cell Science*, *126*(6), 1488–1497. https://doi.org/10.1242/jcs.121657

Strachan, T., & Read, A. (2010). *Human Molecular Genetics* (4th ed.). Garland Science.

Titus, S., Li, F., Stobezki, R., Akula, K., Unsal, E., Jeong, K., … Oktay, K. (2013). Impairment of BRCA1-Related DNA Double-Strand Break Repair Leads to Ovarian Aging in Mice and Humans, *5*(M). https://doi.org/10.1126/scitranslmed.3004925

Vogel, F., & Rathenberg, R. (1975). Spontaneous mutation in man. *Advances in Human Genetics*, *5*, 223–318.

Wei, H., Liu, X., Yuan, J., Li, L., Zhang, D., Guo, X., … Zhang, S. (2015). Age-Specific Gene Expression Profiles of Rhesus Monkey Ovaries Detected by Microarray Analysis. *BioMed Research International*, *2015*. https://doi.org/10.1155/2015/625192

Wong, W. S. W., Solomon, B. D., Bodian, D. L., Kothiyal, P., Eley, G., Huddleston, K. C., …

Niederhuber, J. E. (2016). New observations on maternal age effect on germline de novo mutations. *Nature Communications*, *7*(May 2015), 10486. https://doi.org/10.1038/ncomms10486

Wu, C. I., & Li, W. H. (1985). Evidence for higher rates of nucleotide substitution in rodents than in man. *Proceedings of the National Academy of Sciences*, *82*(6), 1741–1745. https://doi.org/10.1073/pnas.82.6.1741

Zhang, P., Zucchelli, M., Bruce, S., Hambiliki, F., Stavreus-evers, A., & Levkov, L. (2009). Transcriptome Profiling of Human Pre-Implantation Development, *4*(11). https://doi.org/10.1371/journal.pone.0007844