CellPress
REVIEWS

# *De Novo* Mutations Reflect Development and Aging of the Human Germline

J.M. Goldmann,[1] J.A. Veltman,[2,3] and C. Gilissen[1,4,*]

**Human germline *de novo* mutations (DNMs) are both a driver of evolution and an important cause of genetic diseases. In the past few years, whole-genome sequencing (WGS) of parent–offspring trios has facilitated the large-scale detection and study of human DNMs, which has led to exciting discoveries. The overarching theme of all of these studies is that the DNMs of an individual are a complex mixture of mutations that arise through different biological processes acting at different times during human development and life.**

## Highlights

Both the age of the father and the age of the mother are positively correlated with the number of *de novo* mutations (DNMs) in offspring, with the effect size of paternal age being larger.

DNMs associated with paternal and maternal aging each have unique mutational signatures based on their nucleotide substitution spectrum and genomic locations.

DNM clusters have characteristics distinct from nonclustered DNMs, suggesting different underlying mutational causes.

Post-zygotic mutations arising during early embryonic development are a frequent phenomenon and differ from nonmosaic germline DNMs in their mutational spectrum.

A subset of presumed DNMs can be traced back as low-level mosaic mutations in the somatic tissue of a parent. These mutations have a distinct mutation spectrum and can recur in future offspring.

## *De novo* mutations

Human ***de novo* mutations (DNMs,** see Glossary) are **germline** mutations that newly occurred within one generation. While the vast majority of the genome has been inherited from earlier generations, DNMs provide new genetic variation. The consequences of the new genetic mutation can vary widely. While neutral or advantageous mutations might become established in the genome of our species and thereby contribute to human evolution, changes to crucial genetic sequences can also result in misfunctioning of biological systems, resulting in severe disease. One of the earliest known examples of this was Down syndrome, which is caused by a *de novo* trisomy of chromosome 21 [1–3]. In recent years DNMs have been found to be a prominent cause of neurodevelopmental diseases, including intellectual disability, autism, and schizophrenia [4,5]. The unbiased study of *de novo* point mutations in humans was for many years hampered by the lack of techniques to scan the entire genome in a cost-effective way. The introduction of **next-generation sequencing (NGS)** technologies has spurred investigations of DNMs in humans [6]. DNMs can refer to a variety of different mutation types, such as single-nucleotide substitutions, insertions, deletions, and copy-number variants (CNVs). In this review we focus on single-nucleotide mutations and review the progress made in this field since the introduction of WGS, exploring their biology and possible underlying mechanisms (Figure 1, Key Figure), but not the potential pathological consequences.

## The Paternal Age Effect

One of the most prominent findings concerning germline DNMs is the fact that their number increases steadily with the age of the father at conception. The concept of mutations being associated with age was first described by the human geneticist Wilhelm Weinberg in 1912, who observed that sporadic achondroplasia is more common in last-born children than in first-born children [8]. For several sporadic genetic diseases, the association with paternal age is stronger than the association with maternal age [9]. It has been hypothesized that the underlying cause of these associations could be unprecise genome copying during the large number of cell divisions required for continuous sperm cell production in males [8–10]. The first direct observation of DNM rates being higher for offspring of fathers of advanced age was made by WGS of 78 Icelandic parent–offspring trios [11]. This study described that the number of DNMs in the offspring is dependent on the age of the father, with a slope of two additional DNMs per year of life of the father at the time of conception. This association of increased paternal age with increased numbers of DNMs in offspring is generally called the paternal age effect.

The main hypothesis of the underlying cause for the paternal age effect is that mutations arise from incidental copying errors during genome replications [8–11]. Genome replications are frequent in the male germline to ensure continuous generation of sperm cells. The stem cells for sperm production are spermatogonia, located in the seminiferous epithelium. Spermatogonia divide to renew themselves and give rise to spermatocytes, which will eventually differentiate into sperm cells [12]. During aging, the mean number of divisions per spermatogonium continues to increase, with each replication potentially introducing new mutations due to copying errors. Due to this,

[1]Department of Human Genetics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Geert Grooteplein 10, 6525 GA Nijmegen, The Netherlands

[2]Institute of Genetic Medicine, International Centre for Life, Newcastle University, Newcastle upon Tyne, UK

[3]Department of Human Genetics, Donders Centre for Neuroscience, Radboud University Medical Center, Geert Grooteplein 10, 6525 GA Nijmegen, The Netherlands

[4]https://www.radboudumc.nl/en/people/christian-gilissen

*Correspondence: Christian.gilissen@radboudumc.nl

## Key Figure

### Overview of Distinct Classes of Germline *De Novo* Mutations (DNMs)

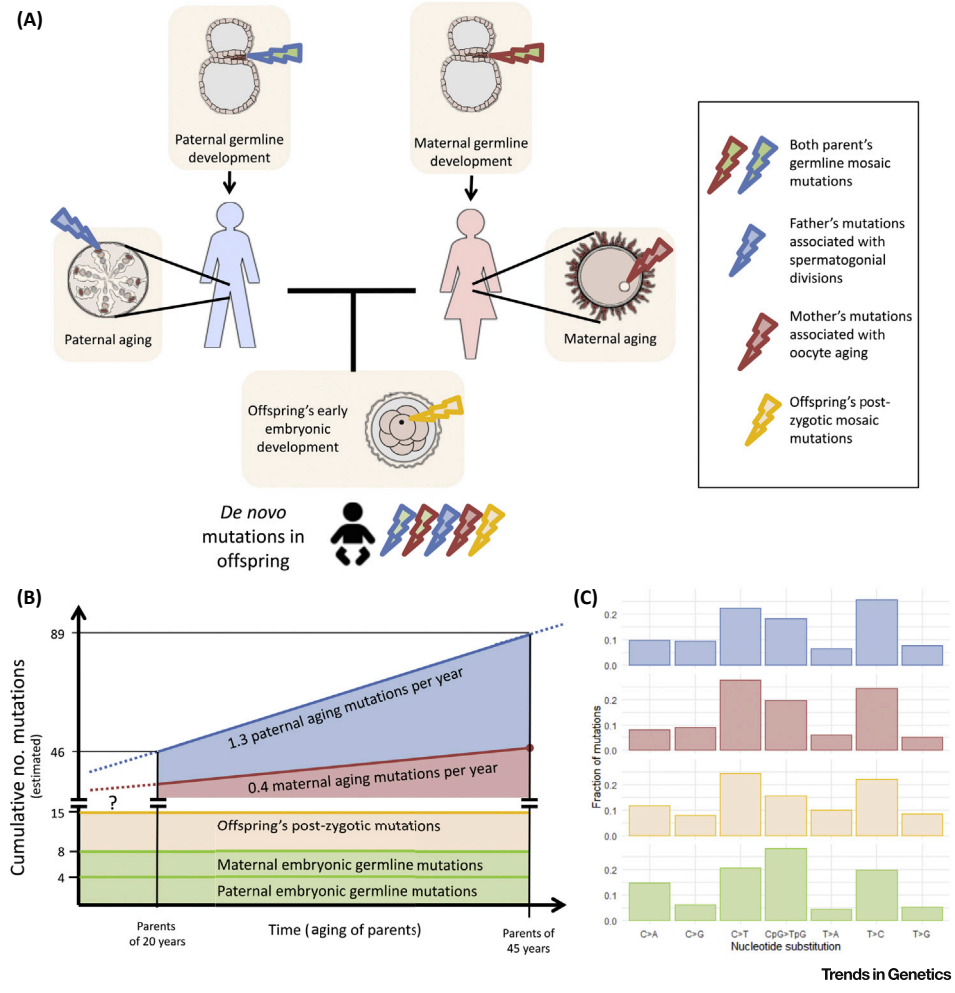


**Figure 1.** (A) Schematic depiction of the biological niches where DNMs arise. The earliest DNMs of an individual arise during the embryonic development of both parents, starting with the specification of primordial germline cells during gastrulation (top schemes; green bolts). During aging of the father, DNMs arise in spermatogonial cells, which replicate continuously to produce sperm (mid-left scheme; blue bolts). Aging of the oocytes in the mother is not accompanied by genome replications, but mutations still arise (mid-right; red bolt). Finally, post-zygotic mutations during early embryonic development of the offspring can affect a large fraction of the offspring's cells (bottom scheme; orange bolt). (B) While aging-associated mutations are more predominant in offspring from parents of advanced age, the number of embryonic DNMs is independent of parental age. Parent–offspring trio sequencing experiments do not allow the analysis of mutations arising during childhood; therefore, uncertainty remains about DNM accrual before the fertile period. (C) Each class of DNM has a distinct mutational spectrum. All numeric values are based on observations by Sasani *et al.* [7].

## Glossary

**Crossover interference**: crossover recombination events formed during meiosis are not randomly distributed along chromosome arms. Usually, only one event occurs per chromosome arm and in the case of two events there is a large distance between them. This phenomenon, that crossovers do not occur in close proximity, is termed the 'interference' of crossovers.

***De novo* mutations (DNMs)**: while literally meaning 'new' mutations, most commonly this refers to germline mutations that are present in most cells of an individual but not present in the majority of its parents' cells. Thereby, *de novo* mutations are the mutations that were introduced in the germline during the course of one generation.

**Double-strand break (DSB)**: a DNA lesion that leads to a complete breach of both phosphate backbones of the DNA molecule.

**Germline**: the developmental lineage of cells from which haploid gametes will be formed. Germline cells will transmit the genome to the next generation. All cells that are not part of the germline are somatic cells.

**Mosaic mutations**: mutations that affect only a subset of the cells of an individual.

**Mutational signature**: a typical 'footprint' mutational spectrum that occurs in many samples and may be associated with a specific mutational cause.

**Mutational spectrum**: a summary statistic of a group of mutations describing the relative frequencies of nucleotide conversions that resulted from the mutation. Six nucleotide conversions can be distinguished in sequencing data (i.e., C/G>A/T, C/G>G/C, C/G>T/A, T/A>A/T, T/A>C/G, T/A>T/A>G/C). In addition, the nucleotide conversions are often stratified by the nucleotides surrounding the affected base, resulting in a spectrum of 96 mutation types.

**Mutation cluster**: multiple mutations that affect the same allele of the same chromosome and occur in close spatial proximity (with intermutational distances smaller than 10–60 kb).

spermatogonia are believed to accumulate mutations during aging, thereby explaining the paternal age effect.

Next to this genome-wide paternal age effect, an additional paternal age effect has been observed for highly specific mutations. Due to particular activating mutations in specific genes, affected spermatogonia lineages can generate more sperm cells than other lineages. As a consequence, the effective germline mutation rate in these genes is up to four orders of magnitude higher than the genome-wide rate. This effect has been termed selfish spermatogonial selection (Box 1). The association of these mutations with paternal age is stronger than the linear genome-wide paternal age effect [13].

Recent studies have confirmed the observation of the general paternal age effect [23–28]. The generation of large-scale datasets allowed deeper characterization of the DNMs, including the analysis of mutational spectra and **mutational signatures** (Box 2). The **mutational spectrum** of human DNMs has been found to mainly comprise two mutational signatures that are known from somatic mutations in cancer [24,29]. These two signatures occur in a wide range of somatic tissues and accumulate mutations in a clock-like manner (i.e., at a constant rate). Therefore, in somatic tissues these signatures are likely to represent the impact of aging on DNA. In line with the replication-error hypothesis, one of these signatures is associated with genome replication [29]. Surprisingly, however, the mutational spectrum of DNMs is not constant and changes with the age of the father at conception. With advancing paternal age, more T>G substitutions and fewer non-CpG C>T substitutions are found in the offspring [27,28]. The existence of such differences suggests that, besides the paternal aging-associated mutations, there is another type of mutation that is independent of paternal age. With increasing age of the father, the balance of these types of mutation shifts and the mutational spectrum becomes biased towards more paternal aging-associated mutations.

## The Paternal Age Effect and Estimations of the Number of Germline Genome Replications

According to the genome-replication hypothesis for the paternal age effect, the number of replications should be directly proportional to the number of DNMs in offspring. The established model for estimating the number of genome replications that occurred during development from zygote to sperm cell includes the following factors [10,34,35]. First, during very early embryogenesis, about ten cell divisions occur until germline cells are specified and the germline is established. Second, during sex organ development, the seminiferous ducts become colonized with germ cells, which requires about 24 germline cell divisions after germ cell specification. During the further fetal period and childhood, the seminiferous epithelium rests and no cell divisions occur. Starting in puberty, spermatogonia will divide continuously to generate sperm cell precursors. It has been estimated that the average spermatogonium divides 23 times per year during this period [36]. Finally, an age-independent number of four genome

**Next-generation sequencing (NGS)**: a group of technologies that allow the reading of the DNA sequence at unprecedented speed and scale.
**Paternal/maternal DNMs**: DNMs that affect the allele of the chromosome inherited from the father/mother, respectively.
**Post-zygotic mutations (PZMs)**: mosaic mutations that arose in early embryonic development.
**Primordial germline cell specification (PCGS)**: the induction of germline cells during early embryonic development.
**Read phasing**: deduction of the allele affected by a mutation from the pattern of sequencing reads. This deduction allows assessment of the parental origin of the affected allele. In addition, it can be used to identify mosaic mutations.
**Somatic cells**: all cells in a human body that are not part of the germline.

### Box 1. Selfish Spermatogonial Selection

The mutations that accumulate in spermatogonia during paternal aging can also have an effect on the fitness of the spermatogonia themselves. While mutations in essential genes might lead to the eradication of the spermatogonial lineage, other mutations can result in selective advantages and lead to clonal expansion; that is, the outgrowth of the specific cell lineage, a process often linked to precancerous lesions. Clonal expansions have been observed in human tissues like bone marrow [14], skin [15], and esophageal epithelium [16], as well as many others [17]. In the human testis, clonal expansion has been observed to be driven by oncogenic mutations of growth factor receptors and components of the RAS–MAPK pathway [13,18]. As a consequence of the expansion of mutated spermatogonia that occurs in the testes of healthy men during aging [19–21], the number of sperm carrying these mutations grows continuously at rates much higher than the general, genome-wide paternal age effect. If these mutations are passed on to offspring, they will cause developmental disorders. One prominent example is achondroplasia, a condition characterized by short stature that is caused by mutation of the fibroblast growth factor receptor 3 (FGFR3) gene. Achondroplasia occurs more frequently than expected based on average mutation rates and its occurrence is strongly associated with paternal age [10,22]. It was shown that these pathogenic mutations are often present as high-level mosaicisms in the testes of the fathers of achondroplasia patients [20,22].

**Box 2. Mutational Signatures**

From the analysis of cancer genomes it is known that exposure to mutagens or failure of DNA repair pathways can result in mutations with specific patterns. An example is the light-associated skin cancers, in which C>T and CC>TT nucleotide substitutions are very common [30]. The advent of NGS technologies allowed the systematic study of mutational patterns in large collections of well-annotated cancer genomes [31–33]. In these analyses, the mutational signatures were analyzed by differentiating the six nucleotide substitution types and stratifying by the nucleotides surrounding the affected base pair, resulting in spectra of 96 mutation types. By mathematically deconvoluting archetypical patterns in the spectra of cancer mutations, these efforts identified 49 single-nucleotide-substitution mutational signatures. These signatures are well annotated and a subset could be linked to underlying mutational mechanisms. The mutational signatures are available online (https://cancer.sanger.ac.uk/cosmic/signatures). Analysis of a mutational spectrum and comparison with the known signatures can hint at the underlying mutational mechanisms.

replications is needed to proceed from spermatogonium to sperm cell. To give an example, the number of genome replications of an average sperm cell of a 30-year-old male can be estimated as ten very early embryonic divisions, plus 24 early embryonic divisions, plus 17 years of sperm production each with 23 divisions (assuming a sperm production onset at the age of 13 years [37]), and finally four replications for spermatogenesis. This is in total 429 replications. Fathers at the age of 20 and 60 years would generate sperm after 199 and 1119 replications, respectively.

However, these estimates for genome replications are not proportional with the observations on DNM accrual (Figure 2). While the number of replications increases by a factor of more than five between the parental ages of 20 and 60 years, the number of DNMs increases by less than a factor of three, as the estimated numbers would be 40 and 91, respectively [28]. Supporting the observations on DNMs, the epidemiological risk of giving birth to a child affected by a developmental disorder (of which a large fraction is caused by DNMs [38]) increases with paternal age only by a factor of less than two (from 0.24% for fathers and mothers both below 22 years up to 0.47% for fathers and mothers both above 42 years) [39]. This mismatch of the observations with the genome replication data has been reported several times and remains unresolved [40–44].

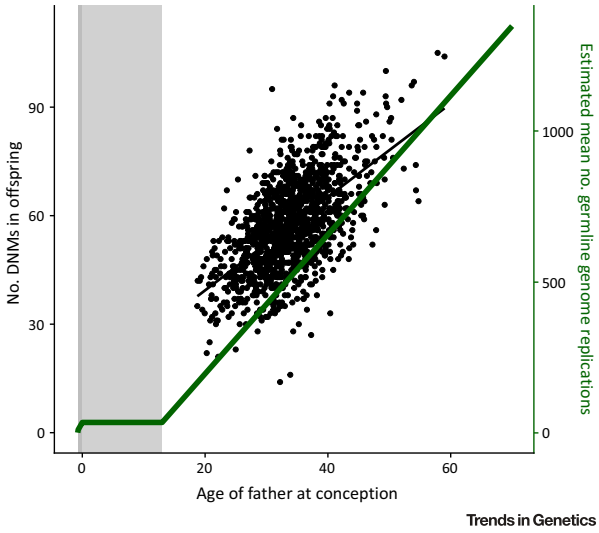


**Figure 2. Estimated Genome Replications in the Male Germline versus Number of *De Novo* Mutations (DNMs) in Offspring.**

During the reproductive period, the number of genome replications is estimated to grow by a higher factor than the number of observed DNMs (see text). Black dots indicate observations of DNM numbers in parent–offspring trios; the black line indicates a linear fit to the data [28]. Green line and green axis indicate the estimated number of genome replications in the male germline. Dark-gray area, prenatal period; light-gray area, childhood.

A crucial factor for the genome-replication model is the estimate of the annual post-pubertal spermatogonium division rate. This estimate of 23 divisions per year is based on observations in the 1960s by Heller and Clermont [36], who radiolabeled the seminiferous epithelium of the testicular ducts and showed that one division cycle lasts 16 days, which would extrapolate to 23 divisions per year. This extrapolation implicitly assumes that there are no interruptions or resting phases between the divisions. However, more recent studies have observed the nature of spermatogonial self-renewal to not be a unidirectional process but rather be driven by stochastic cell transitions, including cell stages without replicative activity [12,45,46]. Such a nonhierarchical model of spermatogonial self-renewal with stochastic, continuous oscillation between cell stages is supported by recent single-cell studies in two ways. First, analyses of single spermatogonial transcriptomes have shown that there is a continuous distribution of cells that do not express replication markers on the one hand and cells that are expressing markers of both replication and sperm differentiation on the other hand [47–51]. Second, spermatogonia have been found to be able to adapt their transcriptomes into both directions, either towards the replication-inactive state or into the amplification and differentiation state [47]. A spermatogonium division model that incorporates such oscillation with dormant cell stages would imply a slower rate of spermatogonium divisions and would therefore provide a better fit with the observations on DNM accrual. Such a model has been proposed and its parameters have been estimated by a fit to DNM accumulation data [43]. A mean spermatogonium division time parameter of 300 days provided the best concordance with the observations. According to this model, the yearly increase of spermatogonium divisions would be not 23 divisions per year, but 0.8. Subsequently, the number of germline divisions for sperm of 20-year-olds and 60-year-olds would be 44 and 76, respectively. This constitutes an increase by a factor of 1.7, which would be similar to the increase in DNM numbers. In addition, a lower rate of spermatogonium divisions would also imply that the human spermatogenesis per-replication mutation rate would be more similar to the same rate in the mouse as well as to the per-replication mutation rates for other phases of human germline biology [52,53]. Taken together, recent observations of DNM accumulation and single-cell transcriptome analyses of the spermatogenic epithelium suggest that estimated spermatogonium division rates may have been based on wrong assumptions.

## The Maternal Age Effect

Until the advent of WGS, DNMs were mostly studied as one contiguous group. However, in the past few years several studies have shown that **maternal DNMs**, although less frequent, contribute a distinct and characteristic source of germline mutations. By applying sequence-based **read phasing**, the parent of origin of the mutated allele can be assessed (Box 3). Studies applying this approach have found that the fraction of DNMs on the maternal allele is about a quarter to a third [23,24]. It has been suggested that this fraction is independent of the age of the parents [44], as maternal DNMs also accumulate during aging, although the rate of accumulation is smaller than for the paternal allele. The maternal age effect was initially detected by correlation [25] and later confirmed using maternal DNMs [26,27]. The mechanism underlying the maternal age effect cannot involve genome replication, as the production of oocytes ceases prenatally and no more genome replications occur. Therefore, the mechanisms that underlie maternal aging-associated DNMs must differ fundamentally from those underlying the paternal aging-associated DNMs. In line with this biological difference, the mutational spectra of maternal and **paternal DNMs** also differ. Maternal DNMs are enriched for C>G mutations and paternal DNMs are enriched for C>T mutations at CpG sites [27,28].

While the precise mechanisms that underlie maternal mutations remain to be understood, some clues were recently identified. First, there was the observation of genome regions highly enriched for maternal DNMs (up to eight times compared with paternal DNMs), including stretches of several million bases on chromosome 8, 9, and 16 [26–28]. The relative frequency of C>G mutations in these regions is significantly higher than among paternal DNMs and even higher than in the remaining maternal DNMs. These genomic regions were previously noticed for their peculiar patterns of recombination in the female germline. Meiotic gene conversions are highly enriched in these regions [54] and meiotic crossovers are less likely to comply with **crossover interference** [55]. This indicates that meiotic recombination could play a role in the formation of maternal DNMs. The second clue

**Box 3. Determining the Origin of a Mutated Allele**

When applying parent–offspring trio sequencing to identify DNMs, the variants are all identified individually such that the constitution of the alleles is unknown. For a subset of regions it is, however, possible to assemble the two haplotype alleles based on the sequencing data. This approach, termed (read) phasing, relies on the fact that two variants occurring in the same sequencing read must come from the same DNA molecule and thus from the same allele. This can be applied to DNMs and nearby heterozygous inherited variants, such that the mutated allele can be assigned to one of the parents. If the nearby variant is present in only one parent, the parent of origin of the mutated allele is determined in this way. With short-read sequencing technologies that generate reads of up to 150 bp, inherited variants must lie relatively close to the DNM to be present in the same sequencing read. Because the variant must also be uniquely attributable to a single parent to be informative, this method successfully determines the parent of origin for only 15–30% of DNMs. Figure I illustrates this approach.

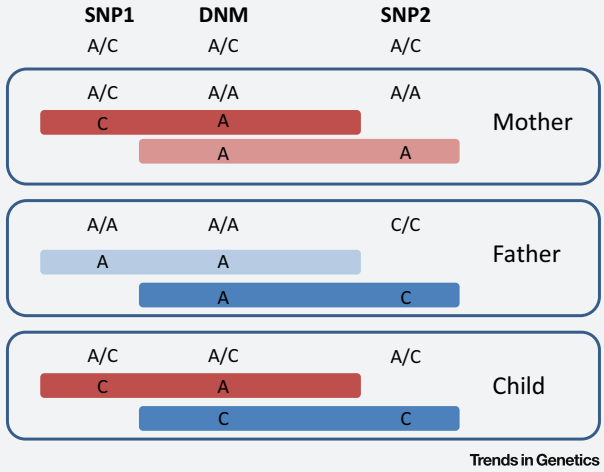


**Figure I. Sequencing Reads Spanning Two SNP Locations and a *De Novo* Mutation (DNM) Location on the Paternal Allele.**

The sequencing reads from mother, father, and child are shown. SNP1 is an informative SNP for the maternal allele; SNP2 is informative for the paternal allele. In the child, the sequencing read that contains the DNM also contains the SNP indicative of the paternal allele.

concerns the under-representation of CpG mutations on the maternal allele compared with the paternal allele. CpG sites in cells of the male germline are methylated at a higher rate than in the female germline, making the sites more prone to spontaneous deamination and C>T transitions [44,56]. While further research on this will be necessary to verify these hypotheses, these findings demonstrate how the biology of a cell lineage shapes the mutational spectrum (see Outstanding Questions).

## DNM Clusters

In 2012 Michaelson *et al.* introduced the concept of **mutation clusters**; that is, mutations on the same DNA molecule within short reciprocal distances, usually below several kilobases [57]. Michaelson *et al.* suggested that these represent a distinct subset of 2–3% of DNMs arising due to a single mutational event, as the incidence of clusters is too high to be explained by the incidental co-occurrence of multiple independent sequential DNMs [23,26,57–61]. The initial suggestion that these mutations represent a distinct subclass has now been substantiated by the finding that clustered DNMs display a distinct nucleotide substitution spectrum, with a significant enrichment of C>G mutations at non-CpG sites [23,26,28,58].

Clustered DNMs also differ from other DNMs in their parent-of-origin characteristics and parental age effects. In sharp contrast to other DNMs, clustered DNMs are equally abundant on paternal and maternal alleles, and their number is more strongly associated with maternal than with paternal age [27,28]. The mutational spectra differ significantly between the parental alleles,

with the maternal allele being more biased to C>G mutations than the paternal, and the distances between the DNMs of a maternal cluster are larger than for paternal clusters. Even more strongly than unclustered maternal DNMs, maternal clustered DNMs are biased to genomic regions prone to maternal meiotic gene conversions (predominantly, again, on chromosomes 8, 9, and 16). In addition, these clusters have been observed close to maternal meiotic gene conversions and close to maternal *de novo* CNVs [27,28]. Together these observations suggest that at least part of the maternal clustered DNMs may be introduced by **double-strand break (DSB)** repair mechanisms that were active in these regions (see Outstanding Questions).

## Embryonic DNMs: Footprints of Germline Development

New mutations not only arise during spermatogenesis and oogenesis, but may occur at any stage of embryonic development. Mutations arising during the embryonal period will be present in only a subset of cells (i.e., mosaics) and depending on the precise timing during development may be present in some tissues but absent in others. A crucial timepoint for the occurrence of embryonic mutations is **primordial germline cell specification (PCGS)** where the germline and somatic tissues are separated. This event occurs about ten cell divisions after fertilization, in the second or third week of embryonic development. Mutations that occur before PCGS may be present as mosaics in all the tissues of an individual. We will refer to these mutations as **post-zygotic mutations (PZMs)**. After this event, the germline is genetically isolated from the soma. Mutations that occur during germline development (after PCGS) will be present only as mosaics in the germline of that individual, and we will refer to these mutations as germline mosaics. Alternatively, mutations occurring during soma development are referred to as somatic mosaicisms.

## Post-zygotic Mutations

PZMs arise during very early embryogenesis before the specification of the germline. These very first divisions of embryogenesis are considered exceptionally mutagenic [62]. The paternal DNA of the zygote has been extremely compacted to fit within the head of the sperm cell, such that it is inaccessible to repair enzymes. Any DNA damage that occurred during the sperm DNA packaging or during the sperm cell's journey still needs repair [63]. Under these circumstances, the zygotic genome is very prone to chromosomal instability, resulting in aneuploidy and copy number variation [64]. Similarly, it is estimated that the per-division single-nucleotide mutation rate before PCGS is roughly twice as high as the mutation rate after PCGS [53].

Since PZMs occur before germline specification, they can affect both somatic and germline cells, and they are sometimes referred to as mixed or combined mosaicism. Therefore, the mutations can be identified in the somatic tissues of that individual (e.g., blood), but may also be transmitted to offspring. In such a case, the offspring will harbor the mutation in all of its cells whereas the parent carries the mutation in only part of its cells. Thus, if an offspring's DNM is present as a low-level mosaic in the somatic tissues of a parent, it must have occurred before PCGS of the parental embryo [24,65–67]. Because of this approach to identify PZMs using parent–offspring trios, they are often referred to as parental mosaicism. Sasani *et al.* applied this approach to blood-derived genome sequencing data from a cohort of 35 parent pairs and their offspring, identifying 475 germline–somatic combined **mosaic mutations** (mean of seven per individual) [7]. Consistent with an early embryonic mutation timing, no significant difference was found between the numbers of PZMs on the paternal allele and the maternal allele, nor did they observe a parental age effect. Contrary to DNMs, there were also no significant differences in the mutational spectra of the paternal and maternal alleles. The rate of seven early embryonic mutations per individual is likely to be an underestimate, as: first, reliable detection of mosaicism is limited by the sensitivity of detection of significant deviations in allele ratios (Box 4); and second, the absence of mosaicism in a sampled tissue does not exclude the presence of mosaicism in other tissues. Earlier studies applying the same approach yielded lower estimates of early embryonic mutations [24,66]. Future large-scale studies might provide more insights into the mutation spectra and biases in the genetic locations of these mutations.

**Box 4. Approaches to Identify Mosaic Mutations**

Techniques for detecting subclonal mutations are developing rapidly [68,69]. When sequencing large numbers of cells with a mosaic mutation, the fraction of sequencing reads with the mutation will deviate from the 50% that is expected for heterozygote variants affecting all cells. Statistical profiling of the distribution of allele fractions from a large number of variants and detection of outliers is an approach to detect PZMs from early embryogenesis [66,70]. Nevertheless, this approach is limited by the statistical power to detect significant outliers and to differentiate them from incidental sequencing errors. Statistical power can be improved by sequencing samples at a very high coverage or by applying molecular barcoding [71]. Besides this approach, monozygotic twins, while sharing the vast majority of their genomes, differ by few variants, which are likely to have arisen from embryonic PZMs [72]. Other approaches to identify PZMs that have recently been demonstrated rely on the use of read phasing (Box 3) [73,74], sampling of multiple tissues from the same individual [75,76], or observations on linkage of the mutation in grand-offspring [62]. Despite all techniques being capable of detecting clones present at high abundance, as theoretically expected for mutations occurring during the very first cell divisions of an embryo, uncertainty about timing of the mutations will remain. Due to loss of cells during development, clonal dynamics, and unequal distribution of alleles among organs, it is difficult to draw conclusions about the exact timing of mosaic mutations.

## Germline Mosaics

The germline is formally established during PCGS, occurring during the third week of embryonic development about ten cell divisions after fertilization, when the embryo is implanted into the uterus wall [12,77]. At that stage, the embryo is a flat, disk-like structure at the interface of two cavities, the amniotic cavity and the yolk sac. The posterior end of this disk is thought to be the location where several cells develop into primordial germline cells [12,77]. After the specification, germline cells migrate to the gonadal ridge and further differentiate, eventually turning into spermatogonia or oogonia.

Embryonic mutations that arise after the germline specification are not expected to affect the **somatic cells** of the same individual. Nevertheless, they can be passed on to any offspring and potentially have implications for their health. Mutations occurring earlier in germline development affect a larger fraction of gametes and therefore the transmission risk is higher. DNMs that recur in several offspring from a parent but that are not detectable in the somatically derived DNA of that person are therefore likely to have arisen during early germline development. Two recent large-scale studies identified such mutations by WGS of large numbers of families [7,78]. Sasani *et al.* investigated large families with more than seven offspring and showed that 3% of DNMs are shared between siblings. They conclude that these mutations stem from parental germline mosaics. These sibling-shared DNMs have a mutation spectrum distinct from nonshared DNMs, with more CpG>CpT mutations and fewer T>C mutations [7,78]. The sibling-shared DNMs were equally divided between paternal and maternal alleles and their numbers were constant across different parental ages [78]. In addition, the genomic regions that are especially vulnerable during maternal aging did not show an excess of sibling-shared mutations. Nevertheless, in parallel with the sexual dimorphism that is established during this phase of germline development, the spectra of sibling-shared mutations on maternal and paternal allele differ. Sibling-shared DNMs on the maternal allele are more likely to be T>C substitutions and less likely to be non-CpG C>T mutations than the same mutations on the paternal allele [78]. All of this suggests that DNMs derived from parental germline mosaicism may define yet another distinct subclass of DNMs that arise through different mutational mechanisms.

## Somatic Mosaics

Whereas germline mosaics affect only the germline of an individual, somatic mosaics also occur after PCGS and affect only the somatic tissues. Therefore, these mutations are not transmittable to offspring. Somatic mutations occur constantly during life, with an estimated rate of one mutation per cell division [52]. Mutations arising after organogenesis are usually restricted to the organs or tissue from which they originate. It is good to note that in traditional trio studies somatic mosaic mutations cannot be distinguished from PZMs. In addition, when a somatic mutation is present in a high fraction of cells, it may be misclassified as a germline DNM [66].

## Concluding Remarks and Future Perspectives

The past few years has seen tremendous progress being made in the study of DNMs and their biology. The overarching theme of all of these studies is that each individual contains a complex mixture of DNMs that arise through different biological processes. These processes act differentially on the genome of paternal and maternal germline cells as well as during the development of an individual. These differences not only are important for our fundamental understanding of biology and human evolution, but also have implications for human health and disease.

The fact that aging is a large contributor to germline mutation has several implications for contemporary and historic populations. Children from parents of advanced age are more likely to suffer diseases caused by DNMs, such as intellectual disability, developmental disorders, autism, and epilepsy [4,39,79–81]. In line with this, studies of historic populations indicated that late-born children on average have lower genetic fitness, as they were less likely to survive infancy and to engage in marriage [82–84]. In contemporary populations, there is a significant trend for potential parents to delay parenthood [85], increasing the probability of genetic disease. This trend will increase the impact of *de novo* genetic diseases on our societies in the future.

For families that gave birth to offspring affected by *de novo* genetic disease, it is of high practical relevance to estimate the chance that the same DNM could recur in a new pregnancy and subsequently cause a next child to be affected as well. The chance of a mutation affecting several offspring critically depends on the type of mutation. Mutations occurring after meiosis are present in only one oocyte or four sperm and are therefore extremely unlikely to affect several offspring. By contrast, mutations that occurred in the early germline development of one of the parents are likely to affect many germ cells and are therefore likely to recur. In addition, the pathological spectrum of diseases caused by embryonic mutations may differ from nonembryonic DNMs [86,98]. Some diseases appear to be preferentially caused by mosaic mutations, like Cornelia de Lange syndrome [87,88], Dravet syndrome [89], and encephalitic epilepsy [90,91]. In other diseases the causative mutations are exclusively found as mosaics, like for instance Proteus syndrome [92]. Another example is dominant mutations in the gene *MECP2* at the X chromosome that give rise to RETT syndrome in females but causes embryonic lethality in males. However, when these mutations occur as PZMs, males may be affected with the same syndrome [93].

For a long time, it has been impossible to deduct the precise timing of a mutation, but with increasing knowledge of the different types of DNMs and their properties it is now possible to provide some probabilistic indications. Jonsson *et al.* implemented an online calculator that is based on observations in their own dataset to estimate recurrence probabilities [78]. The single most meaningful criterion for estimating the recurrence risk is whether the mutation is detectable in any somatic tissues of the parents. In cases of somatic involvement, the mutation predates PCGS in the parents and is likely to recur (as discussed above). In addition, the recurrence risk depends on the parental origin of the mutated allele. As there are more aging-associated DNMs in the male germline, the risk of maternal DNMs arising during embryogenesis and subsequently affecting a large fraction of germline cells is higher. Consequently, maternal DNMs are more likely to recur [78]. Future improvements to such recurrence risk assessment, specifically for paternal mutations, could be to perform semen analysis. This has been demonstrated by Breuss *et al.*, who traced pathogenic DNMs from affected offspring back in paternal semen and profiled the fraction of sperm affected by these mutations [94].

Among the technological developments that will further propel research on DNMs, we want to specifically highlight three that we believe have most potential. The first is the reduction in cost of genome sequencing. This makes it feasible to investigate DNMs on an even larger scale and will eventually lead to more classes of DNMs being detected and characterized, including mosaic mutations (Box 4). The second development is the rise of long-read sequencing that will allow for phasing of a larger fraction of DNMs and thereby better understanding of the differences between male and female germlines. In addition, it will allow the identification of mutations in the less-accessible regions

### Outstanding Questions

The mechanism underlying the maternal age effect cannot involve genome replication, as the production of oocytes ceases and no more genome replications occur. What are the causes that drive the maternal age effect of DNMs?

What causes particular regions in the human genome to be up to eight times enriched for maternal DNMs?

To date, the largest known driver of DNMs in the general, healthy population is aging. Given that in many populations the mean age at reproduction is increasing and subsequently the incidence of *de novo* genetic disease is likely to increase, it is desirable to limit the impact of aging on the genome. How can the mutagenicity of aging be limited?

Hazardous environmental and anthropogenic chemicals like toxins and drugs have long been known to cause mutations. Can such mutations be detected in the germline of the general, healthy population?

In germline biology and during the transmission of the genome, DNA is in several exceptional states. Besides the extremely prolonged prophase of maternal meiosis that lasts for several decades, the packaging of DNA in the head of sperm cells goes along with near-complete stripping of histones and extreme condensation of the DNA. Can we identify mutations caused by this process?

Depending on the genetic background, individuals might vary in the efficiency of their DNA repair pathways. While certain kinds of germline DNA damage are repaired efficiently in one individual, they might have a higher chance of causing mutations in another. Are there genetic components that cause some families to have offspring with DNM numbers higher or lower than the mean trend?

What other biological processes lead to DNMs in the human genome?

of the genome, and reliable identification of more classes of variants, including variations in repeat regions as well as structural variants across different size ranges [95]. Finally, increased accuracy of single-cell sequencing will allow the assessment of single gametes for DNMs such that germline mutations can be studied at unprecedented resolution [96,97].

## Acknowledgments

We thank Godfried van der Heijden for useful discussions.

## References

1. Davenport, C.B. (1932) Mendelism in man. *In Proceedings of the Sixth International Congress of Genetics*. http://www.esp.org/books/6th-congress/facsimile/contents/6th-cong-p135-davenport.pdf
2. Allen, G. (1974) Aetiology of Down's syndrome inferred by Waardenburg in 1932. *Nature* 250, 436–437
3. Lejeune, J. *et al.* (1959) [Study of somatic chromosomes from 9 mongoloid children]. *C. R. Hebd. Seances Acad. Sci.* 248, 1721–1722, (in French).
4. Veltman, J.A. and Brunner, H.G. (2012) *De novo* mutations in human genetic disease. *Nat. Rev. Genet.* 13, 565–575
5. Acuna-Hidalgo, R. *et al.* (2016) New insights into the generation and role of *de novo* mutations in health and disease. *Genome Biol* 17, 241
6. Koboldt, D.C. *et al.* (2013) The next-generation sequencing revolution and its impact on genomics. *Cell* 155, 27–38
7. Sasani, T.A. *et al.* (2019) Large, three-generation CEPH families reveal post-zygotic mosaicism and variability in germline mutation accumulation. *elife* 8, e46922
8. Weinberg, W. (1912) Zur Vererbung des Zwergwuchses. *Arch. Rassen Gesell. Biol.* 9, 710–718, (in German).
9. Penrose, L.S. (1955) Parental age and mutation. *Lancet* 266, 312–313
10. Crow, J.F. (2000) The origins, patterns and implications of human spontaneous mutation. *Nat. Rev. Genet.* 1, 40–47
11. Kong, A. *et al.* (2012) Rate of *de novo* mutations and the importance of father's age to disease risk. *Nature* 488, 471–475
12. Sharma, S. *et al.* (2019) Spermatogonial stem cells: updates from specification to clinical relevance. *Hum. Reprod. Update* 25, 275–297
13. Goriely, A. and Wilkie, A.O.M. (2012) Paternal age effect mutations and selfish spermatogonial selection: causes and consequences for human disease. *Am. J. Hum. Genet.* 90, 175–200
14. Genovese, G. *et al.* (2014) Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N. Engl. J. Med.* 371, 2477–2487
15. Martincorena, I. *et al.* (2015) High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* 348, 880–886
16. Martincorena, I. *et al.* (2018) Somatic mutant clones colonize the human esophagus with age. *Science* 362, 911–917
17. Yizhak, K. *et al.* (2019) RNA sequence analysis reveals macroscopic somatic clonal expansion across normal tissues. *Science* 364, eaaw0726
18. Maher, G.J. *et al.* (2018) Selfish mutations dysregulating RAS–MAPK signaling are pervasive in aged human testes. *Genome Res* 28, 1779–1790
19. Goriely, A. *et al.* (2003) Regulation of cell fate decision of undifferentiated spermatogonia by GDNF. *Science* 287, 1489–1493
20. Goriely, A. *et al.* (2009) Activating mutations in FGFR3 and HRAS reveal a shared genetic origin for congenital disorders and testicular tumors. *Nat. Genet.* 41, 1247–1252
21. Giannoulatou, E. *et al.* (2017) Whole-genome sequencing of spermatocytic tumors provides insights into the mutational processes operating in the male germline. *PLoS One* 12, e0178169
22. Dakouane Giudicelli, M. *et al.* (2008) Increased achondroplasia mutation frequency with advanced age and evidence for G1138A mosaicism in human testis biopsies. *Fertil. Steril.* 89, 1651–1656
23. Francioli, L.C. *et al.* (2015) Genome-wide patterns and properties of *de novo* mutations in humans. *Nat. Genet.* 47, 822–826
24. Rahbari, R. *et al.* (2015) Timing, rates and spectra of human germline mutation. *Nat. Genet.* 48, 126–133
25. Wong, W.S.W. *et al.* (2016) New observations on maternal age effect on germline *de novo* mutations. *Nat. Commun.* 7, 10486
26. Goldmann, J.M. *et al.* (2016) Parent-of-origin-specific signatures of *de novo* mutations. *Nat. Genet.* 48, 935–939
27. Jónsson, H. *et al.* (2017) Parental influence on human germline *de novo* mutations in 1,548 trios from Iceland. *Nature* 549, 519–522
28. Goldmann, J.M. *et al.* (2018) Germline *de novo* mutation clusters arise during oocyte aging in genomic regions with high double-strand-break incidence. *Nat. Genet.* 50, 487–492
29. Alexandrov, L.B. *et al.* (2015) Clock-like mutational processes in human somatic cells. *Nat. Genet.* 47, 1402–1407
30. Pfeifer, G.P. *et al.* (2005) Mutations induced by ultraviolet light. *Mutat. Res. Mol. Mech. Mutagen.* 571, 19–31
31. Nik-Zainal, S. *et al.* (2012) Mutational processes molding the genomes of 21 breast cancers. *Cell* 149, 979–993
32. Alexandrov, L.B. *et al.* (2013) Signatures of mutational processes in human cancer. *Nature* 500, 415–421
33. Alexandrov, L. *et al.* (2019) The repertoire of mutational signatures in human cancer. *bioRxiv*. Published online July 3, 2019. https://doi.org/10.1101/322859
34. Clermont, Y. (1963) The cycle of the seminiferous epithelium in man. *Am. J. Anat.* 112, 35–51
35. Drost, J.B. and Lee, W.R. (1995) Biological basis of germline mutation: comparisons of spontaneous germline mutation rates among *Drosophila*, mouse, and human. *Environ. Mol. Mutagen.* 25, 48–64
36. Heller, C.G. and Clermont, Y. (1963) Spermatogenesis in man: an estimate of its duration. *Science* 140, 184–186
37. Nielsen, C.T. *et al.* (1986) Onset of the release of spermatozoa (spermarche) in boys in relation to age, testicular growth, pubic hair, and height. *J. Clin. Endocrinol. Metab.* 62, 532–535

38. Gilissen, C. *et al.* (2014) Genome sequencing identifies major causes of severe intellectual disability. *Nature* 511, 344–347
39. Deciphering Developmental Disorders Study (2017) Prevalence and architecture of *de novo* mutations in developmental disorders. *Nature* 542, 433–438
40. Ségurel, L. *et al.* (2014) Determinants of mutation rate variation in the human germline. *Annu. Rev. Genomics Hum. Genet.* 15, 47–70
41. Forster, P. *et al.* (2015) Elevated germline mutation rate in teenage fathers. *Proc. Biol. Sci.* 282, 20142898
42. Gao, Z. *et al.* (2015) Interpreting the dependence of mutation rates on age and time. *PLoS Biol* 14, e1002355
43. Scally, A. (2016) Mutation rates and the evolution of germline structure. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 371, 20150137
44. Gao, Z. *et al.* (2019) Overlooked roles of DNA damage and maternal age in generating human germline mutations. *Proc. Natl Acad. Sci. U. S. A.* 116, 9491–9500
45. Hara, K. *et al.* (2014) Mouse spermatogenic stem cells continually interconvert between equipotent singly isolated and syncytial states. *Cell Stem Cell* 14, 658–672
46. Krieger, T. and Simons, B.D. (2015) Dynamic stem cell heterogeneity. *Development* 142, 1396–1406
47. Guo, J. *et al.* (2018) The adult human testis transcriptional cell atlas. *Cell Res* 28, 1141–1157
48. Green, C.D. *et al.* (2018) A comprehensive roadmap of murine spermatogenesis defined by single-cell RNA-seq. *Dev. Cell* 46, 651–667.e10
49. Hermann, B.P. *et al.* (2018) The mammalian spermatogenesis single-cell transcriptome, from spermatogonial stem cells to spermatids. *Cell Rep* 25, 1650–1667.e8
50. Wang, M. *et al.* (2018) Single-cell RNA sequencing analysis reveals sequential cell fate transition during human spermatogenesis. *Cell Stem Cell* 23, 599–614.e4
51. Guo, J. *et al.* (2017) Chromatin and single-cell RNA-seq profiling reveal dynamic signaling and metabolic transitions during human spermatogonial stem cell development. *Cell Stem Cell* 21, 533–546.e6
52. Milholland, B. *et al.* (2017) Differences between germline and somatic mutation rates in humans and mice. *Nat. Commun.* 8, 15183
53. Lindsay, S.J. *et al.* (2019) Striking differences in patterns of germline mutation between mice and humans. *Nat. Commun.* 10, 4053
54. Halldorsson, B.V. *et al.* (2016) The rate of meiotic gene conversion varies by sex and age. *Nat. Genet.* 48, 1377–1384
55. Campbell, C.L. *et al.* (2015) Escape from crossover interference increases with maternal age. *Nat. Commun.* 6, 6260
56. Kobayashi, H. *et al.* (2013) High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. *Genome Res* 23, 616–627
57. Michaelson, J.J. *et al.* (2012) Whole-genome sequencing in autism identifies hot spots for *de novo* germline mutation. *Cell* 151, 1431–1442
58. Besenbacher, S. *et al.* (2016) Multi-nucleotide *de novo* mutations in humans. *PLoS Genet* 12, e1006315
59. Schrider, D.R. *et al.* (2011) Pervasive multinucleotide mutational events in eukaryotes. *Curr. Biol.* 21, 1051–1054
60. Terekhanova, N.V. *et al.* (2013) Prevalence of multinucleotide replacements in evolution of primates and *Drosophila*. *Mol. Biol. Evol.* 30, 1315–1325
61. Yuen, R.K. *et al.* (2016) Genome-wide characteristics of *de novo* mutations in autism. *NPJ Genom. Med.* 1, 16027
62. Harland, C. *et al.* (2016) Frequency of mosaicism points towards mutation-prone early cleavage cell divisions. *bioRxiv*. Published online October 9, 2016. https://doi.org/10.1101/079863
63. Rathke, C. *et al.* (2014) Chromatin dynamics during spermiogenesis. *Biochim. Biophys. Acta* 1839, 155–168
64. Vanneste, E. *et al.* (2012) Aneuploidy and copy number variation in early human development. *Semin. Reprod. Med.* 30, 302–308
65. Campbell, I.M. *et al.* (2014) Parental somatic mosaicism is underrecognized and influences recurrence risk of genomic disorders. *Am. J. Hum. Genet.* 95, 173–182
66. Acuna-Hidalgo, R. *et al.* (2015) Post-zygotic point mutations are an underrecognized source of *de novo* genomic variation. *Am. J. Hum. Genet.* 97, 67–74
67. Campbell, I.M. *et al.* (2014) Parent of origin, mosaicism, and recurrence risk: probabilistic modeling explains the broken symmetry of transmission genetics. *Am. J. Hum. Genet.* 95, 345–359
68. Forsberg, L.A. *et al.* (2016) Mosaicism in health and disease – clones picking up speed. *Nat. Rev. Genet.* 18, 128–142
69. Dou, Y. *et al.* (2018) Detecting somatic mutations in normal cells. *Trends Genet* 34, 545–557
70. Huang, A.Y. *et al.* (2014) Postzygotic single-nucleotide mosaicisms in whole-genome sequences of clinically unremarkable individuals. *Cell Res* 24, 1311–1327
71. Acuna-Hidalgo, R. *et al.* (2017) Ultra-sensitive sequencing identifies high prevalence of clonal hematopoiesis-associated mutations throughout adult life. *Am. J. Hum. Genet.* 101, 50–64
72. Dal, G.M. *et al.* (2014) Early postzygotic mutations contribute to *de novo* variation in a healthy monozygotic twin pair. *J. Med. Genet.* 51, 455–459
73. Freed, D. and Pevsner, J. (2016) The contribution of mosaic variants to autism spectrum disorder. *PLoS Genet* 12, e1006245
74. Ju, Y.S. *et al.* (2017) Somatic mutations reveal asymmetric cellular dynamics in the early human embryo. *Nature* 543, 714–718
75. Bae, T. *et al.* (2018) Different mutational rates and mechanisms in human cells at pregastrulation and neurogenesis. *Science* 359, 550–555
76. Huang, Y. *et al.* (2018) Distinctive types of postzygotic single-nucleotide mosaicisms in healthy individuals revealed by genome-wide profiling of multiple organs. *bioRxiv*. Published online April 26, 2018. https://doi.org/10.1101/309005
77. Tang, W.W.C. *et al.* (2016) Specification and epigenetic programming of the human germ line. *Nat. Rev. Genet.* 17, 585–600
78. Jónsson, H. *et al.* (2018) Multiple transmissions of *de novo* mutations in families. *Nat. Genet.* 50, 1674–1680
79. Wiener-Megnazi, Z. *et al.* (2012) Advanced paternal age and reproductive outcome. *Asian J. Androl.* 14, 69–76
80. Taylor, J. *et al.* (2019) Paternal-age-related *de novo* mutations and risk for five disorders. *Nat. Commun.* 10, 3043
81. Krupp, D.R. *et al.* (2017) Exonic mosaic mutations contribute risk for autism spectrum disorder. *Am. J. Hum. Genet.* 17, 241
82. Arslan, R.C. *et al.* (2017) Older fathers' children have lower evolutionary fitness across four centuries and in four populations. *Proc. Biol. Sci.* 284, 20171562
83. Hayward, A.D. *et al.* (2015) Fitness consequences of advanced ancestral age over three generations in humans. *PLoS One* 10, e0128197

84. Nybo Andersen, A.-M. and Urhoj, S.K. (2017) Is advanced paternal age a health risk for the offspring? *Fertil. Steril.* 107, 312–318
85. Mills, M. *et al.* (2011) Why do people postpone parenthood? Reasons and social policy incentives. *Hum. Reprod. Update* 17, 848–860
86. Biesecker, L.G. and Spinner, N.B. (2013) A genomic view of mosaicism and human disease. *Nat. Rev. Genet.* 14, 307–320
87. Huisman, S.A. *et al.* (2013) High rate of mosaicism in individuals with Cornelia de Lange syndrome. *J. Med. Genet.* 50, 339–344
88. Ansari, M. *et al.* (2014) Genetic heterogeneity in Cornelia de Lange syndrome (CdLS) and CdLS-like phenotypes with observed and predicted levels of mosaicism. *J. Med. Genet.* 51, 659–668
89. Yang, X. *et al.* (2017) Genomic mosaicism in paternal sperm and multiple parental tissues in a Dravet syndrome cohort. *Sci. Rep.* 7, 15677
90. Stosser, M.B. *et al.* (2018) High frequency of mosaic pathogenic variants in genes causing epilepsy-related neurodevelopmental disorders. *Genet. Med.* 20, 403–410
91. Møller, R.S. *et al.* (2019) Parental mosaicism in epilepsies due to alleged *de novo* variants. *Epilepsia* 60, epi.15187.
92. Lindhurst, M.J. *et al.* (2011) A mosaic activating mutation in AKT1 associated with the Proteus syndrome. *N. Engl. J. Med.* 365, 611–619
93. Topçu, M. *et al.* (2002) Somatic mosaicism for a MECP2 mutation associated with classic Rett syndrome in a boy. *Eur. J. Hum. Genet.* 10, 77–81
94. Breuss, M. *et al.* (2017) Quantification of autism recurrence risk by direct assessment of paternal sperm mosaicism. *bioRxiv*. Published online October 24, 2017. https://doi.org/10.1101/208165
95. Mantere, T. *et al.* (2019) Long-read sequencing emerging in medical genetics. *Front. Genet.* 10, 426
96. Wang, J. *et al.* (2012) Genome-wide single-cell analysis of recombination activity and *de novo* mutation rates in human sperm. *Cell* 150, 402–412
97. Tran, Q.T. *et al.* (2019) Chromosomal scan of single sperm cells by combining fluorescence-activated cell sorting and next-generation sequencing. *J. Assist. Reprod. Genet.* 36, 91–97
98. Campbell, I.M. *et al.* (2015) Somatic mosaicism: implications for disease and transmission genetics. *Trends Genet.* 31, 382–392