

## Decoding germline de novo point mutations

Anne Goriely

Weatherall Institute of Molecular Medicine, University of Oxford

[Anne.Goriely@imm.ox.ac.uk](mailto:Anne.Goriely@imm.ox.ac.uk)

**Analysis of a large whole-genome sequencing dataset of 36,441 high-quality de novo mutations (DNMs) that arose in 816 family trios<sup>1</sup> provides an unprecedented view into the landscape of de novo mutations in the germline. This work both refines and challenges some of the views previously held on the nature and origin of DNMs.**

-----

Mutations are the source of all sequence variation in our genomes. Considering how important they are to species evolution and disease, it is surprising how little we know about the mechanisms controlling the occurrence and molecular spectrum of mutations, or even the rate at which they are introduced in our genome. A study of the largest set of germline de novo point mutations available so far, reported by Goldmann *et al*<sup>1</sup> in this issue, goes some way to bridging this gap in knowledge.

### We're all mutants

With the advent of next-generation sequencing, it has become possible to directly measure the intergenerational mutation rate across the human genome. In 2012, a landmark study<sup>2</sup> presented whole-genome sequencing data of 78 family trios (father-mother-child) which indicated that each newborn carries between 30-100 (average of 60) sequence alterations, also known as de novo mutations (DNMs), that were not present in their parents' DNA. Because DNMs arise in the parental germline (sperm or egg) before fertilization, their occurrence is intimately linked to the biology of germ cells. While all female eggs are produced during fetal development and do not divide after birth, sperm are continuously produced over a man's reproductive lifetime. Because the number of stem cell divisions increases as men age, it was postulated that DNMs must occur due to DNA replication errors that progressively accumulate in ageing males<sup>3</sup> (Figure 1). This was precisely what the 2012 study<sup>2</sup> described: the total number of DNMs in a child strongly correlated with the age of the father at conception, increasing by ~1-2 DNMs for every additional year of paternal age, while mothers contributed a small fraction (~10 DNMs), regardless of their age<sup>2</sup>. Since similar results have been obtained by several other studies<sup>4,5</sup>, the consensus with regard to germline DNMs - and therefore to sporadic disease risk - is that the main culprit is the error-prone process of DNA replication cycles associated with advanced paternal age.

### Sexing DNMs

The strength of the Goldmann *et al.* study is undoubtedly its size, reporting the analysis of whole-genome data from 816 trios, 10-times larger than the 2012 study<sup>2</sup>. In total, the authors identify 36,441 DNMs, an average of ~45 mutations per child (a rate lower than

previously described, because only high-confidence DNMs are considered). The real value afforded by the much larger scope of this work lies in using a computational approach to unambiguously assign a significant proportion of DNMs to their parental chromosomes of origin (in total 7,216 DNMs, many more than previous studies<sup>2,5</sup>). Using this sexed-dataset, the authors validate some previous findings, including the male-biased origin of DNMs - with ~80% of mutations originating during spermatogenesis - and the paternal age-effect of ~1 novel mutation per additional year of father's age. This stratified analysis provides the resolution needed to discover previously undetected parental-specific patterns. Earlier studies had lacked the power to detect a small but significant age-effect on maternally-derived DNMs, with mothers contributing ~1 mutation for every 4 additional years of maternal age. Other patterns also emerge from these sex-resolved DNMs: specific regions of the genome are enriched in maternally-derived mutations, while molecular signatures and the sequence-specific context in which maternal and paternal DNMs arise show significant differences (Figure 1).

### **Maintaining genome integrity**

Although these findings suggest distinct causative mechanisms in the origin of DNMs in the male and female germlines, the actual processes involved remain unclear. As multicellular organisms with a long generation time, our germlines act as gatekeepers to protect the integrity of heritable information across generations. To do so, they have evolved strategies to curtail mutation rates that likely involve mechanisms preventing lesions from arising as well as a complex repair machinery to correct errors upon damage. While replication errors occurring during stem cell divisions provide the most likely explanation for the origin of most paternally-derived DNMs, the description of a maternal age-effect with distinct molecular signatures implies that mutations also arise through non-replicative mechanisms. Gao et al<sup>6</sup> recently proposed that poorly repaired non-replicative DNA lesions would result in a steady, clock-like accumulation of mutations over time, a prediction that is vindicated by the present finding of a maternal-age effect. Although it may be unsettling to realize that we all carry up to 100 DNMs, at the whole-genome scale, this translates to a mutation risk as low as  $\sim 1 \times 10^{-8}$  at any given genomic location. This is several orders of magnitude lower than the rates observed for mutations in somatic tissues<sup>7</sup>, suggesting that overall the DNA replication and repair machineries work efficiently in germ cells. However, parental age-effects indicate that the accuracy of these mechanisms declines with age, albeit at a different pace for men and women; this ought to be a concern for our reproductively aging population. Another question raised by the present study is why do some regions of the genome show an enrichment of maternally-derived mutations? As these regions encompass large genes with tumour-suppressor function, perhaps some DNMs provide a survival advantage to mutant eggs in ageing women. Selfish mechanisms that promote proliferation of mutant stem cells and lead to progressive enrichment of sperm carrying pathogenic mutations are a well-known phenomenon in the male germline<sup>8</sup>.

These exciting findings should stimulate hypothesis-driven research through collaborative efforts. In the cancer field, large-scale genome analysis has found ~30 distinct mutational signatures across a spectrum of cancer types and provided a powerful approach to understanding tumor biology and the impact of specific mutagens<sup>9</sup>. It is hoped that similar large-scale strategies will soon be applied to germline DNMs. As many trio genomes have

been sequenced, efforts to generate a meta-dataset of sexed DNMs should be pursued. With increased power and further data stratification, more detailed mutational patterns will likely emerge. These may in turn correlate to factors such as parental life history or ethnic and genetic differences. This meta-analysis should also provide a unique opportunity to study basic germline biology and help not only understand the contribution of specific cellular processes to disease risk but also how these strategies have shaped the evolution of our genome.

1. Goldmann J.M *et al. Nat Genet* doi:10.1038/ng.3597 (2016)
2. Kong A. *et al. Nature* **488**, 471-475 (2012)
3. Crow J.F. *Nat Rev Gen* **1**, 40-47 (2000)
4. Francioli L.C. *et al. Nat Genet* **47**, 822-826 (2015)
5. Rahbari R. *et al. Nat Genet* **48**, 126-133 (2016)
6. Gao Z. *et al. PLoS Biol.* **14**, e1002355 (2016)
7. Vogelstein B *et al. Science* **339**, 1546-1558 (2013)
8. Goriely A. & Wilkie AOM *Am J Hum Genet* **90**, 175-200 (2012)
9. Helleday, T. *et al. Nat. Rev. Genet.* **15**, 585-598 (2014)

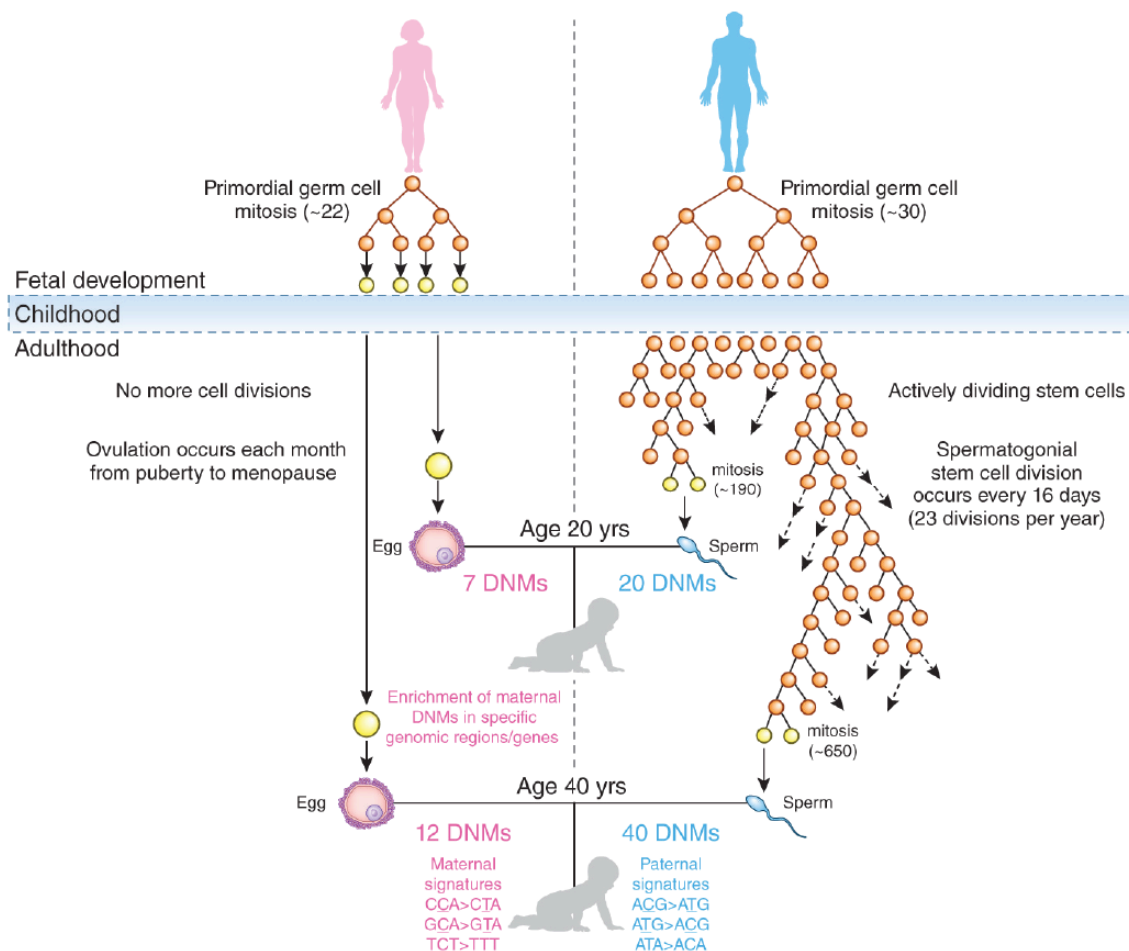


Figure 1: Gametogenesis differs in females (left/pink) and males (right/blue). The sperm produced by a 20-year-old has gone through ~190 cell divisions (mitosis). This number

increases to ~650 by the age of 40 years. As eggs do not replicate after birth, these differences are likely to explain the relative contribution of parental DNMs in the progeny. However, maternal and paternal DNM numbers increase with parental ages and show sex-specific mutational patterns. Orange = actively dividing stem cells; Yellow = differentiating gametes undergoing meiosis.

/631/208 Biological sciences/Genetics

/631/208/514 Biological sciences/Genetics/sequencing