Mutation rates and the evolution of germline structure

Aylwyn Scally
Department of Genetics, University of Cambridge, CB2 3EH, UK

Abstract

Genome sequencing studies of *de novo* mutations in humans have revealed surprising incongruities with our understanding of human germline mutation. In particular, the mutation rate observed in modern humans is substantially lower than that estimated from calibration against the fossil record, and the paternal age effect in mutations transmitted to offspring is much weaker than expected from our longstanding model of spermatogenesis. I consider possible explanations for these discrepancies, including evolutionary changes in life history parameters such as generation time and the age of puberty, a possible contribution from undetected post-zygotic mutations early in embryo development, and changes in cellular mutation processes at different stages of the germline. I suggest a revised model of stem cell state transitions during spermatogenesis, in which 'dark' gonial stem cells play a more active role than hitherto envisaged, with a long cycle time undetected in experimental observations. More generally I argue that the mutation rate and its evolution depend intimately on the structure of the germline in humans and other primates.

Keywords
Germline, mutation rate, human evolution

The germline mutation rate

All evolutionary processes depend on the flow of genetic information from one generation to the next, and as with any signal, errors in transmission can occur. The rate at which this happens is called the germline mutation rate, and is of central importance to evolutionary genetics. Not only is it key to interpreting genomic differences between individuals and populations, it also determines the timescale by which we can relate genetic data to other evidence for the evolutionary past. This relationship is not straightforward however, because although in evolutionary genetic theory the mutation rate often plays the role of a fundamental constant, in truth it evolves like any other trait, and can differ by orders of magnitude between species (1).

Estimates of the mutation rate in humans have varied according to the data and methods available. The first were made even before the nature of the DNA molecule had been established (2,3), and so were indirect and restricted to mutations causing phenotypic differences, such as at dominant disease loci. Subsequent estimates were based on phylogenetic comparisons between species, with divergence times drawn from fossil evidence. More recently, developments in genome sequencing technology have enabled mutation rate estimates based on counting *de novo* mutations, comparing closely related individuals in parent-offspring trios or larger pedigrees (reviewed in (4,5)).

In principle, phylogenetic and *de novo* estimates represent different aspects of the same approach, counting genetic differences accumulated over a number of generations. For evolutionary analyses a *de novo* estimate seems at first glance more attractive because it avoids the circularity implicit in phylogenetic calibration, particularly when comparing genetic data against fossil dates. However, the first such estimates in human trios yielded a value of 0.5×10^{-9} bp^{-1} yr^{-1} for single-nucleotide mutations, almost half the established phylogenetic rate, and thus implying a substantial lengthening of the evolutionary timescale if applied across all hominoid lineages (4). While such a revision may be warranted in places, particularly for recent events within the genus *Homo* and the speciation of the African great apes, a longer timescale for older events is difficult to reconcile with the primate fossil record. For example, with this rate the 4.7% sequence divergence between apes and old-world monkeys (6) implies a genetic divergence time 47 Myr ago, and hence speciation ~40 Myr ago (assuming a reasonably large ancestral population), whereas the fossil record seems consistent with a divergence no more than 25-30 Myr ago (7).

Several explanations for this disagreement have been proposed, including the possibility that *de novo* estimates have failed to correctly quantify false positives or inaccessible regions of the genome (5). However, while there are caveats to any approach, more than a dozen subsequent *de novo* studies have consistently produced similarly low values (5). This includes one study based on more distantly related individuals (8), and while other forthcoming pedigree-based estimates may lead to some adjustment (for reasons discussed below), it seems unlikely that methodological considerations alone will close the gap between phylogenetic and *de novo*-estimated rates. Furthermore, additional evidence supporting a low germline mutation rate in modern humans comes from comparisons of ancient and modern DNA (9), and a lower rate is arguably more compatible with archaeological evidence for the timing of recent events such as the divergence of Native American and East Asian populations (10).

This paper explores three alternative explanations for the rate discrepancy and discusses factors underlying the germline mutation rate which may have led to its evolution on shorter or longer timescales. Firstly, I discuss the possibility that mutation rates may have slowed due to life history changes during the last 20 million years of hominoid evolution. Secondly, I consider whether aspects of the cellular genealogy of the germline might have led to a substantial number of mutations going undetected in trio sequencing experiments. Finally, I discuss how stem cell processes in spermatogenesis affect the germline mutation rate and how our model for this might be reconciled with recent experimental observations.

Life history changes during hominoid evolution

One possible explanation for the discrepancy between mutation rate estimates is that rates themselves may have changed during hominoid evolution. Since they are observed to differ between species across large evolutionary distances, a slowdown on this scale is not implausible *a priori* (11,12). Indeed, great apes have evolved in several ways over this time, notably increasing in body mass (13). This itself leads to an explanation for the putative slowdown based on a change in generation time (defined as the average time from zygote to zygote along a genetic lineage), since life history parameters such as generation time scale with body mass across a wide range of mammal species (14,15). Consider a simplistic model where the per-generation mutation rate μ_{gen} is constant and the mutation rate per year μ scales inversely with generation time: $\mu = \mu_{\text{gen}} / t_{\text{gen}}$. Then an increase in generation time by a factor of almost two could account for the necessary reduction in yearly rate from $\sim 1 \times 10^{-9}$ bp^{-1} yr^{-1} in the past to 0.5×10^{-9} bp^{-1} yr^{-1} today (Figure 1).

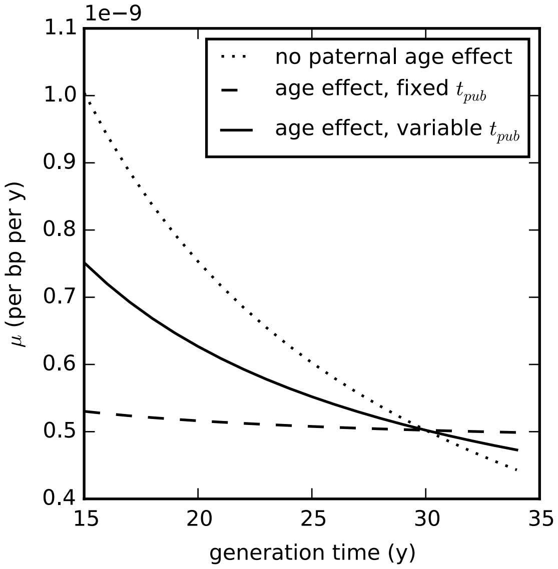

Figure 1: Models of human mutation rate slowdown with changing life history parameters. *Dotted*, simple scaling of mutation rate with generation time; *dashed*, including a paternal age effect but with fixed age of puberty; *solid*, including a paternal age effect and with age of puberty scaling with generation time, assuming $t_{\text{pub}} = 14$ yr when $t_{\text{gen}} = 30$ yr. Overall rates per bp are scaled to be 0.5×10^{-9} bp^{-1} yr^{-1} when $t_{\text{gen}} = 30$ yr.

However as previous studies have noted (5,16) , this model is too simplistic, for in supposing that the number of mutations per generation is independent of t_{gen} it ignores the fact that older fathers tend to pass on more mutations to their offspring than younger fathers. This phenomenon, the paternal age effect, is a consequence of the fact that cell-division replication errors are the major source of germline mutation, and whereas in both sexes there are several divisions associated with embryonic development prior to gametogenesis, spermatogenesis in males involves a process of continuous further cell division throughout reproductive life. Hence the older the father, the more cell divisions his gametes will have passed through, and the more errors accumulated. By contrast in oogensis a stock of primary oocytes is generated within the developing embryo, each of which is held in stasis until the final two meiotic divisions leading to ovulation later in life.

Empirical measurements of the paternal age effect in *de novo* sequencing studies have found that the mean number of mutations passed on by fathers grows linearly with age, roughly doubling between the ages of 20 and 40 years (17–19). This would seem to largely mitigate the generation time effect on mutation rates (5,16). Consider a straightforward extension to the model presented above: for an autosomal lineage (which spends equal time in males and females) we have $\mu_{\text{gen}} = (\mu_{\text{gen,f}} + \mu_{\text{gen,m}}) / 2$, where $\mu_{\text{gen,f}}$ is the female mutation rate per generation and $\mu_{\text{gen,m}}$ the male rate. Then

$$\mu_{\text{gen,m}} = \mu_{\text{dev,m}} + \mu_{\text{s}}(t_{\text{gen}} - t_{\text{pub}})$$

where $\mu_{\text{dev,m}}$ is the per-generation rate of mutations acquired during embryonic and juvenile development in males, μ_{s} is the yearly rate of mutation during spermatogenesis, and t_{pub} is the timing of puberty. We assume no age effect in mutations passed on by females, characterizing them by a single parameter $\mu_{\text{gen,f}}$. Parameters in this model can be taken from the experimental data of Kong *et al.* (2012) (17), who estimated whole-genome values of $\mu_{\text{gen,f}} = 14.2$ and $\mu_{\text{s}} = 2.01$ per yr; $\mu_{\text{dev,m}}$ is estimated by assuming the same mutation rate per cell division in males and females and that $\mu_{\text{gen,f}}$ and $\mu_{\text{dev,m}}$ correspond to 30 and 37 cell divisions respectively (see below for a discussion of these assumptions).

If t_{pub} is fixed then even a substantial change in generation time has relatively little effect on the yearly mutation rate under this model, as shown in Figure 1. However, the assumption of a fixed age of puberty is itself almost certainly invalid, since like other life history parameters the age of male sexual maturity scales with body mass across the primates, and variation within extant primates suggests a strong correlation ($R^2 = 0.84$) with t_{gen} (Figure 2). Assessments of sexual maturity can vary and may not coincide with the onset of spermatogenesis in every case (20). Nevertheless if we incorporate a linear scaling of t_{pub} with t_{gen} we recover much of the generation time effect, in the sense that an increase in t_{gen} from 15 to 30 yr now corresponds to a reduction in μ by a factor of 1.5 (Figure 1).

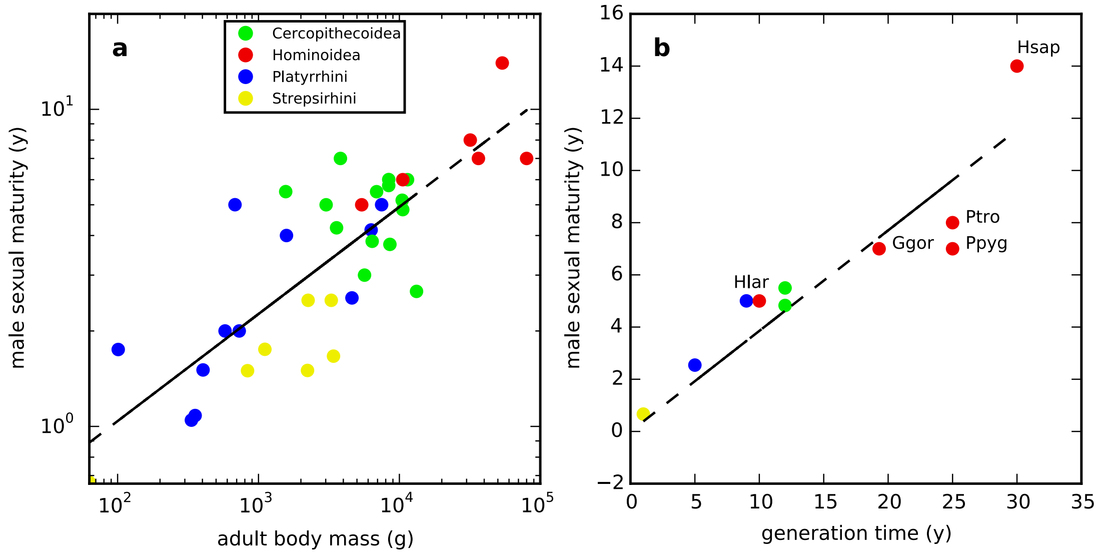

Figure 2. Variation of male age of sexual maturity with body mass and generation time in extant primates. (**a**) Variation with body mass suggests a scaling coefficient of 0.34. Data from (21,22). (**b**) A zero-intercept regression of age of sexual maturity on t_{gen} has slope 0.38. Data from (6,21,23–28). Hominoids are labeled: Hsap, human; Ptro, chimpanzee; Ggor, gorilla, Ppyg, orangutan; Hlar, gibbon.

Hidden germline mutations in trio sequencing

An alternative explanation for the discrepancy in rates may lie in how *de novo* sequencing experiments relate to the cellular genealogy of the germline, and the definition of germline mutation rate as the mean number of mutations acquired on a germline lineage from zygote to zygote. Mutations on somatic lineages are important in the context of diseases such as cancer, but such lineages do not as a rule extend beyond lifespan of the organism and thus make no direct contribution to evolutionary genetic processes. However, the detection of *de novo* mutations in trios is based on sequencing somatic cells in parents and offspring, not zygotes (or even other germ cells). To understand the implications of this and how these experiments relate to what we want to measure, we need to consider the cellular genealogy of the germline within a family (Figure 3).

Germ cell specification – the process by which certain cellular lineages are set aside as germ cells – occurs in mammals around the time of gastrulation. Following invagination of the epiblast, a number of cells originating there find a niche in the wall of the yolk sac and subsequently migrate as primordial germ cells (PGCs) to the gonadal ridge. Many somatic lineages also differentiate around the same time, and also have their origins within the epiblast. In humans this specification process occurs about two weeks after fertilization, or ~15 cell divisions (29). Thereafter, germ cell lineages undergo several further divisions in preparation for gametogenesis: approximately 15 more divisions in females and 20-24 in males (29,30). Thus in total there are about 30-40 mitotic divisions from fertilization to puberty, at which point in males the population of gonial stem cells (GSCs) is established in the testes, and the primary oocytes have been formed in females. From then on the male and female gametogenetic processes differ markedly, with GSCs replicating continuously in adult males to maintain the germ cell lineages and support gamete production.

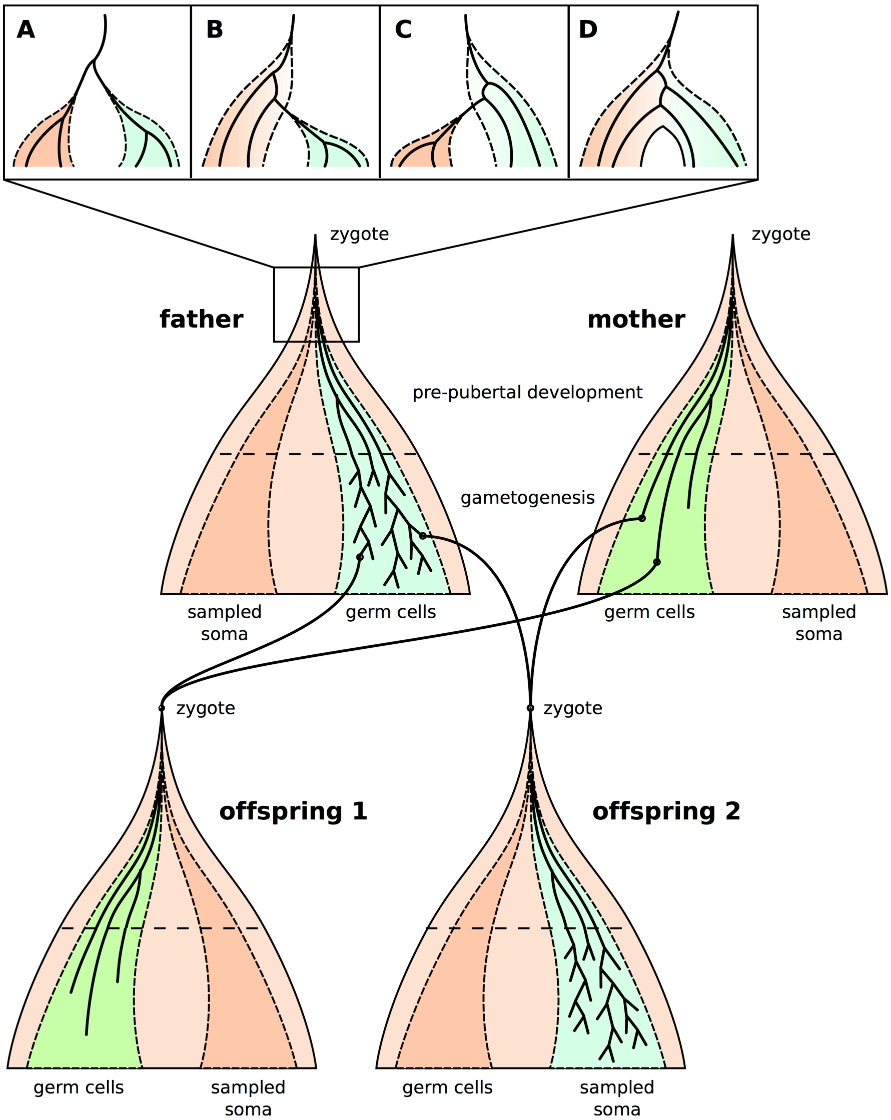
A
B
C
D
zygote
father
zygote
mother
pre-pubertal development
gametogenesis
sampled soma
germ cells
germ cells
sampled soma
zygote
zygote
offspring 1
offspring 2
germ cells
sampled soma
germ cells
sampled soma

Figure 3. Cellular genealogies in a two-offspring family. Solid black lines represent cellular lineages; germ cell populations are shaded green (in females) or blue (in males), somatic cells are shaded orange. Darker somatic populations represent the cellular ancestors of somatic cells sampled in a sequencing experiment. Panels A-D show possible configurations of germ and sampled somatic cellular lineages at the early post-zygotic stage: A, any cell ancestral to sampled somatic and germ cells is ancestral to all such cells; B, cells may be ancestral to all germ cells but only some sampled somatic cells; C, cells may be ancestral to all sampled somatic cells but only some germ cells; D, cells may be ancestral to some germ cells and some sampled somatic cells (meaning that some germ cells may be more closely related to some somatic cells than to other germ cells, and vice versa).

Given this structure, the fact that *de novo* sequencing estimates are based on sequencing somatic rather than germ cells creates a potential for error. For example, in comparing parents and offspring, mutations arising early in the somatic cellular genealogy of the offspring may be counted as *de novo* germline mutations (false positive errors), and early mutations in either parent on lineages having both somatic and germline descendants may not be recognized as *de novo* (false negatives).

Some of these cases may be excluded or recovered by careful filtering based on the fraction of somatic cells in which they are present (31). However there may be a class of early post-zygotic mutations which cannot be accounted for in this way, depending on when and how the divergence of germ cell and sequenced somatic lineages occurs. Prior to the completion of this divergence, early embryonic cells may be ancestral both to germ and somatic cells within the organism, and mutations occurring then may be shared by some or all cells in either genealogy (Figure 3). Such 'hidden' mutations could contribute a component to the germline mutation rate which is undetectable in parent-offspring comparisons, and whose size depends on the number of cell divisions and the per-cell division mutation rate at this early stage (32). For example, it has been suggested that the first few post-zygotic cell divisions might be particularly mutagenic, based on the high level of chromosomal instability and other errors found in early IVF embryos and the frequency of early pregnancy loss after conception (33–35).

The potential for hidden mutations depends on the distribution of cell fates within the epiblast (for which much of our understanding derives from studies in mice or non-human primates). It may also depend to some extent on which somatic cells are sequenced. For example, compared to cells sampled from multiple tissues or from blood, cells derived only from one tissue or region of the body may descend from a smaller number of lineages at any given stage in development. As a consequence, configurations A and C for the divergence of germ and somatic lineages in Figure 3 may be more likely for such cells, potentially increasing the number of early cell divisions in which mutations would be hidden. As an aside, the observation that parental mosaicism in blood is correlated with recurrence risk (36) suggests that lineage ancestries for these cells at least are mixed in humans (case D in Figure 3) (37). Lineage tracing experiments on mouse oocytes suggest that a similar situation exists in mice across a range of somatic cell types, notwithstanding a degree of lineage clustering by cell type (38).

Might a hidden mutation component explain the discrepancy between phylogenetic and *de novo* rate estimates? Various considerations suggest that this is unlikely, subject to further data. Firstly, although hidden mutations are impossible to detect in single-generation experiments, comparisons over many generations should be sensitive to mutations on all 'internal' segments, including all hidden mutations except at the root and leaves of the pedigree. If hidden mutations make a substantial contribution to the germline mutation rate, we might expect pedigree-based estimates to be higher than those made in trios. Two such studies have been published to date, of which one did not differ significantly from trio-based estimates (8) and the other obtained a value 33% higher (39). Forthcoming studies may clarify this picture.

Secondly, a large hidden contribution should lead to a correspondingly high rate of within-family recurrence of genetic diseases associated with *de novo* mutations. This would be in addition to rates of recurrence due to shared gametic ancestry following germ cell specification, for which previous models have estimated recurrence rates of <<1% for mutations of paternal origin (i.e. most mutations) and ~4% for those of maternal origin (36,40,41). By definition, hidden mutations occurring in a parent will be present in all of his or her gametes. Thus if hidden mutations constitute a fraction ϕ of all germline mutations, the probability of recurrence due to such mutations will be ϕR, where R is sibling relatedness. If hidden mutations occurred at a rate similar to the observed *de novo* mutation rate, so that the total *de novo* rate matched the phylogenetic rate, we would expect to see at least 25% recurrence of autosomal-linked diseases ($R = 0.5$). Recurrence at sex-linked loci depends on the sexes of offspring, but even in brothers of female offspring we would expect a recurrence rate of at least 12.5% for diseases caused by *de novo* mutations on the X chromosome (R = 0.25).

Clinically estimated recurrence rates depend on the disorder involved and the nature of the causative mutation or mutations. Some disorders such as Duchene muscular dystrophy show recurrence rates as high as 14% (42), but estimates are generally less than 1% (43,44). However, such estimates may not necessarily reflect the recurrence rates of single-nucleotide variants (SNVs) as counted in *de novo* sequencing experiments and in phylogenetic comparisons. Even where they relate to clinical genetic data, such data often include structural mutations and chromosomal abnormalities whose origins may tend to differ from those of *de novo* SNVs. In particular, chromosomal rearrangements may be enriched for meiotic errors (45), whereas the apparent linearity of the paternal age effect suggests that germline SNVs are dominated by mitotic events. Where clinical estimates are based on phenotypic recurrence, uncertainty arises in modeling the relationship between genotype and phenotype, the number of loci involved and controlling for possible environmental factors. Additionally, some phenotypes may be difficult to diagnose consistently, particularly where there is already a diagnosis in siblings, and further potential bias arises from stoppage, whereby parents of an affected child are less likely to have additional children (46,47). Some of these considerations suggest that clinical estimates might underestimate the true recurrence of *de novo* germline mutations. However, the degree of underestimation would have to be at least an order of magnitude to be consistent with a substantial contribution of hidden mutations to the germline mutation rate.

Another effect of hidden mutations would be to inflate the male-female mutation rate ratio as measured in trio comparisons. If hidden mutations occur with equal probability in males and females, and if the male-female ratio of observed (non-hidden) mutations is α_{obs}, it is straightforward to show that the true male-female ratio α is bounded above by α_{obs} and given by

$$\alpha = ((1 - \alpha_{\text{obs}})\phi + 2\alpha_{\text{obs}}) / (2 - (1 - \alpha_{\text{obs}})\phi)$$

The value α_{obs} estimated in trios by Kong *et al.* (2012) (17) was 3.9. Alternative estimates based on comparing X-chromosome and autosomal genetic diversity within or between species (which should reflect all germline mutations) have typically fallen in the range 3–7 (48). Values at the lower end of this range could suggest a moderate contribution from hidden mutations (rearranging the above expression, $\alpha = 3$ and $\alpha_{\text{obs}} = 3.9$ implies $\phi = 0.15$), but a value higher than α_{obs} would be inconsistent with the model presented here. In this case it may be that hidden mutations are not equally likely in males and females, which would be surprising given that most or all of the cell divisions involved occur prior to the onset of somatic sex differentiation in the embryo. However there are also several factors which can substantially bias the rate estimated from X-autosomal comparisons, including selection, sex-biased demography and differences between male and female generation times (49–51). Moreover, just as the overall germline mutation rate may have varied over evolutionary timescales, so too might the gametogenetic factors contributing to male mutation bias, which would further affect estimates based on genetic diversity.

Changes in the structure of spermatogenesis

The importance of paternally transmitted mutations focuses attention on spermatogenesis as a key factor affecting the germline mutation rate. The established model of human spermatogenesis is based on long-standing experimental observations of the seminiferous epithelium (the environment within the testes where spermatogenesis occurs) (52,53). Yves Clermont observed several types of spermatogonial cell in humans, differing in their appearance and degree of staining with hematoxylin and eosin (54). Two of these types correspond to self-renewing (GSC) states (55), and based on their staining are generally referred to as pale (A_p) and dark (A_d) spermatogonia. However in Clermont's observations only A_p cells were seen to actively divide, each doing so every 16 days to produce a new A_p

cell and a progenitor spermatocyte which he termed B-spermatogonia. The latter subsequently undergo two further mitotic divisions and meiosis to produce up to 32 spermatozoa (55) in a process lasting 48 days. Thus A_p cells are widely regarded as the active spermatogenetic population, and A_d cells are thought to comprise a pool of reserve stem cells, to be drawn upon only when the active population has failed or is damaged. Figure 4 (model 1) illustrates this model in terms of the cell states and transitions involved.

One possible explanation for a slowdown in mutation rate would therefore be an increase in the cycle time of the seminiferous epithelium during hominoid evolution, leading to fewer mutations acquired during spermatogenesis for a typical adult male. Such a change is equivalent to varying μ_s in the model discussed above, and Figure 5 shows the effect on germline mutation rate, assuming that puberty also scales with generation time as previously discussed. The seminiferous epithelial cycle time in monkeys varies between 9 and 11 days (56), and if the cycle time in ancestral great apes was similar to this, the change since then would correspond to a mutation rate slowdown by almost a factor of two (dashed line in Figure 5), and perhaps sufficient to explain the discrepancy in mutation rates.

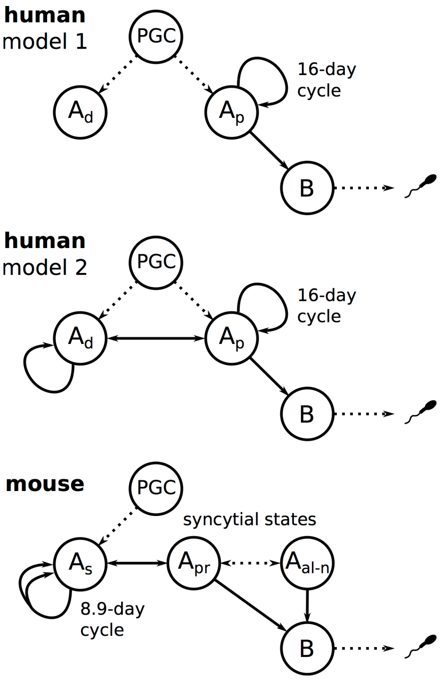

Figure 4. Cell states and possible transitions in models of human and mouse spermatogenesis, from primordial germ cell (PGC) to spermiogenesis. Dotted lines represent transitions involving one or more non-replicating intermediate states (for example, the initial transition from PGC to A_p is sometimes thought to pass through a temporary A_d state). Two human models are shown: model 1 is the established model originally due to Clermont (52), and model 2 an alternative model discussed in the text. In the mouse model, the double arrow from A_s to itself indicates that cell division is symmetric: $A_s \rightarrow A_s + A_s$.

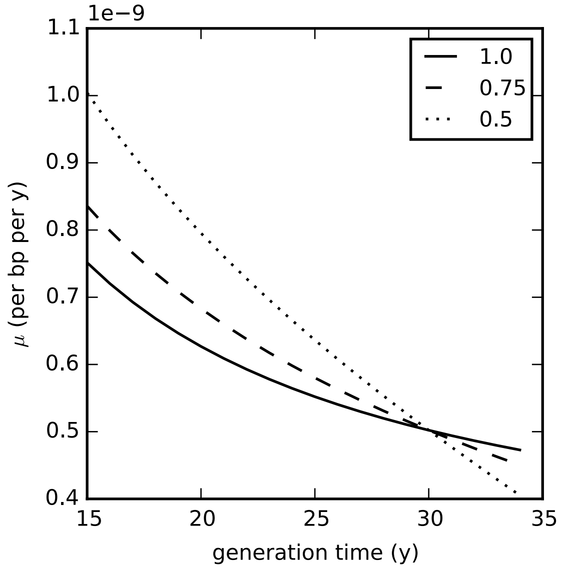

Figure 5: Models of human mutation rate slowdown in which the seminiferous epithelial cycle time increases with generation time and age of puberty. *Dotted line*, cycle time is 0.5 times the present length when generation time is 15 y; *dashed line*, cycle time is 0.75 times the present length when generation time is 15 y; *solid line*, no change in cycle time (identical to the solid line in Figure 1). Overall rates per bp are scaled to be 0.5×10^{-9} bp^{-1} y^{-1} when t_{gen} = 30 y.

However there is a problem with this model as presented, in that lineages resulting in gametes produced by a 30-year-old male would have passed through ~400 cell divisions since fertilization, meaning we would expect a roughly tenfold increase in the number of mutations passed on to offspring at age 30 compared to those at puberty. Yet sequencing studies have consistently measured only a twofold increase from early to late adulthood (17–19). This discrepancy, noted also by Ségurel *et al.* (5), suggests that aspects of the model for spermatogenesis need to be revised.

One possibility is that per-cell-division mutation rates are much higher at earlier developmental stages than during spermatogenesis. For example, there is evidence that somatic cell division mutation rates are substantially higher than those in germ cells (57) and it may be that changes in the environment or phenotype of cells at germ cell specification are accompanied by improved mechanisms of DNA replication error correction. In order to account for a weak paternal age effect, the mutation rate in the first 15 cell divisions (prior to specification) would need to be a factor of ~20 higher than in subsequent divisions, which is near the limit of the range of reported estimates for germline and somatic cells (1). Note also that this excludes any hidden mutations of the kind discussed above, although such mutations would also be made more likely by an elevated post-zygotic mutation rate. Alternatively, elevated per-cell-division rates may last for a longer time, perhaps until the onset of spermatogenesis or shortly thereafter. If we assume a higher rate applies to the first 40 divisions then it need only be higher by a factor of ~9. A recent study of transmitted mutations in a large cohort including some teenage fathers suggested that the very earliest stage of spermatogenesis may be more mutagenic (19). If true, this might reflect a shift to lower mutation rates once spermatogenesis is fully established. Early spermatogenesis is known to differ in some respects from the process later on: for example daily sperm production volumes are ~10 times lower in teenage males than in men 20-30 years old (58).

Another possibility is that the apparent 16-day cycle of the seminiferous epithelium is only part of the picture and that germline lineages actually experience a longer cell cycle time for some or all of their passage through spermatogenesis. This would imply a more complex structure for GSC state transitions and the number of self-renewing states in which they can exist. It is of course possible to conceive of many such models, but a relatively straightforward extension of the existing model would be for the A_d cells to play a more prominent role. If they replicate with a longer cycle time than was detectable in Clermont's data, and if transitions are possible between the A_d and A_p states, then germline lineages could spend some or even most of their time in the A_d state during spermatogenesis (Figure 4, model 2). By reducing the number of germline cell divisions, this could account for a weaker than expected paternal age effect.

Potential evidence for such a model comes from several sources. Within primates, investigations of spermatogonial renewal in monkeys after exposure to radioactive or contraceptive agents (59,60) have shown that A_p cells may be able to transition to A_d without undergoing cell division. If this occurs under normal conditions then the A_d state could play a role other than that of a dormant and nonproliferative reserve. Other evidence comes from comparison with spermatogenesis in mice, which although differing in several respects does share many basic features with primates (61) (Figure 4). GSCs in mice can exist in a singular state (A_s) or in various syncytial states wherein the nuclei share a common cytoplasm, either paired (state A_{pr}) or in longer alignments of *n* cells (states $A_{al\text{-}n}$) (62). Recent experimental studies suggest that progenitor spermatocytes may be produced from divisions of any of these states (61,63), but the degree of commitment to (or likelihood of) differentiation may be greater in the A_{al} state. An analogy can be drawn with the process in humans, based both on function and on expression levels of several molecular markers, in which the A_s state corresponds to A_d and the A_{pr} and A_{al} states correspond to A_p, (61,64). Moreover, a model of stochastic transition between the A_s, A_{pr} and A_{al} states, in which intracellular bridges are broken as well as created, has been shown to fit the dynamics of GSC populations in mice (63,64). The analogous set of transitions in humans and other primates would fit the alternative model shown in Figure 4.

By making certain assumptions about the possible transitions between GSC states, we can estimate what cycle time for the A_d state would produce the observed paternal age effect. Previous approaches to modelling stem cell systems have sometimes assumed homeostatic equilibrium as a way of estimating or constraining model parameters (e.g. (63)), but it may be that spermatogenesis is better represented as a near-equilibrium process. For example there is notable variation in daily sperm production (DSP) with age: in humans mean DSP decreases steadily in older men, dropping by a factor of two from age 20 to 80 years (58). To capture these non-equilibrium aspects we can use a finite state simulation in which transitions between cell states have both a probability parameter and an associated transition time (Figure 6). The basic assumption of this model is that the transition $A_p \rightarrow A_d$ occurs with some probability during each A_p cycle. For the reverse transition, since A_p is observed to divide asymmetrically ($A_p \rightarrow A_p + S$) I assume for now that A_d behaves similarly ($A_d \rightarrow A_d + A_p$); an alternative choice ($A_d \rightarrow A_p$) is discussed below. I also introduce a cell death state which both regulates the process (since otherwise A_d replication will lead to unbounded GSC proliferation) and ensures that gamete production declines with age. Cell death likely plays a role in regulating many stem cell systems (65), and in spermatogenesis GSC replication occurs only within a niche at the basement membrane of a seminiferous tubule for which several cellular and other environmental factors may be essential. In particular the availability of Sertoli cells, somatic cells which play both a structural and regulatory role in gamete production, is thought to be a critical factor (53,58). Cell death probabilities in this model are an abstraction representing the typical availability of and contention for these critical factors.

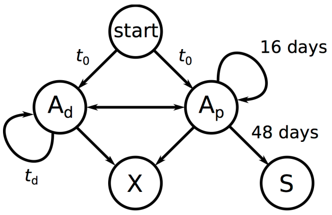

Figure 6. A possible finite state model for spermatogenesis. Germline lineages are traced through the model, entering at the START state, from which they can transition to either of the GSC states A_d or A_p. The cell death state (X) can be reached from either GSC state, whereas we assume here that spermiogenesis (S) is possible only from A_p. Transitions from the GSC states correspond to cell divisions, so the model describes a branching process representing an entire genealogy of spermatozoa, each descended from a single primordial germ cell (PGC) and having an associated age and total number of cell divisions (depth) from the root of the genealogy. In the established model (model 1 in Figure 4), the transition $A_p \rightarrow A_d$ has zero probability under normal conditions.

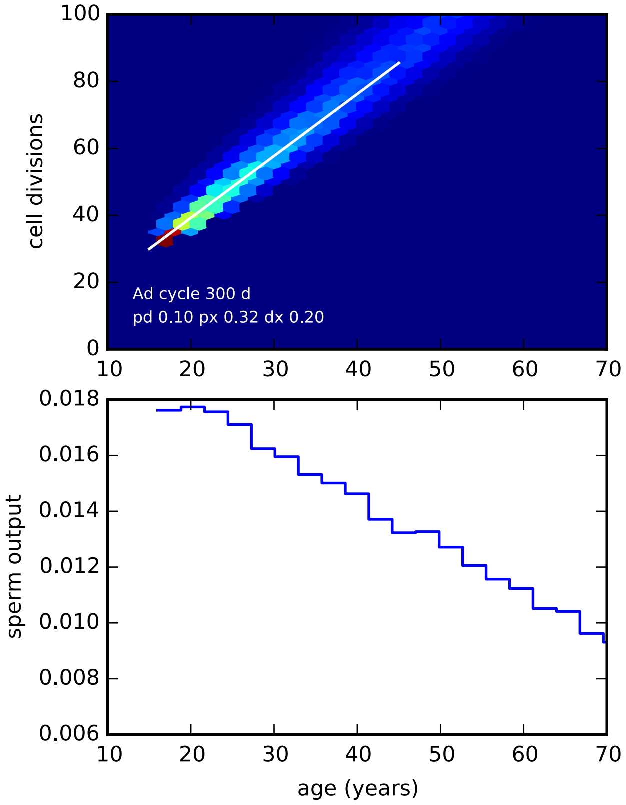

Figure 7. Output from an implementation of the model in Figure 6, for parameters fitting the observed paternal age effect and decline in sperm production. (Code available on request.) *Upper panel***:** Joint distribution of paternal age and number of cell divisions since the zygote in simulated spermatozoa (colours indicate relative frequency from zero (dark blue) to high (red)) for a set of parameters in which the A_d cycle time t_d = 300 days and initial transition times t_0 were sampled uniformly from the range 5000–5500 days, $P(A_p \rightarrow A_d) = 0.1$, $P(A_p \rightarrow X) = 0.32$, $P(A_d \rightarrow X) = 0.2$. The white line shows the slope of the paternal age effect observed in humans. *Lower panel*: Simulated relative daily sperm output as a function of age.

A simple parameter-space search, fitting simulated output of this model to the observed slopes of the paternal age effect and DSP age profile, estimates an A_d cycle time of around 300 days and values of 20-30% for transition probabilities to the cell death state (Figure 7). Replication of A_d cells on this timescale would likely not have been observed in Clermont's or subsequent experiments, although transitions $A_p \rightarrow A_d$, which in this example are predicted to occur in 10% of A_p cycles, might be detectable.

Other model assumptions are possible, and may result in different parameter estimates. For example, in a model where the $A_d \rightarrow A_p$ transition occurs without cell division, so that the A_d cycle is essentially a quiescent interlude before GSCs return to the active A_p state, a similar procedure fitting the observed paternal age effect estimates a cycle time of around 750 days (data not shown). However the point here is not that parameters can be inferred under one model or another, but the fact that including an active role for the A_d cells allows models which are compatible both with longstanding observations of the seminiferous epithelium and with recent measurements of the paternal age effect. More sophisticated models might also include feedback or global regulation mechanisms other than cell death (53,56), age-related changes in cell-division mutation rate and spermiogenetic efficiency, and perhaps phenomena such as selfish spermatogonial selection (66). At present experimental data is limited, but more extensive data including trio sequencing on population-wide scales will provide a better basis for exploration of spermatogenetic models along these or similar lines.

Discussion

I have argued that the discrepancy between phylogenetic and *de novo* estimates of the mutation rate is more likely due to a genuine evolutionary slowdown than to methodological errors or the failure of trio sequencing experiments to detect early post-zygotic mutations. Nevertheless, the latter factors may be present at some level and thus contribute to the discrepancy, meaning that the magnitude of any slowdown may be less than was first hypothesized. Also, while we may be confident that rates have slowed at some point during primate evolution, our inference of the timing, extent and number of ancestral lineages involved in such a slowdown is determined largely by the fossil record and the confidence with which we can constrain speciation events, particularly within the hominoids. Initial attempts to reconcile the rate discrepancy were concerned with the plausibility of a slowdown affecting all four great ape lineages in parallel and to the same degree, given that their branch lengths from an outgroup such as macaque do not differ substantially (50). However, if newer interpretations of the fossil record were to admit a more ancient speciation time of 20 Mya or more between the ancestors of orangutans and other great apes, they would be consistent with an earlier slowdown affecting only the stem hominoid lineage, and this would suffice to explain the current data without requiring parallel evolution.

More broadly, and regardless of the extent to which rates may have changed in recent primate evolution, the processes considered here are relevant to evolutionary genetic analyses across the mammalian tree and beyond. Previous studies have proposed life history variation as an explanation for mutation rate change, but it is clear that such explanations need to involve more biologically sophisticated models incorporating factors such as varying pubertal age and sex-dependent parameters (5,51). Mutation rate change may also be due to evolution in the underlying cellular processes and genealogical structure of the germline, particularly in gametogenesis. Here too, recent experimental data are incongruous with existing models of spermatogenesis and the strength of the paternal age effect. I have focused on potential variation in cell-division mutation rates at different developmental stages and on the stem cell states involved in spermatogenesis. Other issues not touched on include the relative importance of spontaneous mutation processes (67), potential evidence for a maternal age effect (68), and the evolution of regulatory factors controlling gametogenesis (53). Progress to date in addressing these questions has been difficult in part because of the challenge of obtaining experimental data on human germline processes: some techniques can only be applied to non-human models, and genome sequence data for human *de novo* mutations has previously been limited. However the potential now exists for large-scale genome sequencing of somatic and germ cells and experimental studies of non-human and human stem cell systems. In combination with computational modeling approaches such as that presented here (and widely used in previous studies to explore stem cell population dynamics (36,63,69–71)), these developments will facilitate a better understanding of mutation processes and the evolution of the human germline.

References

1. Lynch M. Evolution of the mutation rate. *Trends Genet*. 2010 Aug 30; 26(8):345–52.
2. Haldane J. The rate of spontaneous mutation of a human gene. *J Genet*. 1935; (3)
3. Nachman MW. Haldane and the first estimates of the human mutation rate. *J Genet*. 2004 Dec; 83(3):231–3.
4. Scally A, Durbin R. Revising the human mutation rate: implications for understanding human evolution. *Nat Rev Genet*. 2012 Sep 11; 13(10):745–53.
5. Ségurel L, Wyman MJ, Przeworski M. Determinants of Mutation Rate Variation in the Human Germline. *Annu Rev Genomics Hum Genet*. 2014 Aug 31; 15(1):47–70.
6. Carbone L, Alan Harris R, Gnerre S, Veeramah KR, Lorente-Galdos B, Huddleston J, *et al.* Gibbon genome and the fast karyotype evolution of small apes. *Nature*. 2014 Sep 10;513(7517):195–201.
7. Stevens NJ, Seiffert ER, O'Connor PM, Roberts EM, Schmitz MD, Krause C, *et al.* Palaeontological evidence for an Oligocene divergence between Old World monkeys and apes. *Nature*. 2013 May 15;497(7451):611–4.
8. Campbell CD, Chong JX, Malig M, Ko A, Dumont BL, Han L, *et al.* Estimating the human mutation rate using autozygosity in a founder population. *Nat Genet*. 2012 Sep 23;44(11):1277–81.
9. Fu Q, Li H, Moorjani P, Jay F, Slepchenko SM, Bondarev A a., *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature*. 2014 Oct 22; 514(7523):445–9.
10. Schiffels S, Durbin R. Inferring human population size and separation history from multiple genome sequences. *Nat Genet*. 2014 Aug; 46(8):919–25.
11. Wu CI, Li WH. Evidence for higher rates of nucleotide substitution in rodents than in man. *Proc Natl Acad Sci*. 1985 Mar 1; 82(6):1741–5.
12. Wilson Sayres MA, Venditti C, Pagel M, Makova KD. Do variations in substitution rates and male mutation bias correlate with life-history traits? A study of 32 mammalian genomes. *Evolution*. 2011 Oct; 65(10):2800–15.
13. Fleagle JG. *Primate Adaptation and Evolution*, Second Edition. Academic Press; 1998.
14. Jones JH. Primates and the Evolution of Long, Slow Life Histories. *Curr Biol*. 2011 Sep; 21(18):R708–17.
15. de Magalhaes JP, Costa J, Church GM. An Analysis of the Relationship Between Metabolism, Developmental Schedules, and Longevity Using Phylogenetic Independent Contrasts. *Journals Gerontol Ser A Biol Sci Med Sci*. 2007 Feb 1;62(2):149–60.
16. Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, *et al.* Rate of de novo mutations and the importance of father's age to disease risk. *Nature*. 2012 Aug 22; 488(7412):471–5.
17. Thomas GWC, Hahn MW. The Human Mutation Rate Is Increasing, Even As It Slows. *Mol Biol Evol*. 2014 Feb 1; 31(2):253–7.
18. Sun JX, Helgason A, Masson G, Ebenesersdóttir SS, Li H, Mallick S, *et al.* A direct characterization of human mutation based on microsatellites. *Nat Genet*. 2012 Aug 23; 44(10):1161–5.
19. Forster P, Hohoff C, Dunkelmann B, Schurenkamp M, Pfeiffer H, Neuhuber F, *et al.* Elevated germline mutation rate in teenage fathers. *Proc R Soc B Biol Sci*. 2015 Mar 22; 282(1803):20142898–20142898.

20. Luetjens CM, Weinbauer GF. Functional assessment of sexual maturity in male macaques (*Macaca fascicularis*). *Regul Toxicol Pharmacol*. 2012 Aug; 63(3):391–400.
21. Saltzman W, Tardif SD, Rutherford JN. Hormones and Reproductive Cycles in Primates. *Hormones and Reproduction of Vertebrates*. Elsevier; 2011. p. 291–327.
22. Kingsley SR. Physiological development of male orang-utans and gorillas. *Orang-utan Biology*. New York, NY: Oxford University Press; 1988. p. 123–31.
23. Langergraber KE, Prufer K, Rowney C, Boesch C, Crockford C, Fawcett K, *et al.* Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *Proc Natl Acad Sci*. 2012 Aug 13; 1–6.
24. Molur S, Brandon-Jones D, W. Dittus W, Eudey A, Kumar A, Singh M, *et al. Status of South Asian Primates: Conservation Assessment and Management Plan (C.A.M.P.) workshop report*. 2003. viii+432 p.
25. Rogers J, Kidd KK. Nucleotide polymorphism, effective population size, and dispersal distances in the yellow baboons (*Papio hamadryas cynocephalus*) of Mikumi National Park, Tanzania. *Am J Primatol*. 1996; 38(2):157–68.
26. Milton K, Lozier JD, Lacey E a. Genetic structure of an isolated population of mantled howler monkeys (*Alouatta palliata*) on Barro Colorado Island, Panama. *Conserv Genet*. 2009 Apr; 10(2):347–58.
27. Tardif S, Carville A, Elmore D, Williams LE, Rice K. Reproduction and breeding of nonhuman primates. *Nonhuman Primates in Biomedical Research*. 2012. p. 197–249.
28. Schneider N, Chikhi L, Currat M, Radespiel U. Signals of recent spatial expansions in the grey mouse lemur (Microcebus murinus). *BMC Evol Biol*. 2010; 10(1):105.
29. Drost JB, Lee WR. Biological basis of germline mutation: Comparisons of spontaneous germline mutation rates among drosophila, mouse, and human. *Environ Mol Mutagen*. 1995; 25(S2):48–64.
30. Crow JF. The origins, patterns and implications of human spontaneous mutation. *Nat Rev Genet*. 2000 Oct;1(1):40–7.
31. Acuna-Hidalgo R, Bo T, Kwint MP, van de Vorst M, Pinelli M, Veltman JA, *et al.* Post-zygotic Point Mutations Are an Underrecognized Source of De Novo Genomic Variation. *Am J Hum Genet*. 2015 Jul;97(1):67–74.
32. Harland C. Frequency of mosaicism points towards mutation prone early cleavage cell divisions. Biology of Genomes. Cold Spring Harbor; 2015.
33. Vanneste E, Voet T, Le Caignec C, Ampe M, Konings P, Melotte C, *et al.* Chromosome instability is common in human cleavage-stage embryos. *Nat Med*. 2009 May 26; 15(5):577–83.
34. Chow JF, Yeung WS, Lau EY, Lee VC, Ng EH, Ho P-C. Array comparative genomic hybridization analyses of all blastomeres of a cohort of embryos from young IVF patients revealed significant contribution of mitotic errors to embryo mosaicism at the cleavage stage. *Reprod Biol Endocrinol*. 2014; 12(1):105.
35. Macklon NS, Geraedts JPM, Fauser BCJM. Conception to ongoing pregnancy: the “black box” of early pregnancy loss. *Hum Reprod Update*. 2002 Jul 1; 8(4):333–43.
36. Campbell IM, Stewart JR, James RA, Lupski JR, Stankiewicz P, Olofsson P, *et al.* Parent of Origin, Mosaicism, and Recurrence Risk: Probabilistic Modeling Explains the Broken Symmetry of Transmission Genetics. *Am J Hum Genet*. 2014 Oct; 95(4):345–59.
37. Samuels M, Friedman J. Genetic Mosaics and the Germ Line Lineage. *Genes*. 2015 Apr 17; 6(2):216–37.
38. Reizel Y, Itzkovitz S, Adar R, Elbaz J, Jinich A, Chapal-Ilani N, *et al.* Cell Lineage Analysis of the Mammalian Female Germline. Barsh GS, editor. *PLoS Genet*. 2012 Feb 23; 8(2):e1002477.
39. Palamara PF, Francioli LC, Wilton PR, Genovese G, Gusev A, Finucane HK, *et al*. Leveraging Distant Relatedness to Quantify Human Mutation and Gene-Conversion Rates. *Am J Hum Genet*. 2015; 97(6):775–89.
40. Hartl DL. Recurrence risks for germinal mosaics. *Am J Hum Genet*. 1971 Mar; 23(2):124–34.

41. Wijsman EM. Recurrence risk of a new dominant mutation in children of unaffected parents. *Am J Hum Genet*. 1991 Apr; 48(4):654–61.

42. Bakker E, Veenema H, Den Dunnen JT, van Broeckhoven C, Grootscholten PM, Bonten EJ, *et al.* Germinal mosaicism increases the recurrence risk for "new" Duchenne muscular dystrophy mutations. *J Med Genet*. 1989 Sep 1; 26(9):553–9.

43. Röthlisberger B, Kotzot D. Recurrence risk in de novo structural chromosomal rearrangements. *Am J Med Genet Part A*. 2007 Aug 1; 143A(15):1708–14.

44. McGaughran J, Sinnott S, Moal FD-L, Wilson M, Mowat D, Sutton B, *et al.* Recurrence of Mowat-Wilson syndrome in siblings with the same proven mutation. *Am J Med Genet Part A*. 2005 Sep 1; 137A(3):302–4.

45. Poot M, Haaf T. Mechanisms of Origin, Phenotypic Effects and Diagnostic Implications of Complex Chromosome Rearrangements. *Mol Syndromol*. 2015 Aug 15; 6(3):110–34.

46. Guo S-W. Inflation of Sibling Recurrence-Risk Ratio, Due to Ascertainment Bias and/or Overreporting. *Am J Hum Genet*. 1998 Jul; 63(1):252–8.

47. Slager SL, Foroud T, Haghighi F, Spence MA, Hodge SE. Stoppage: An issue for segregation analysis. *Genet Epidemiol*. 2001 Apr; 20(3):328–39.

48. Wilson Sayres M A., Makova KD. Genome analyses substantiate male mutation bias in many species. *BioEssays*. 2011 Dec; 33(12):938–45.

49. Vicoso B, Charlesworth B. Evolution on the X chromosome: unusual patterns and processes. *Nat Rev Genet*. 2006 Aug 1; 7(8):645–53.

50. Scally A, Dutheil JY, Hillier LW, Jordan GE, Goodhead I, Herrero J, *et al.* Insights into hominid evolution from the gorilla genome sequence. *Nature*. 2012 Mar 8; 483(7388):169–75.

51. Amster G, Sella G. Life history effects on the molecular clock of autosomes and sex chromosomes. *bioRxiv*. 2015 Aug.

52. Clermont Y. Two classes of spermatogonial stem cells in the monkey (Cercopithecus aethiops). *Am J Anat*. 1969 Sep; 126(1):57–71.

53. Schlatt S, Ehmcke J. Regulation of spermatogenesis: An evolutionary biologist's perspective. *Semin Cell Dev Biol*. 2014 May; 29:2–16.

54. Clermont Y. The cycle of the seminiferous epithelium in man. *Am J Anat*. 1963 Jan;112(1):35–51.

55. Ehmcke J, Schlatt S. A revised model for spermatogonial expansion in man: lessons from non-human primates. *Reproduction*. 2006 Nov 1; 132(5):673–80.

56. Luetjens CM, Weinbauer GF, Wistuba J. Primate spermatogenesis: new insights into comparative testicular organisation, spermatogenic efficiency and endocrine control. *Biol Rev Camb Philos Soc*. 2005; 80(3):475–88.

57. Lynch M. Rate, molecular spectrum, and consequences of human mutation. *Proc Natl Acad Sci*. 2010 Jan 19; 107(3):961–8.

58. Amann RP. The Cycle of the Seminiferous Epithelium in Humans: A Need to Revisit? *J Androl*. 2008 Apr 17; 29(5):469–87.

59. Fouquet JP. Renewal of spermatogonia in the monkey (Macaca fascicularis). *Biol Reprod*. 1986 Aug 1; 35(1):199–207.

60. O'Donnell L. Impairment of Spermatogonial Development and Spermiation after Testosterone-Induced Gonadotropin Suppression in Adult Monkeys (Macaca fascicularis). *J Clin Endocrinol Metab*. 2001 Apr 1; 86(4):1814–22.

61. Hermann BP, Sukhwani M, Hansel MC, Orwig KE. Spermatogonial stem cells in higher primates: are there differences from those in rodents? *Reproduction*. 2010 Mar 1; 139(3):479–93.

62. de Rooij DG, Griswold MD. Questions About Spermatogonia Posed and Answered Since 2000. *J Androl*. 2012 Nov 1; 33(6):1085–95.

63. Hara K, Nakagawa T, Enomoto H, Suzuki M, Yamamoto M, Simons BD, *et al.* Mouse Spermatogenic Stem Cells Continually Interconvert between Equipotent Singly Isolated and Syncytial States. *Cell Stem Cell*. 2014 May; 14(5):658–72.

64. Krieger T, Simons BD. Dynamic stem cell heterogeneity. *Development*. 2015 Apr 15; 142(8):1396–406.

65. Print CG, Loveland KL. Germ cell suicide: new insights into apoptosis during spermatogenesis. *BioEssays*. 2000 Apr 24; 22(5):423–30.

66. Maher GJ, Goriely A, Wilkie OM. Cellular evidence for selfish spermatogonial selection in aged human testes. *Andrology*. 2014 May; 2(3):304–14.

67. Gao Z, Wyman MJ, Sella G, Przeworski M. Interpreting the dependence of mutation rates on age and time. *bioRxiv*. 2015 Jul 24;

68. Dunlop CE, Telfer EE, Anderson R a. Ovarian germline stem cells. *Stem Cell Res Ther*. 2014 Jan; 5(4):98.

69. Qin J, Calabrese P, Tiemann-Boege I, Shinde DN, Yoon S-R, Gelfand D, *et al.* The Molecular Anatomy of Spontaneous Germline Mutations in Human Testes. *PLoS Biol*. 2007 Sep; 5(9):e224.

70. Giannoulatou E, McVean G, Taylor IB, McGowan SJ, Maher GJ, Iqbal Z, *et al.* Contributions of intrinsic mutation rate and selfish selection to levels of de novo HRAS mutations in the paternal germline. *Proc Natl Acad Sci*. 2013 Dec 10; 110(50):20152–7.

71. de Rooij DG, van Beek ME a B. Computer Simulation of the Rodent Spermatogonial Stem Cell Niche. *Biol Reprod*. 2013 May 31; 88(5):131–131.