1.5 Mutation: The ultimate source of genetic variation.

DNA is an exquisitely robust data storage system: a typical baby is born with just mutation one per ∼100 million base pairs (that's about 70 genome-wide). Nonetheless, mutations are central to our story, as they are the source of all genetic variation, for good and bad, enabling evolution and causing disease.

Figure 1.57: **A mutation** *is a change to the genome sequence: in this example a C→T mutation (also G→A on the other strand). A mutation at a single position, like this, is referred to as a* ***single nucleotide mutation*** *or* ***point mutation****.*

The existential challenge of DNA storage and replication. Each cell in your body carries a single precious copy of your genome. Errors in the genomes of your germline cells (the cells that produce gametes) can cause genetic diseases in your children; errors in somatic cells (cells of the body) can lead to cancer or other diseases of aging. Thus, safeguarding the integrity of the genome is a fundamental requirement for all cells.

And yet, every genome copy suffers a constant barrage of **DNA damage**: i.e., events that create molecular alterations, or **lesions**, in the DNA. But as we shall see, the vast majority of these lesions are repaired by DNA repair pathways. Only a tiny fraction of these result in **mutations** – i.e., events in which DNA repair or proofreading fails, resulting in permanent (uncorrectable) changes to the genome sequence of a cell.

It's estimated that a typical cell suffers 70,000 lesions per day[89]! The metabolic processes playing out continuously within each cell produce a variety of small nasty molecules such as reactive oxygen species that can cause DNA damage [a]. Meanwhile, hydrolysis reactions can cleave chemical bonds in DNA [90]. External mutagens including x-rays and gamma-rays, UV radiation (in exposed skin), and mutagenic chemicals such as nicotine, alcohol, and asbestos cause further damage.

The resulting lesions include many possible nucleotide modifications including addition and removal of methyl groups, deamination and depurination (in which a base is released), chemical modifications such as pyrimidine dimers in which adjacent thymines or cytosines form inappropriate covalent bonds parallel to the DNA helix.

Other damage events can cause breaks in the DNA molecule, including single strand breaks (one strand of the double helix breaks, while the other strand stays intact) or – much worse – double strand breaks, in which the helix breaks apart completely. Double strand breaks must be repaired rapidly to maintain cell viability. It's been estimated that a mammalian cell suffers 55,000 single strand breaks and 25 double strand breaks per day [91].

Moreover, the genome of 6 billion base pairs must be copied at every cell division. **DNA replication** provides further opportunity for errors, either when copying damaged sites that have not yet been corrected, or by errors introduced in the copying process itself. A typical cell in your body is descended through tens to hundreds of cell divisions – each involving

[a] *There is a large literature on mechanisms of DNA damage and repair, but most of this will be outside our scope in this book, except where it intersects with our main themes.*

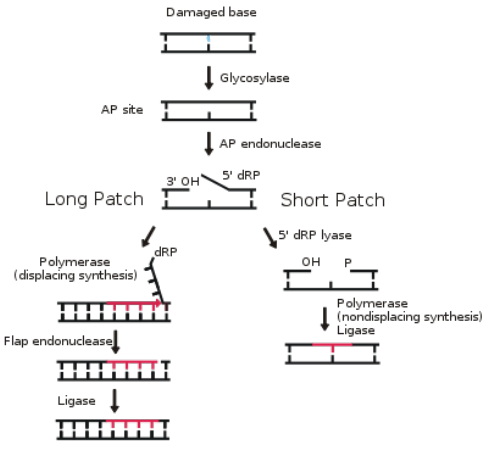

Figure 1.58: **Example of a DNA repair pathway:** *Base Excision Repair. Here, a damaged base (blue) is removed, and patched (red bases), using the other strand as template.* ***Cells suffer thousands of DNA lesions per day, but nearly all are repaired using pathways including this.*** *Credit: Amazinglarry [Link] Public Domain.*

genome copying – since you were a single fertilized egg.

Clearly, protecting the genome against decades of spontaneous chemical damage and mutagens, and accurate DNA replication through trillions of cell divisions during one's lifetime, is an existential challenge. Multicelled organisms couldn't survive if all this DNA damage resulted in permanent changes in genomes.

Consequently, cells have evolved an exquisitely complex molecular machinery of proteins responsible for detection and repair of spontaneous DNA damage, and for highly accurate DNA replication and proofreading. And when DNA damage is so severe that it cannot be repaired – as can happen with double strand breaks – there are alternate pathways for programmed cell death.

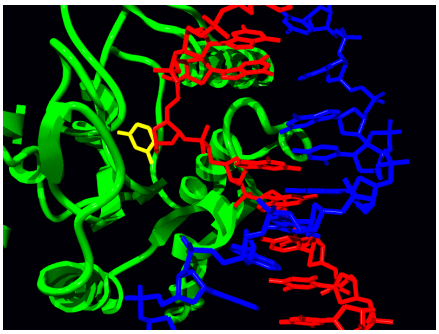

Figure 1.59: **Molecular structure showing repair of a damaged base**. *The DNA strands are red and blue. The repair enzyme uracil glycosylase is in green; it has flipped a nonstandard base (uracil, in yellow) out of the red strand prior to removal and correction by Base Excision Repair.* *Credit: TimVickers [Link] Public Domain.*

As we shall see, the repair and proofreading pathways are absolutely gob-smackingly effective, with germline mutation rates on the order of one per billion base pairs per year.

Mutations and evolution. *That's very impressive...* – you say – *But aren't some mutations good? Don't we also need mutations to enable adaptation?* Yes indeed, this is true. A tiny fraction of mutations are advantageous and, over thousands of years, these are the drivers of evolutionary change. Mutation enables what Darwin called "descent with modification": if there were no mutation, there would be no modification – and no evolution.

This suggests a paradox: On average, mutation is bad for individuals, but in the long-term mutations are necessary for species to adapt and survive. As we shall discuss later, natural selection acts mainly on short-term effects – in this case, the direct fitness cost of mutation – and lacks the foresight to consider possible future benefits to the species (Chapter 2.6). Thus, selection generally favors mutation rates to evolve as low as reasonably possible; fortuitously these rates are still high enough to enable adaptation [92] [93].

In the remainder of the chapter we discuss the rates and mechanisms of mutations.

Germline mutation rates. In animals, there is a strict separation between cells of the **germline** (which produce gametes–eggs or sperm), and the **soma** or **somatic cells** (which produce the body of the organism).

Mutations arise in both types of tissues, but they have very different implications: germline mutations can be passed on to future generations and, as such, they touch nearly every topic in this book; somatic mutations are not passed on, but can lead to cancer and potentially other diseases of aging.

Detecting germline mutations. We can detect ***de novo*** (new) germline mutations by sequencing genomes of families. The example below shows the sequencing of a **family trio**: both parents and a child. If the child has

an allele that is not present in either parent, this must have arisen by mutation, most likely in the germline of a parent [94]:

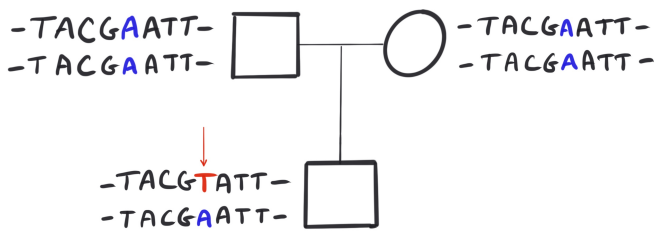

Figure 1.60: **Single nucleotide mutation in a child.** *De novo mutations can be detected by genome sequencing of family trios: here, both parents are homozygous for A, while the child is a T/A heterozygote.*

Starting around 2010, with access to 2nd-generation sequencing, there has been a series of studies characterizing mutation rates in a variety of populations [95]. Several of the most extensive studies have come from the genetics company DeCODE, based in Iceland [96]. The plot below, from DeCODE, shows a histogram of the number of single nucleotide mutations per child, across a large sample of families:

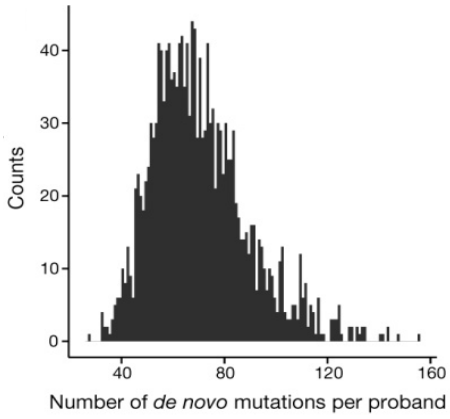

Figure 1.61: **Number of new mutations per child.** *The plot shows the distribution across children in many families.*

Credit: Figure 1d of Hákon Jónsson et al (2017). [Link] Used with permission.

As you can see, a typical child inherits about 70 single nucleotide mutations. Hmmm... does that seem like an awful lot of mutations to you? Well, bear in mind that only about 1% of the genome is protein coding, so a typical child will have about 0–1 mutations in protein coding regions, and perhaps a couple more in regulatory regions. Most of these will not have detectable effects. It's been estimated that about 1.5% of children are born with a loss-of-function mutation, such as premature stop, in a highly constrained gene. Such mutations are a major cause of childhood developmental disorders [97].

We are now ready to estimate the genome-wide mutation rate. The human genome is about 3.1 GB, but in the study above they could only get high-quality sequence data for about 2.68 GB (i.e., excluding repetitive regions). Remembering that each child gets two genome copies (one from each parent) we can estimate the average mutation rate as the average number of mutations divided by the sequenceable genome size [b]:

$$\frac{70 \text{ muts}}{2 \cdot (2.68 \times 10^9) \text{ bp}} = 1.3 \times 10^{-8} \text{ muts/bp} \qquad (1.1)$$

[b] *The human mutation rate is about* 1.3×10^{-8} *mutations per base pair per generation, or just slightly more than one mutation per 100 Mb. This is a fundamental parameter, and useful to remember.*

We'll see shortly that mutation rate increases linearly with the age of the parents; this estimate is for an average parental age of 30. Equivalently this corresponds to a mutation rate of about 4.0×10^{-10} per base pair per year of the parent's ages.

DNA replication is remarkably accurate. At this point I like to emphasize that DNA storage and replication is just remarkably, astonishingly, accurate. The DNA in your parent's germ cells was stored for 2–4 decades or more, and replicated hundreds of times, with an aggregate error rate of just one point mutation per 100 million base pairs [98]!

To put this in perspective, compare this to the process of copying books. Before the invention of the printing press, medieval scribes used to make hand copies of the Bible and other texts. The Bible contains about 700,000 words, or about 3.5 million letters. So to be as accurate as DNA replication, a scribe would have to copy almost 30 Bibles with just a single letter mistake. (In truth, hand-copying of texts was notoriously error-prone and medieval scholars were known to grumble about the "foolish" mistakes of their scribes [99].)

Figure 1.62: **Medieval book copying (1148).** *A scribe would have needed to copy 30 Bibles with just a single mistake to be as accurate as the transmission of human DNA from one generation to the next.* *Credit: British Library article [Link]; Digitized Worms Bible [Link]; Public Domain.*

More mutations in older parents; more mutations in dads. In 1912, the German doctor Wilhelm Weinberg (of Hardy-Weinberg fame) reported that children with a skeletal defect called achondroplasia had older-than-average fathers. During the following 60 years, similar patterns were seen for several severe dominant diseases: namely, that the risk of disease increases with parental age, and especially with the ages of fathers.

Although it was not yet possible to sequence the mutations directly, these disease cases were interpreted as arising from *de novo* mutations in the parents. Here's an example from a 1987 paper, before the genes for achondroplasia and most other diseases were known:

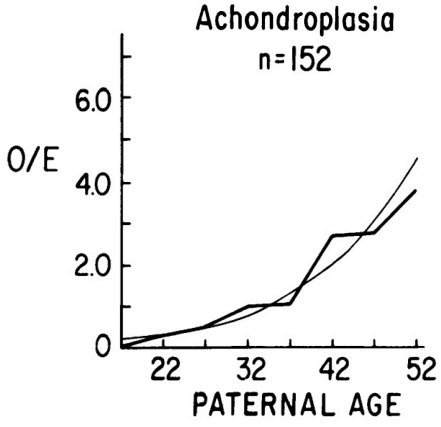

Figure 1.63: **Pre-genome era evidence that mutation rates increase with age (1987).** *The plot shows that rates of achondroplasia increase with paternal age; the y-axis is prevalence in each age-bin, divided by mean prevalence.* *Credit: Modified from Fig. 1 of Neil Risch et al (1987). [Link]*

These, and other, observations were taken as indirect evidence that the *de novo* mutation rate increases with age, and is higher in fathers [100] [101].

A century after Weinberg's work, this hypothesis was confirmed, with sequencing studies showing that

- The number of mutations per child increases roughly linearly with

parental age, with a much higher slope in dads;
• Dads have higher mutation rates at all ages (by a ratio of 3:1 if dad and mum are the same age [102]), as you can see below:

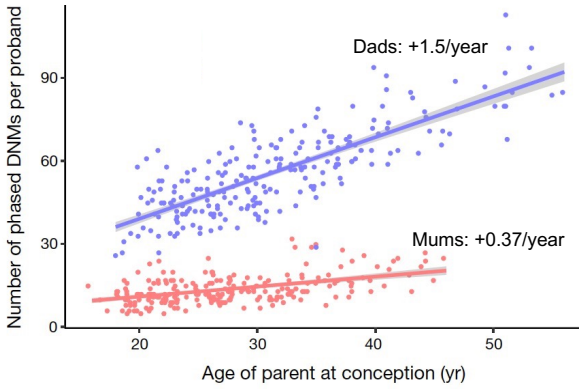

Figure 1.64: **Numbers of new mutations from each parent, as a function of parental age.** *Each data point shows counts for a single child. The increase in mutation counts as a function of parental age is statistically significant in both sexes.* *Credit: Modified Figure 1e of Hákon Jónsson et al (2017). [Link]*

This plot emphasizes that the kids of older parents (especially older dads) inherit a lot more mutations. In fact, if you look back at the histogram of the number of mutations inherited per child – ranging from around 40–120 – most of this variation can be explained by differences in the ages of the parents (plus random sampling variation) [103].

Mutation rates for other types of variation, including STRs and structural variants, are also higher in males than in females [104]. But there's one important exception to this rule: **chromosomal segregation errors** – such as Down Syndrome, caused by transmission of three copies of Chromosome 21 – **are mainly from meiosis errors in mums**. We'll cover this at the end of the chapter.

For decades, the higher mutation rates in males were believed to reflect the fact that there are more cell divisions in the male germline than in the female germline. But as we'll discuss shortly, recent work suggests that most mutations are due to spontaneous damage rather than cell divisions.

Mutation rates in somatic tissues. So far we have been talking about the rates of inherited (i.e., germline) mutations. Mutations also occur within the tissues of our bodies; these are important as drivers of cancer, and also likely contribute to some diseases of aging [105]. How do somatic mutation rates compare to germline mutation rates? How do they vary with age, and across tissues?

It turns out that it is much more difficult to study somatic mutations, than germline mutations. Since mutations occur very rarely in the genome (as low as 1 mutation per 10^8 base pairs), the sequencing error rates have to be extremely low, otherwise errors will outnumber true mutations detected. For studying inherited mutations, we can generally assume that all cells in a tissue sample from a child will carry the same mutations. In contrast, for somatic mutations, mutations that occur early in development may be carried by most or all of the cells in a tissue, but mutations

that occurred recently may be carried by only one or a few cells. Thus, to get accurate somatic mutation rate estimates we need to be able to detect mutations that are present in a single DNA molecule.

Recent techniques based on so-called *duplex sequencing* make this possible [106]. In short, both strands of the same DNA helix are used as independent templates for PCR amplification and sequencing. Variant nucleotides are only confirmed as mutations if they are observed from both strands.

Using these methods, recent work has provided the first direct measurements of somatic mutation rates [107]. As you see below, these tend to accumulate roughly linearly with age, similar to germline mutations. Overall rates are roughly 20–50 times higher than for germline, though still exceedingly low in absolute terms [108].

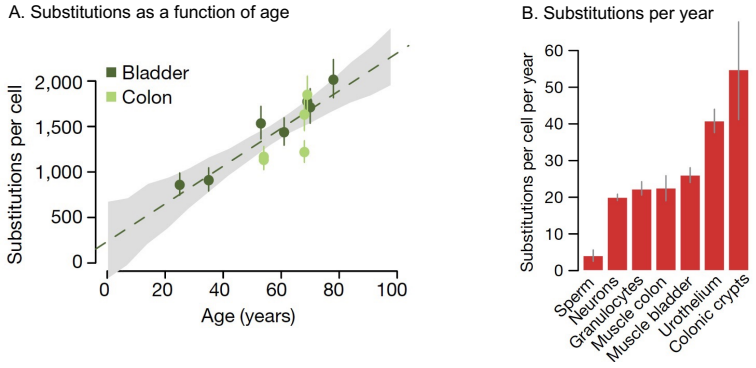

Figure 1.65: **Mutation accumulation in somatic tissues. A.** *Average numbers of mutations per cell in individuals of different ages for bladder and colon.* **B.** *Rates of mutation accumulation per year in different tissue types.*
Credit: From Figure 3 of Federico Abascal et al (2021). [Link] Used with permission.

One other key observation is that the mutation rates in different tissues are not strongly related to rates of cell division, suggesting that a large fraction of mutations arise from spontaneous damage instead of errors in DNA replication. For example, in the plot above, cortical neurons and urothelial cells undergo little or no cell division, but nonetheless have fairly typical mutation rates [109]. If you recall that a typical cell is estimated to suffer $\sim$70,000 genome lesions per day, this implies that only around one lesion per million actually results in mutation.

Mutation rates in cancer. There's one important exception to the rule that human mutation rates are very low: cancer.

Cancer refers to a collection of diseases in which somatic cells start to replicate in uncontrolled manner. In healthy tissues, cell replication and cell death are tightly constrained. As we shall see in Chapter 4.3, cancers are evolving systems that gain the ability to expand without the usual constraints. Typically, the transition into a full-blown cancer state involves multiple mutations across a collection of genes that suppress or enhance cell division. Mutations that enable faster cell division or metastasis are selectively favored within a developing cancer, even though they are severely detrimental to survival of the patient.

Consequently, some cancers arise cells with particularly high rates of exogenous damage; for example, skin cells that are exposed to UV light suf-

fer high rates of DNA damage (this is why you should wear sunscreen!). Secondly, many cancers actually evolve high mutation rates, by gaining mutations in DNA repair or proofreading genes. Cancer cells with higher mutation rates are more likely to accumulate additional key mutations that increase rates of cell division or metastasis. This leads to indirect positive selection on so-called *mutator* genotypes. The plot below shows numbers of mutations (per megabase) across a broad range of cancer types. At the high end, these numbers are 100–1000-fold higher than in healthy somatic tissue.

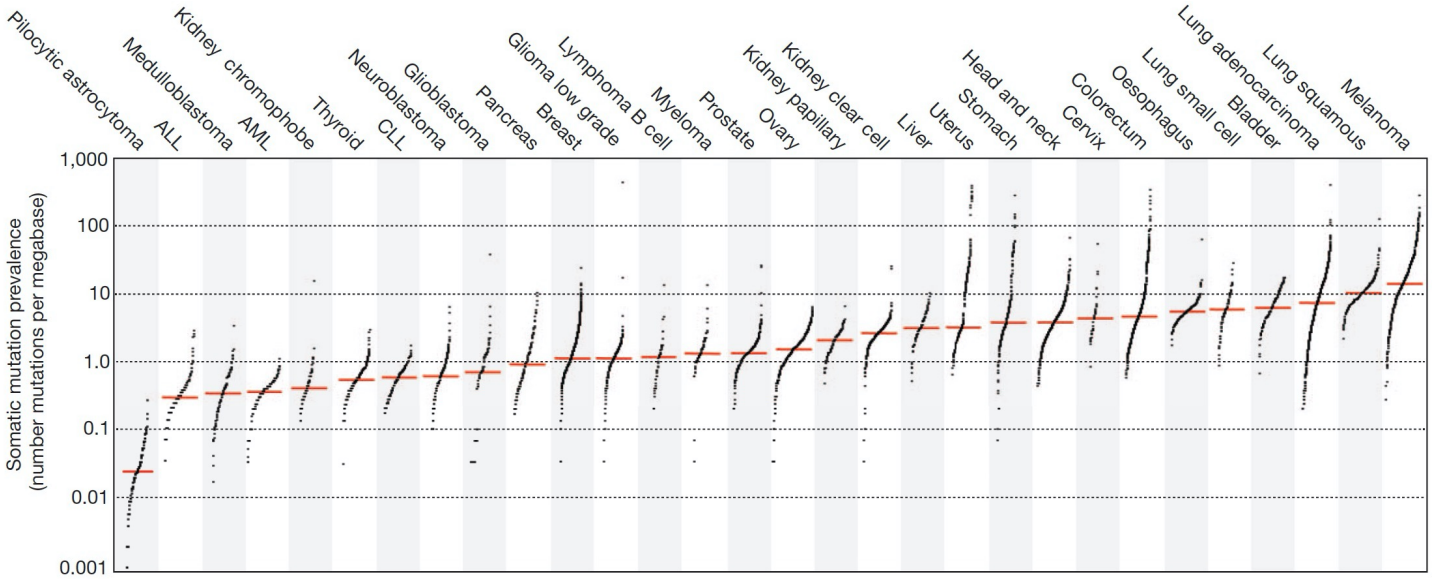

Figure 1.66: High mutation prevalence in cancer. *Mutation rates per megabase for different cancer types. Each data point shows the rate in a different patient; horizontal red lines show the median for each cancer. Notice that for many cancers the numbers are in the range of* 1–100 *mutations per megabase; higher than mutation numbers in healthy somatic tissue (*∼0.1–1 *per Mb).* *Credit: Figure 1 of Ludmil Alexandrov et al (2013). [Link] Used with permission.*

Types and mechanisms of mutations. Up to now, we've been focusing on single nucleotide mutations, and ignoring distinctions between different types of mutations. But you'll remember that the genome contains many different types of variation – including indels, STRs, and structural variants. These different types of mutations occur at widely varying rates, and this fact greatly influences the distribution of genetic variation and disease.

The table below shows estimates of germline rates for important subtypes of single nucleotide mutations, as well as a range of other events [110]. As we will explain shortly, the molecular mechanisms vary widely across different types of mutations, leading to widely varying rates.

As you can see below, single nucleotide mutations make up the majority of all mutations, but some other types of errors – notably STRs – occur at very high rates in particular sequence contexts. Meanwhile, although

structural variants occur at low rates, they are important because they can affect large genomic regions within a single mutational event:

Mutation Type	Number	Rate per site
Single Nucleotide	70	1.2×10^{-8}
Transition (non-CpG)	35	6.2×10^{-9}
Transversion (non-CpG)	21	3.8×10^{-9}
CpG Transition	12	1.1×10^{-7}
CpG Transversion	1	9.6×10^{-9}
Mitochondrial DNA	0.01	6×10^{-7}
Small Indel (70% deletion; 30% insertion)	~5	8.2×10^{-10}
Short Tandem Repeat* (30% contraction; 70% expansion)	>85	5×10^{-5} (average)
Structural Changes* (60% deletion; 30% duplication†)	>0.16	
Aneuploidy (live birth#)	~1/160	

Table 1.5: Genome-wide mutation mutation counts and rates. *Estimated average numbers of mutations per child, genomewide; all numbers are approximate and assume an average parental age of 30. Note that although structural mutations are relatively rare, they often affect tens of kilobases or more of DNA sequence.* *Estimates are from short-read data and detect a restricted subset of mutations, especially for structural variants. STR rates vary widely across motif lengths and types. †Other structural variants not listed include TE insertions and more-complex events. #Aneuploidy rates at fertilization are much higher. Credit: modified from an unpublished table by Ziyue Gao.

We'll now give a brief overview of types of mutations and mechanisms; this is a large and complex area, so my goal here is to give you an introduction to some of the key points, and not to be comprehensive.

Single nucleotide mutations: Transitions, transversions, and CpGs. Starting with single nucleotide ("point") mutations, the first key classification are transitions and transversions. To understand these, recall the chemical structure of DNA. Each rung in the DNA ladder contains a purine (A or G) paired with a pyrimidine (C or T). Purines have two rings, and pyrimidines have one ring. A **transition** switches one purine for another (on the other strand that's switching between pyrimidines); a **transversion** switches from purine to pyrimidine, or vice versa.

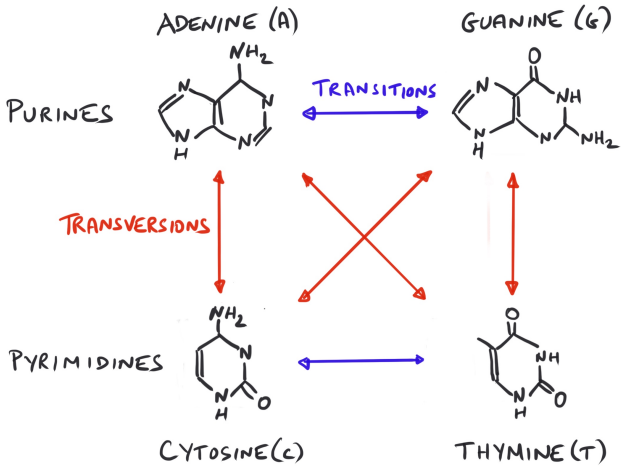

Figure 1.67: **Transitions and Transversions.** *Transition mutations switch between purines (2 rings) or between pyrimidines (1 ring); transversions switch between types. Transitions switch between similar molecules and occur at higher rates than transversions.*

The reason this matters is that because the purines (and similarly the

pyrimidines) resemble each other, the most frequent errors are transitions: i.e., they swap between purines, or between pyrimidines. If all possible point mutations occurred at equal rates, we would expect only 1/3 of mutations to be transitions (count the blue arrows versus red arrows, above). But transitions occur at nearly twice the rate of transversions, so that around 2/3 of point mutations are transitions.

There's one special type of point mutation that is particularly important: **CpG mutation**. In many organisms, including mammals, cytosine can optionally carry a methylation side group. In mammals this occurs almost exclusively when C and G occur at successive nucleotides: i.e., 5'–C–G–3', known as a "CpG". (The 'p' in CpG represents the phosphate that connects successive nucleotides on the same strand, and distinguishes this from the base pairing of G and C on opposite strands.) CpG methylation plays a critical role in preventing undesirable gene expression; consequently most CpGs in the genome are methylated except near the transcription start sites of expressed genes.

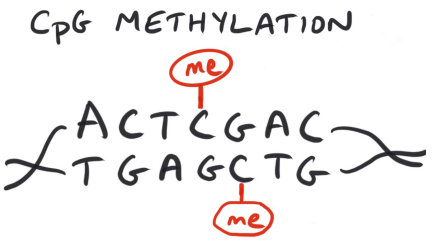

Figure 1.68: **CpG Methylation**. *Most cytosines in a CpG context carry an extra methyl group; cytosine methylation plays an essential role in gene silencing in mammals. Methylated Cs are highly mutable.*

This is relevant here because methylated cytosines can spontaneously convert to thymine. If these are not properly repaired, they cause C→T mutations. These mutations occur at a very high rate, ∼20-fold higher than other transition mutations.

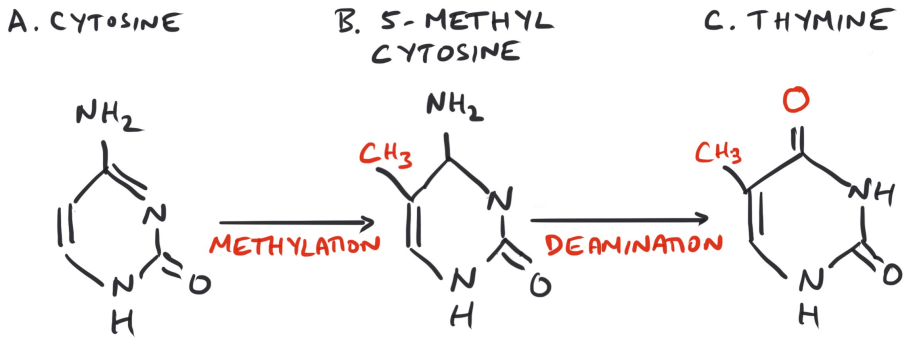

Figure 1.69: **Chemical structure of cytosine methylation and mutation.** *Most cytosines (**A**) become methylated (**B**) when they are in a CpG context. Methylated-C is prone to spontaneous deamination, which results in thymine (**C**). The thymine would be opposite a G on the other strand which tells the cell it must be repaired. Rare failures to repair the T result in C→T mutations.*

Another important special category is for **mitochondrial DNA**, in which mutation rates are even higher than at CpGs. Mitochondria evolved from bacterial symbionts early in the evolution of eukaryotes; they still maintain a small circular genome of 16 kb with 37 genes. DNA repair pathways in mitochondria are more limited than in the nuclear genome, and one important pathway (mismatch repair) may be absent [111]. Consequently the point mutation rate for mitochondrial DNA is about 50-fold higher than in the main genome, at about 6×10^{-7} per base pair per generation [112]. As we shall see later, the high mutation rate of mitochondrial DNA made it an important target of study for early work on human origins, when DNA sequencing was technically challenging and expensive (Chapters 3.2 and 3.3).

Short Tandem Repeats. Some of the highest error rates in the genome occur at short tandem repeats (STRs). Recall that STRs consist of long strings of a repeated motif such as CACACACA.... It turns out that it's difficult for cells to copy these long strings accurately. The main type of error consists of adding or subtracting one repeat. STR mutation rates

have been estimated at a rate of 3×10^{-4} per STR per generation for two-nucleotide repeats, and 1×10^{-3} for four-nucleotide repeats–making these mutation rates as much as a hundred thousand-fold higher than for single nucleotides[113]. Due to their extremely high mutation rate, STRs are highly variable from person-to-person. For this reason STRs are the most commonly-used genetic marker for "DNA fingerprinting" in forensics.

The high rates of STR mutation are due to a process known as **replication slippage**, in which one strand loops out during DNA replication, leaving one or more repeats unpaired:

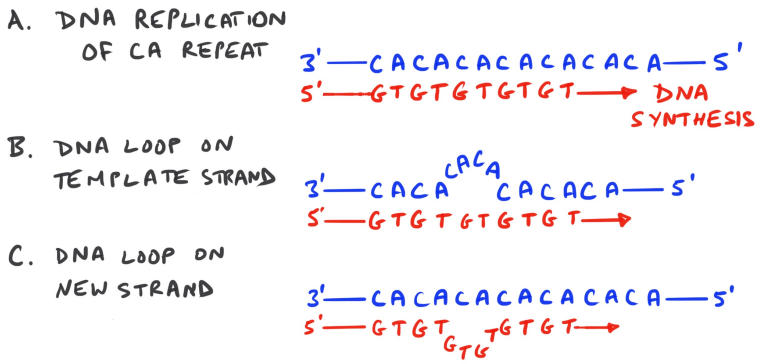

Figure 1.70: **Replication slippage model of STR mutation.** *During DNA replication one strand bubbles out to form a short single-stranded loop, with standard DNA base-pairing on either side of the loop. This causes either loss (B) or gain (C) of repeats, depending on whether the loop is on the template or replicating strand.*

Some STRs play important roles in functional variation and disease: for example in Chapter 1.3 we discussed an STR of CAG repeats within the coding sequence of the Huntingtin gene. The number of repeats is highly mutable; that's ok as long as the number of repeats stays within the normal range – up to 35 in this gene – but longer STRs cause Huntington's disease. Similarly, noncoding STRs sometimes affect gene regulation and contribute to complex traits [114].

Structural variants. Lastly, structural variants–including deletions, duplications, inversions, and more-complex changes in copy number – are another major feature of genetic variation and disease risk, and of genome evolution over longer timescales.

Broadly speaking, most structural mutations are likely due to a few main processes including **recombination errors** [115] and **DNA replication errors** [116]. For both of these processes, repetitive sequences play important roles in confusing the cellular machinery, leading to structural mutations. Alternatively, other events may arise from erroneous **double strand break repair** of damaged DNA, which does not require large-scale sequence homology and hence is not strongly clustered in repetitive regions [117].

The first of these mechanisms, i.e., recombination errors, occurs through a process called **NAHR (non-allelic homologous recombination)**. In NAHR, a DNA sequence that is repeated within a genomic region causes misalignment of homologous sequences during meiosis (i.e., the sequences are *homologous*, meaning that they are (nearly) identical copies of a single original sequence, but *non-allelic*, meaning that they are from distinct chromosomal locations. If a recombination event occurs within the mis-

aligned sequences, this leads to structural changes.

As can be seen here, deletions and duplications can be viewed as alternative products of NAHR, when the repeats are oriented in the same direction. Alternatively, NAHR between inverted repeats leads to inversions.

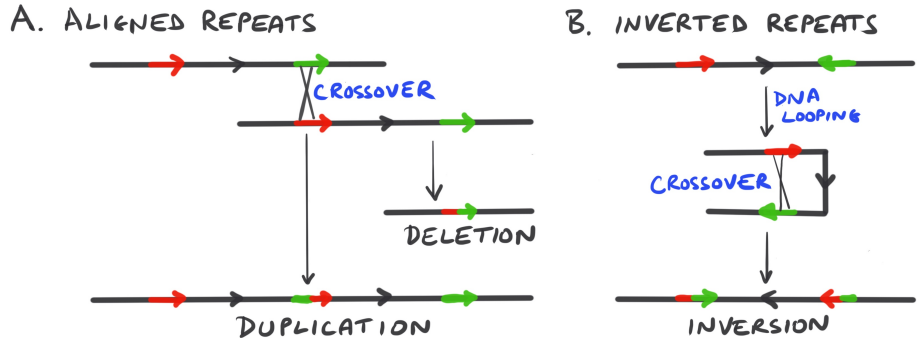

Figure 1.71: **NAHR model.** *A genomic region contains a pair of repeated elements, in green and red. If these misalign during meiosis, cross-over events lead to rearrangements: either deletion/duplication products if the elements are oriented in the same direction, or inversions if they are oriented in reverse orientations. In B, a loop has formed allowing incorrect cross-over between sequences within the same chromosome.* *Redrawn from Figure 1 of Lee and Lupski (2006) [Link]*

As a rule, deletions and duplications are more likely to have functional consequences, because they change **gene dosage** (i.e., the number of copies of genes contained within the region). Since expression of a gene is, usually, roughly proportional to its dosage, this can have functional consequences including possibly genetic diseases, if the affected region contains so-called **dosage-sensitive** genes. In contrast, inversions do not change copy number, and are less likely to cause major effects.

One genomic region that is susceptible to NAHR is at a locus known as 17p11.2 that is responsible for a pair of neurological disorders [c]. The cartoon below shows that the DNA sequence marked in white (CMT1a-REP) appears twice in the region, separated by 1.4 Mb. As discussed above, the two Chromosome 17 homologs can misalign at the repeated region during meiosis; if recombination takes place this, produces both a duplication and a deletion product. This event occurs at a rate[118] of about 10^{-4} to 10^{-5}, which is low in absolute terms, but far more frequent than specific point mutations.

[c] *The 17p11.2 notation uses a classical naming system for chromosome regions that were visible by microscopy prior to the genome sequencing era. This indicates a locus on the p-arm of Chromosome 17 at cytological band 11.2.*

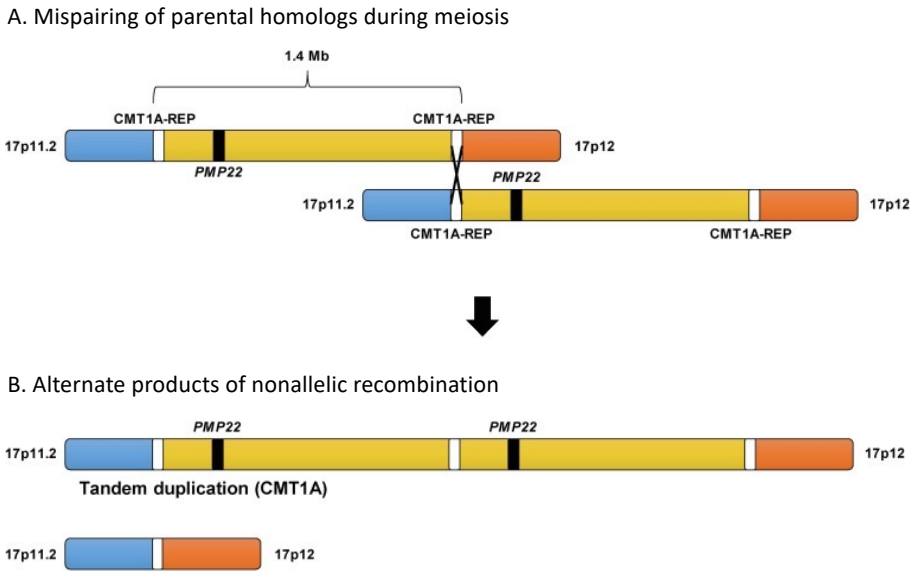

Figure 1.72: **NAHR mechanism at the Charcot-Marie-Tooth (17p11.2) locus.** *The CMT locus is shown as a series of colored blocks for each of the two parental homologs. A duplicated sequence (in white, labeled CMT1a-REP) is present twice, 1.4 Mb apart.* **A.** *During meiosis the duplicated region can mispair, potentially leading to NAHR.* **B.** *NAHR can produce two products: either the entire region is duplicated, or deleted.* *Credit: Modified Figure 1 from Harrison Pantera et al. 2020 [Link] Used with permission.*

The affected region contains a dosage-sensitive gene named PMP22 which

encodes a peripheral nerve myelin protein. Individuals with the duplication (leading to over-expression of PMP22) suffer from a peripheral neuropathy called Charcot-Marie-Tooth, while individuals with the deletion (and under-expression of PMP22) have a different neuropathy with distinct symptoms [119] [120].

The example above illustrates a common mechanism in which large low-copy repeats surround a dosage-sensitive gene, driving recurrent genetic disorders. That is a relatively simple example, but because repeats can drive structural mutations, repeat-dense regions can become crucibles of repeated structural mutations. In some genomic regions, the different haplotypes vary greatly in terms of overall structure, repeat content and orientation [121]. One such example is shown below:

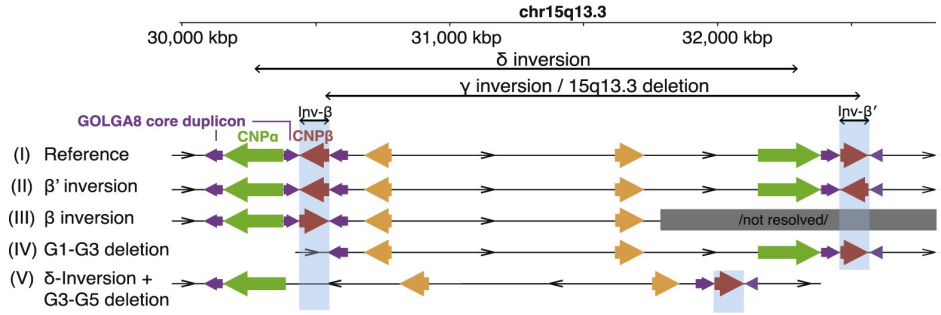

Figure 1.73: **Complex repeat structure at the Prader-Willi/Angelman Syndrome (15q13.3) locus.** *Repeated sequences are shown as directed colored arrows. Based on sequencing of healthy individuals, authors identified five common haplotypes that differ in content or orientation of repeat units.* ***Notice here the diversity of repeat structures and orientations across common haplotypes, typical of repeat-rich regions of the genome.*** *Credit: From Figure 5c of David Porubsky et al. 2022 [Link] CC-BY-NC)*

This region is also noteworthy because it is home to a pair of deletion syndromes called Prader-Willi and Angelman Syndromes [d]. These are caused by deletions that occur between the two maroon arrows, labeled CNPβ. Based on the NAHR model that we discussed above, the authors propose that this deletion may be restricted to haplotypes where the CNPβ arrows point in the same direction, as in Haplotype II.

In the last part of the chapter, we return to some broader topics about the overall patterns and distributions of mutations.

The puzzle of male-driven mutation. Early in this chapter we discussed the point that most mutations (around 75–80% in humans) come from fathers [122]? Why?

For decades there was a standard explanation for this. The key idea was that DNA replication during cell division is the main driver of mutation, and there are many more cell divisions in the male germline than in the female germline, as follows.

Both males and females go through about 30 rounds of cell division early on, as the embryo is developing. If the developing embryo is a girl, they develop into nearly-mature egg cells, and then stop development. Later, when she is an adult, each egg cell completes development right before it is released in ovulation. In contrast, in males, the germ cells that produce sperm stop dividing until shortly before puberty, but after puberty they continue to divide throughout life.

[d] *It's outside our main focus here but the 15q13.3 deletion syndromes have a remarkable inheritance pattern. When the deletion is inherited from the mother, it causes Angelman syndrome (developmental disabilities and motor defects); when it comes from the father, it causes Prader-Willi Syndrome (chronic overeating and related health issues). The difference arises because the region is imprinted: the gene UBE3 is only expressed from the maternal copy and loss of UBE3 causes Angelman; in contrast, SNRPN is only expressed from the paternal copy and its loss causes Prader-Willi [Link].*

This argument helps to explain why older dads transmit more mutations than younger dads. It's also tempting to explain the excess of mutations from dads compared to mums simply as a result of the far greater number of cell divisions in males.

But this calculation also suggests another prediction, namely that the *fraction* of paternal mutations should increase as the parents get older, because the male germline accumulates more and more cell divisions with age, while the female germline does not. Using modern data we can test this. Do we see this pattern in the data? Unfortunately for the number-of-cell-divisions model, we do not:

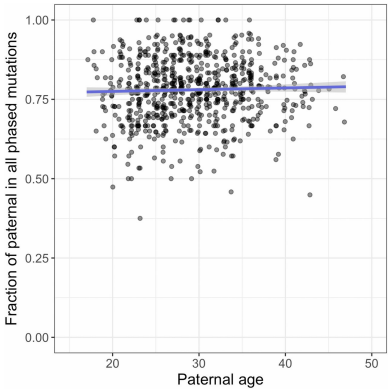

Figure 1.74: **Fraction of paternal mutations as a function of dad's age.** *Each point shows data for one child; the blue line is the regression fit. Under the model where this is controlled by number of cell divisions we would expect a strong positive slope; the fact that the slope is flat argues against this model.* *Note that the parental ages are matched in this analysis. Credit: Figure 1 from Ziyue Gao et al, 2019 [Link].*

As you see above, the fraction of mutations from dads is around 80% at all ages. Moreover, we now have data from many different mammals, with a wide range of generation times. In all these species males have more germline cell divisions than females, but the precise ratio varies widely depending on the specific details of development, age at puberty and reproduction. But, oddly enough, the fraction of paternal mutations is remarkably similar across all these species. These show only a very weak increase with generation time, across species whose generation times range from weeks to decades.

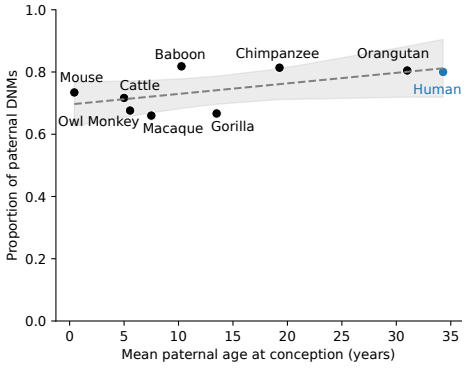

Figure 1.75: **Proportion of paternal mutations in different mammals as a function of generation time.** *This proportion is surprisingly consistent around 75% even though these species vary greatly in terms of generation time, and the ratio of male:female germline cell divisions.* *Credit: Kindly modified by Marc de Manuel Montero and Felix Wu, based on Figure 3B from Felix Wu et al, 2020 [Link] CC BY 4.0*

These results, as well as other analyses [123], suggest that the standard story based on number of cell divisions is not correct. Instead they point to a model where most mutations are not due to DNA replication, but are caused by DNA damage, accumulating steadily with age. (Remember that cells suffer thousands of lesions a day, and if only a tiny fraction of

that is not properly repaired it results in mutations.) We know that there is at least some contribution from non-replicative mutations (i.e., caused by damage), because the mutation rate increases with age even in mothers (albeit slower than in dads), even though mother's germ cells are not dividing.

In summary, the current data suggest that non-replicative mutations are the main driver of germline mutation, but we have to assume that these rates are about 3x higher in testes than in ovaries. It's not known yet why the rate is so much higher in testes, although this does seem to be a broadly conserved feature across at least mammals, birds and reptiles [124].

The puzzle of chromosome segregation errors. There's one huge exception to the rule that genome errors are rare, and male-biased, and that's for **aneuploidy** – i.e., cases where a cell does not carry the correct set of chromosomes: i.e., 23 pairs for a diploid human cell [e] [f].

In sharp contrast to mutations, aneuploidy is inherited mainly from mothers, especially older mothers. For example, around 93% of Down Syndrome cases (3 copies of Chromosome 21) come from chromosomal errors in the egg [125]. Furthermore, the rate of Down Syndrome increases dramatically with the age of the mum: from less than 0.1% in 20-year old mothers to around 1% at age 40 and 3% at age 45.

[e] *Aneuploidy is not a mutational process, but we cover it in this chapter under the broad umbrella of the types of genome alterations that can be transmitted to a zygote.*

[f] *See also Chapter 1.3 for more about aneuploidy.*

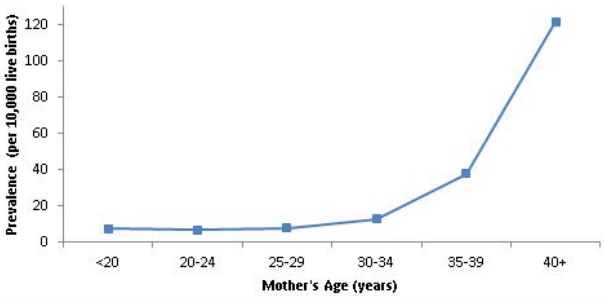

Figure 1.76: **The prevalence of Down Syndrome (Trisomy 21) increases rapidly with mother's age.** *Credit: CDC educational materials. [Link]. Public Domain.*

And Down Syndrome is really just the tip of the iceberg: it's possible for oocytes to carry gains or losses of any of the chromosomes. However, for most other possible aneuploidies, the resulting embryos fail to develop properly, let alone to survive to full term pregnancy.

It turns out that in older women a strikingly large fraction of oocytes carry at least one aneuploidy: by about age 45, more than 50% of oocytes in a typical woman carry chromosomal defects [126] [127]. In contrast, aneuploidy rates in sperm are around 1–4% [128]. As well as causing chromosomal disorders including Down Syndrome, these high rates of aneuploidy are a lead cause of infertility among older women:

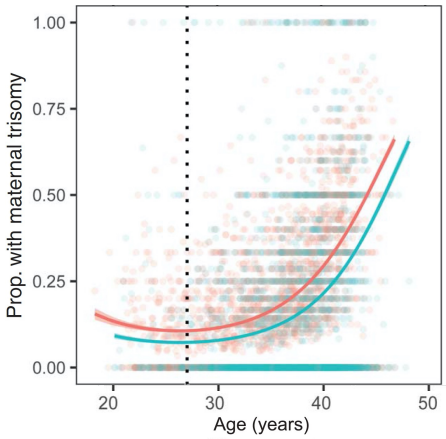

Figure 1.77: **High rates of trisomic oocytes in older women.** *The fitted curves show total trisomy rates across all chromosomes, at two developmental timepoints. Plotted points are for individual patients. As you can see, trisomy increases rapidly after age 35.* *Credit: From Figure 1e of Jennifer Gruhn et al (2019) [Link] Used with permission.*

Earlier in this chapter I emphasized how extraordinarily accurate DNA storage and replication are. Thus, by contrast, female meiosis is remarkably error-prone.

The molecular mechanisms for this are currently an active research area [129], but in broad strokes they are related to a very curious aspect of how egg cells develop in mammals. During fetal development, female germ cells migrate to the ovaries, where they undergo several rounds of mitotic cell division. A subset of the cells then enter meiosis to produce mature oocytes. Recall that meiosis is a process involving two rounds of cell division that produce haploid gametes.

Oddly enough, in normal female development, egg maturation halts in the middle of the first round of cell division, known as Meiosis 1. The oocytes must then wait, *for decades (!)*, until they are re-activated prior to ovulation. At that point, the homologous chromosomes are pulled apart to complete Meiosis 1. Meiosis 2 is completed later, upon fertilization.

While they are waiting to complete Meiosis 1, the homologous chromatids are tethered together by a protein complex called a kinetochore, as well as at crossover sites (which result in recombination). The chromatids sit in this tethered configuration for up to 40+ years until they are pulled apart by the meiotic spindle to complete cell division. It seems that multiple components of the meiotic machinery may deteriorate with age, including the kinetochore, and the assembly of the meiotic spindle [130] [131].

So, from an evolutionary point of view, why are aneuploidy rates in female meiosis so high? Curiously, it does not seem that female meiosis has evolved to minimize the rates of aneuploidy. A first line of evidence comes from analysis of crossover points. Crossovers in females are set up during fetal development and play an essential role in stabilizing the homologous chromatids for the completion of Meiosis 1. In human females (but not in males), about 25% of crossover sites are not fully assembled, and these incomplete cross-overs are a major driver of trisomy 21 [132].

Secondly, the meiotic spindles (which pull the chromatids apart) are actually less stable in human oocytes than in other mammals [133]. Somewhat perplexingly, this is because human oocytes do not express a key spindle-stabilizing protein, KIFC1, used by other mammals (and also used in

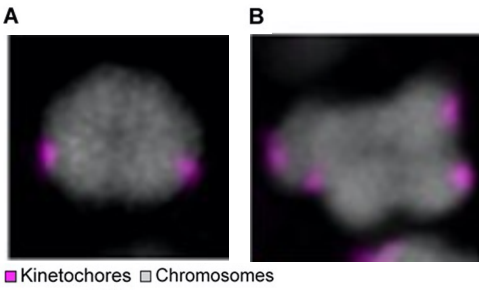

Figure 1.78: **Kinetochores (purple) can drift apart in human oocytes.** *In Meiosis 1 each chromosome pair consists of two pairs of sister chromatids; each of the four chromatids has its own kinetochore (purple).* **(A)** *When pairs of chromatids are tightly bound, there are two purple dots, one for each pair of sister chromatids;* **(B)** *when the sister chromatids drift apart, all four kinetochores can be seen. Separation of kinetochores increases with age and is thought to contribute to aneuploidy.* *Credit: Figure 2 from Agata Zielinska et al, 2015. [Link]*

human mitotic cells). This hints that human spindles have specifically evolved to be unstable.

It's not yet clear why meiosis may have evolved to be more error-prone than strictly necessary. One intriguing type of explanation is that oocytes are known to be susceptible to the evolution of "selfish" centromeres that hijack the process of meiosis to increase their chances of transmission (known as centromeric drive). Error-prone meiosis may evolve as either a consequence of centromeric drive, or as an antidote to it. For more on this, see [134]. A second type of explanation notes that maximizing fertility may not always lead to higher female fitness, especially in humans and other primates, which makes high investments in each offspring. In this hypothesis, since most aneuploidy leads to failure of implantation, aneuploidy serves to lower female fertility in an age-dependent fashion [135].

In summary, the genome is astonishingly well-protected against mutations; most inherited mutations come from fathers, and mutation rates increase with parental age in both sexes. In contrast, most aneuploidy comes from meiosis errors in older mothers, for reasons that are still not entirely clear.

In the next section of the book we will talk about the inheritance of mutations within families and within populations. Some mutations are inherited within populations for thousands of generations, or even eventually spread throughout an entire species.

Notes and References.

[89] Tubbs A, Nussenzweig A. Endogenous DNA damage as a source of genomic instability in cancer. Cell. 2017;168(4):644-56

[90] Gates KS. An overview of chemical processes that damage cellular DNA: spontaneous hydrolysis, alkylation, and reactions with radicals. Chemical Research in Toxicology. 2009;22(11):1747-60

[91] See Tubbs et al (2017), above.

[92] This paragraph touches on several complex topics. In most cases, natural selection pushes mutation rates to be as low as possible; exceptions include so-called 'mutator strains' in bacteria, as well as cancers, which generally evolve high mutation rates. There is presumably some molecular or physiological limit to how low mutation rates can be (it's also been argued that there may be a metabolic cost to having arbitrarily accurate DNA repair). However, Michael Lynch has argued that multi-celled organisms are generally not close to any fundamental limit because natural selection becomes ineffective when the mutation rate is low-enough. For reasons we'll explain in Chapter 2.6, this means that mutation rates are mainly determined through an interaction between selection and effective population size.

Sung W, Ackerman MS, Miller SF, Doak TG, Lynch M. Drift-barrier hypothesis and mutation-rate evolution. Proceedings of the National Academy of Sciences. 2012;109(45):18488-92

[93] I should also point out that it's an over-simplification to say that evolution does not act on long-term effects. As a thought experiment, imagine a species with a magical repair pathway that lowers the mutation rate to zero. In the short term, this new repair pathway would presumably be favored, as there would be no fitness cost due to mutations. But in the long term, this species could not adapt to changing environments, and would likely eventually go extinct.

[94] In practice, when we do genome sequencing, we're actually sequencing from a somatic tissue (usually blood). So this study-design potentially over-estimates the *de novo* mutation rate by including somatic mutations in the child. We can get a more accurate estimate by sequencing 3-generation pedigrees: we know that 50% of germline mutations should be transmitted to a grandchild in the third generation. It turns out that the 2- and 3-generation estimates are quite similar as few mutations occur early enough in somatic development to appear as heterozygous sites in sequencing of bulk tissue while not contributing to the germline.

[95] Roach JC, Glusman G, Smit AF, Huff CD, Hubley R, Shannon PT, et al. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. Science. 2010;328(5978):636-9;

Ségurel L, Wyman MJ, Przeworski M. Determinants of mutation rate variation in the human germline. Annu Rev Genomics Hum Genet. 2014;15(1):47-70

[96] 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;488(7412):471-5

Jónsson H, Sulem P, Kehr B, Kristmundsdottir S, Zink F, Hjartarson E, et al. Parental influence on human germline de novo mutations in 1,548 trios from Iceland. Nature. 2017;549(7673):519-22

[97] Agarwal I, Fuller ZL, Myers S, Przeworski M. Relating pathogenic loss-of function mutations in humans to their evolutionary fitness costs. bioRxiv. 2022

[98] Great thread about how amazing DNA replication is: [Link].

[99] E.g., Amos van Baalen writes about medieval copying errors; in one cited example: *In his Latin poem 'On Scribes', the English scholar Alcuin of York (c. 740–804) admonishes scribes to "take care not to insert their silly remarks" and that "their hands not make mistakes through foolishness".* [Link].

[100] Weinberg W. Zur vererbung des zwergwuchses. Arch Rassen-u Gesel Biolog. 1912;9:710-8

Crow JF, Denniston C. Mutation in human populations. Advances in Human Genetics 14. 1985:59-123

Risch N, Reich E, Wishnick M, McCarthy J. Spontaneous mutation and parental age in humans. American Journal of Human Genetics. 1987;41(2):218

[101] It was also inferred from studies of sequence evolution of the X, Y and autosomes, that mutation rates are higher in males; eg

Shimmin LC, Chang BHJ, Li WH. Male-driven evolution of DNA sequences. Nature. 1993;362(6422):745-7

[102] Gao Z, Moorjani P, Sasani TA, Pedersen BS, Quinlan AR, Jorde LB, et al. Overlooked roles of DNA damage and maternal age in generating human germline mutations. Proceedings of the National Academy of Sciences. 2019;116(19):9491-500

[103] About 70% of the variance in de novo mutation count is explained by parental age

Kaplanis J, Ide B, Sanghvi R, Neville M, Danecek P, Coorens T, et al. Genetic and chemotherapeutic influences on germline hypermutation. Nature. 2022;605(7910):503-8.

[104]Structural variation: Belyeu JR, Brand H, Wang H, Zhao X, Pedersen BS, Feusier J, et al. De novo structural mutation rates and gamete-of-origin biases revealed through genome sequencing of 2,396 families. The American Journal of Human Genetics. 2021;108(4):597-607. STRs: Mitra I, Huang B, Mousavi N, Ma N, Lamkin M, Yanicky R, et al. Patterns of de novo tandem repeat mutations and their role in autism. Nature. 2021;589(7841):246-50

[105]One emerging theme in cancer biology is that most aging tissues are susceptible to clonal expansions of specific cell lineages with proliferative advantages. An example where this contributes to aging is through clonal expansions in immune cells and their link to CAD:

Jaiswal S, Natarajan P, Silver AJ, Gibson CJ, Bick AG, Shvartz E, et al. Clonal hematopoiesis and risk of atherosclerotic cardiovascular disease. New England Journal of Medicine. 2017;377(2):111-21

[106]Kennedy SR, Schmitt MW, Fox EJ, Kohrn BF, Salk JJ, Ahn EH, et al. Detecting ultralow-frequency mutations by Duplex Sequencing. Nature Protocols. 2014;9(11):2586-606

[107]Abascal F, Harvey LM, Mitchell E, Lawson AR, Lensing SV, Ellis P, et al. Somatic mutation landscapes at single-molecule resolution. Nature. 2021;593(7859):405-10

[108]To put this in context, the highest mutation rate of nearly 60 per year implies around 1 mutation per 100 million base pairs.

[109]See again Abascal et al (2021)

[110]Single nucleotide variation: Kong et al (2012), Jonsson et al (2017); Indels: Jonsson et al (2017); Structural variation: Belyeu et al (2021); STRs: Sun et al (2012), Mitra et al (2021), Steely et al (2021), Kristmundsdottir et al (2023). Mitochondrial DNA: Fu et al (2013)–converted from rate per year assuming a generation time of 30 years. References not given previously:

Sun JX, Helgason A, Masson G, Ebenesersdóttir SS, Li H, Mallick S, et al. A direct characterization of human mutation based on microsatellites. Nature Genetics. 2012;44(10):1161-5

Steely CJ, Watkins S, Baird L, Jorde L. The Mutational Dynamics of Short Tandem Repeats in Large, Multigenerational Families. bioRxiv. 2021

Kristmundsdottir S, Jonsson H, Hardarson MT, Palsson G, Beyter D, Eggertsson HP, et al. Sequence variants affecting the genome-wide rate of germline microsatellite mutations. Nature Communications. 2023;14(1):3855

Fu Q, Mittnik A, Johnson PL, Bos K, Lari M, Bollongino R, et al. A revised timescale for human evolution based on ancient mitochondrial genomes. Current Biology. 2013;23(7):553-9

[111]Fontana GA, Gahlon HL. Mechanisms of replication and repair in mitochondrial DNA deletion formation. Nucleic Acids Research. 2020;48(20):11244-58

[112]Fu et al (2013), cited above.

[113]Sun et al (2014), cited above

[114]Gymrek M, Willems T, Guilmatre A, Zeng H, Markus B, Georgiev S, et al. Abundant contribution of short tandem repeats to gene expression variation in humans. Nature Genetics. 2016;48(1):22-9

[115]Carvalho C, Lupski JR. Mechanisms underlying structural variant formation in genomic disorders. Nature Reviews Genetics. 2016;17(4):224-38

[116]The second major class of mechanisms is due to errors in DNA replication and repair. These are much more complicated than NAHR, and involve a variety of different pathways. These include mis-templating of repetitive regions during DNA replication, or during repair of replication errors. See Carvalho and Lupski (2016) and see:

Ottaviani D, LeCain M, Sheer D. The role of microhomology in genomic structural variation. Trends in Genetics. 2014;30(3):85-94

[117]These mechanisms involve non-homologous end joining or micro-homology mediated end joining. See:

Kidd JM, Graves T, Newman TL, Fulton R, Hayden HS, Malig M, et al. A human genome structural variation sequencing resource reveals insights into mutational mechanisms. Cell. 2010;143(5):837-47

[118]I cannot find a rate estimate, but the prevalence of CMT is about 1/2500 births, and the 17p11.2 locus is reported to be responsible for nearly half of cases.

[119]Hereditary Neuropathy with Liability to Pressure Palsies

[120]The Charcot-Marie Tooth locus was the first genetic disorder to be found that is usually due to structural variation, in 1992:

Roa BB, Garcia CA, Pentao L, Killian JM, Trask BJ, Suter U, et al. Evidence for a recessive PMP22 point mutation in Charcot–Marie–Tooth disease type 1A. Nature Genetics. 1993;5(2):189-94

An interesting footnote to the story is that the PMP22 gene was discovered by a team led by James Lupski. Lupski, a pioneer in studies of structural variation, is himself affected by Charcot-Marie Tooth syndrome; however Lupski's genome sequence showed that his own symptoms are due to mutations in a different gene: described here: [Link], and here:

Lupski JR, Reid JG, Gonzaga-Jauregui C, Rio Deiros D, Chen DC, Nazareth L, et al. Whole-genome sequencing in a patient with Charcot–Marie–Tooth neuropathy. New England Journal of Medicine. 2010;362(13):1181-91

[121]Porubsky D, Höps W, Ashraf H, Hsieh P, Rodriguez-Martin B, Yilmaz F, et al. Recurrent inversion polymorphisms in humans associate with genetic instability and genomic disorders. Cell. 2022;185(11):1986-2005

[122]Key recent work on this problem comes from Molly Przeworski's lab: Gao et al (2019), cited above, and:

Gao Z, Wyman MJ, Sella G, Przeworski M. Interpreting the dependence of mutation rates on age and time. PLoS biology. 2016;14(1):e1002355

Wu FL, Strand AI, Cox LA, Ober C, Wall JD, Moorjani P, et al. A comparison of humans and baboons suggests germline mutation rates do not track cell divisions. PLoS Biology. 2020;18(8):e3000838,

de Manuel M, Wu FL, Przeworski M. A paternal bias in germline mutation is widespread across amniotes and can arise independently of cell divisions. bioRxiv. 2022

[123]Wu et al (2020) and de Manuel et al (2022), cited above.

[124]The ratio is around 3:1 in mammals and 2:1 in birds and reptiles: de Manuel et al (2022) [Link]

[125]Vraneković J, Božović IB, Grubić Z, Wagner J, Pavlinić D, Dahoun S, et al. Down syndrome: parental origin, recombination, and maternal age. Genetic Testing and Molecular Biomarkers. 2012;16(1):70-3

[126]Kuliev A, Zlatopolsky Z, Kirillova I, Spivakova J, Janzen JC. Meiosis errors in over 20,000 oocytes studied in the practice of preimplantation aneuploidy testing. Reproductive biomedicine online. 2011;22(1):2-8

[127]Gruhn et al (2019), from which the figure is taken, proposes that the small uptick at younger ages is a real effect, and is due to a distinct signature of Meiosis 1 errors that declines with age; however this a very weak signal compared to the primary signature of increased aneuploidy at older ages.

Gruhn JR, Zielinska AP, Shukla V, Blanshard R, Capalbo A, Cimadomo D, et al. Chromosome errors in human eggs shape natural fertility over reproductive life span. Science. 2019;365(6460):1466-9

[128]Greaney J, Wei Z, Homer H. Regulation of chromosome segregation in oocytes and the cellular basis for female meiotic errors. Human Reproduction Update. 2018;24(2):135-61

[129]This section greatly simplifies a complex field. For more on this, you can start with: Greaney et al (2018), cited above;

Nagaoka SI, Hassold TJ, Hunt PA. Human aneuploidy: mechanisms and new insights into an age-old problem. Nature Reviews Genetics. 2012;13(7):493-504

Webster A, Schuh M. Mechanisms of aneuploidy in human eggs. Trends in cell biology. 2017;27(1):55-68

[130]Zielinska AP, Holubcova Z, Blayney M, Elder K, Schuh M. Sister kinetochore splitting and precocious disintegration of bivalents could explain the maternal age effect. Elife. 2015;4:e11389

Patel J, Tan SL, Hartshorne GM, McAinsh AD. Unique geometry of sister kinetochores in human oocytes during meiosis I may explain maternal age-associated increases in chromosomal abnormalities. Biology Open. 2016;5(2):178-84

[131]One interesting aspect of this is that cross-overs play an important role in tethering the sister chromatids. Even though the crossovers (i.e., recombination events) are set up during fetal development, it turns out that children of older mothers have more maternal crossovers. This suggests that oocytes with more cross-overs are more likely to be non-aneuploid, and thus to produce successful pregnancies.

Wang S, Hassold T, Hunt P, White MA, Zickler D, Kleckner N, et al. Inefficient crossover maturation underlies elevated aneuploidy in human female meiosis. Cell. 2017;168(6):977-89

[132]Wang et al (2017), cited above.

[133]So C, Menelaou K, Uraji J, Harasimov K, Steyer AM, Seres KB, et al. Mechanism of spindle pole organization and instability in human oocytes. Science. 2022;375(6581):eabj3944

Bennabi I, Terret ME, Verlhac MH. Meiotic spindle assembly and chromosome segregation in oocytes. Journal of Cell Biology. 2016;215(5):611-9

[134]Centromeric drive:

Zwick ME, Salstrom JL, Langley CH. Genetic variation in rates of nondisjunction: association of two naturally occurring polymorphisms in the chromokinesin nod with increased rates of nondisjunction in Drosophila melanogaster. Genetics. 1999;152(4):1605-14

Malik HS. The centromere-drive hypothesis: a simple basis for centromere complexity. Centromere. 2009:33-52

Kursel LE, Malik HS. The cellular mechanisms and consequences of centromere drive. Current opinion in cell biology. 2018;52:58-65

Lampson MA, Black BE. Cellular and molecular mechanisms of centromere drive. In: Cold Spring Harbor symposia on quantitative biology. vol. 82. Cold Spring Harbor Laboratory Press; 2017. p. 249-57

Hurst LD. Selfish centromeres and the wastefulness of human reproduction. PLoS Biology. 2022;20(7):e3001671

[135]This model notes that aneuploidy can increase the gap between successive children to allow greater maternal care for each child, and to reduce fertility in older women who might otherwise care for their existing children or grandchildren. In this view, incomplete crossovers are a feature, not a bug of the system. It's hard to rule out this type of explanation, but it strikes me as a rather clumsy physiological mechanism to regulate fertility. Wang et al (2017), cited above.