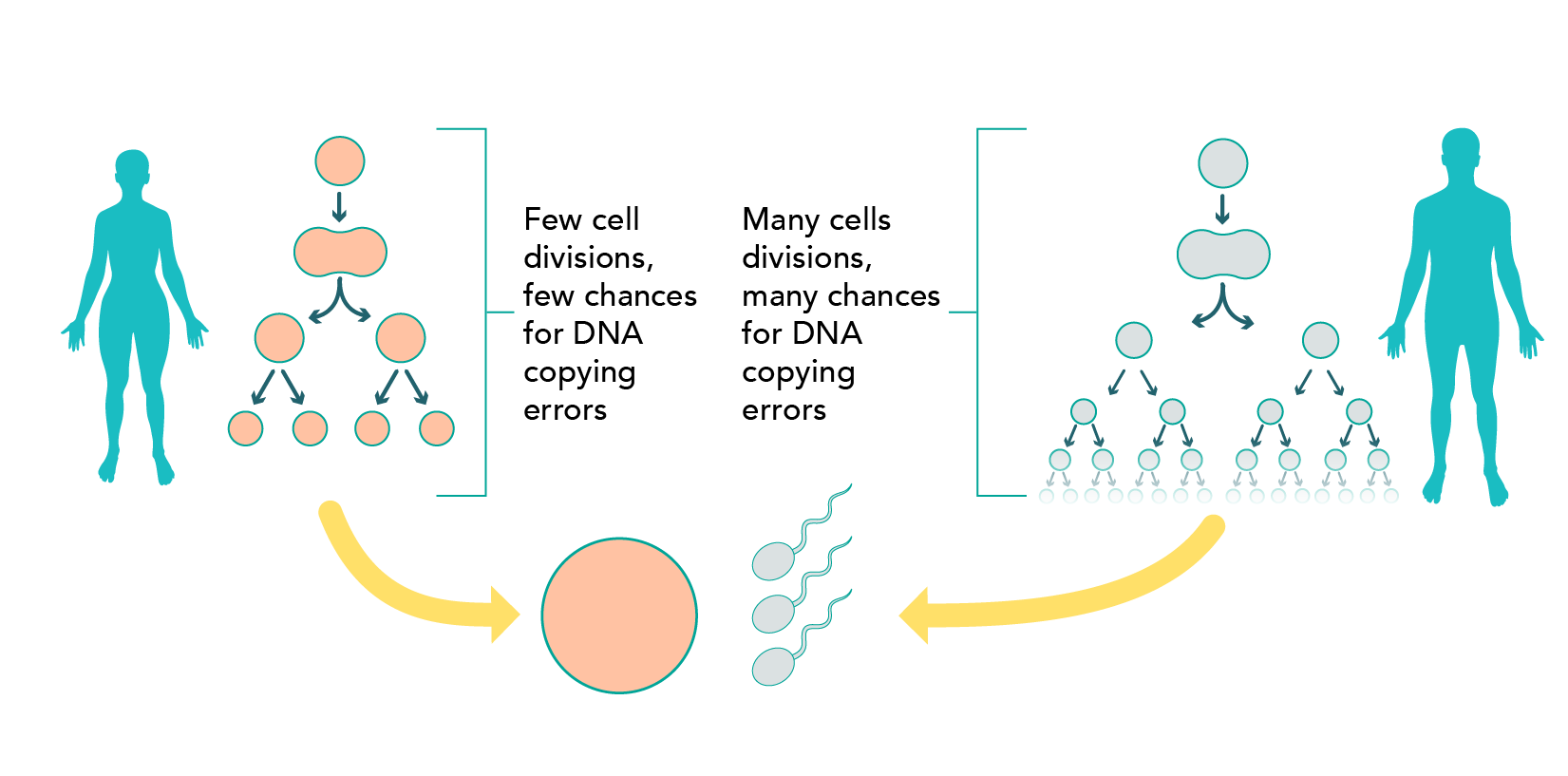
Timeline

Description automatically generated

**Digging Data: Mutation**

Mutations – changes in the genetic sequence of DNA or RNA – are the raw material for evolution. Natural selection, genetic drift, and other evolutionary processes act on genetic variation – and that genetic variation starts with mutation. Even if a genetic variant is introduced to a population through migration, ultimately, that variant got its start as a mutation. So understanding where, when, why, and how often mutations occur is key in understanding how evolution happens. Today, quick and inexpensive DNA sequencing means biologists can take a closer look than ever before at the process of mutation. What they are learning is sometimes surprising.

In humans, each baby has around 70 brand new or “de novo” mutations. De novo mutations occur in the reproductive cells of parents and are passed on to the child. Evidence suggests that most de novo mutations in a child come from the sperm that helped create that child, and relatively few mutations come from the egg. Biologists thought this made sense. In humans, beginning at puberty, the cells that produce sperm divide (and copy their DNA) throughout adulthood, leading to vast numbers of sperm. In contrast, in a person with ovaries, all the DNA copying leading up to egg production is completed before that person is even born.[[1]](#footnote-1) The cells that produce sperm just go through many more cycles of DNA replication and cell division than do the cells that produce eggs.[[2]](#footnote-2) If most mutations happen because a cell makes an error when it copies its DNA, producing a new DNA strand that differs slightly in sequence compared to the original, then we’d expect sperm to be the source of most new mutations. All those cell divisions in the cells that eventually lead to sperm provide many opportunities for copying mistakes to occur.



Biologists Felix Wu, Alva Strand, Laura Cox, Carole Ober, Jeffrey Wall, Priya Moorjani, and Molly Przeworski work at different universities and research centers, but they all wanted to know, **is it *really* the case that most new mutations in humans are caused by copying errors when cells divide?** Some evidence had already suggested that copying errors are not the whole story when it comes to heritable mutations.

A person smiling for the camera

Description automatically generated with medium confidence A person smiling for the camera

Description automatically generated with medium confidence  A person smiling for the camera

Description automatically generated with medium confidence

A person wearing glasses

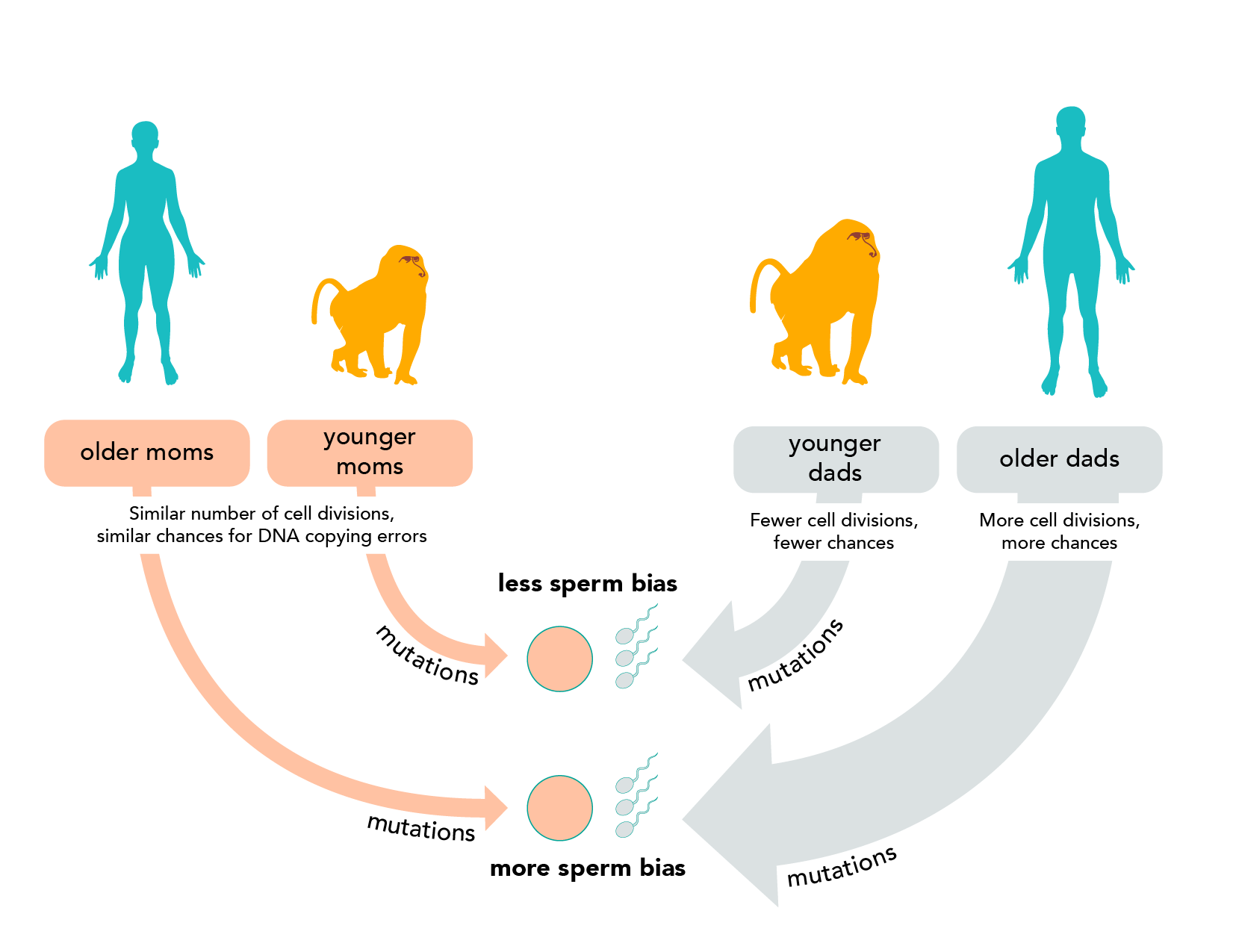
Description automatically generated with medium confidenceA person smiling for the camera

Description automatically generated with medium confidenceA person with blonde hair

Description automatically generated with low confidence

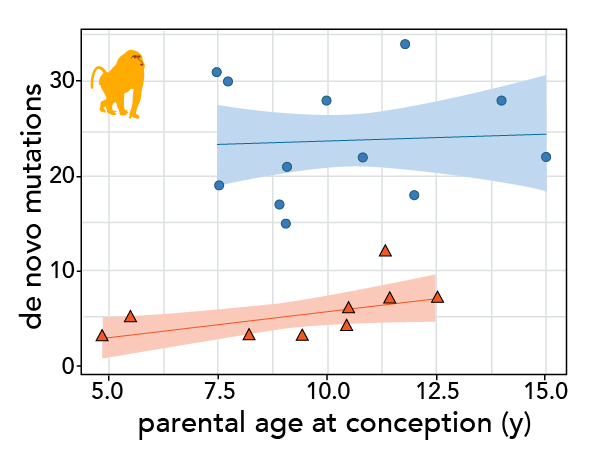
*The research team. Top row: Felix Wu, Alva Strand, Laura Cox, Carole Ober. Bottom row: Jeff Wall, Priya Moorjani, Molly Przeworski. Photos reproduced with permission from the individuals.*

**Hypothesis**: The research team set out to test the hypothesis that **most de novo mutations occur because of copying errors**. They looked at what percent of new mutations in a baby come from DNA carried by sperm, as opposed to from DNA carried by the egg; we’ll call this “sperm bias” in new mutations. If new mutations mostly come from copying errors, then the number of mutations contributed by eggs should not be much affected by the age of the person contributing the egg (since, in primates, all the DNA copying that leads up to egg production occurs before that future parent is even born). However, the number of mutations contributed by the sperm *should* increase with the age of the genetic father (since DNA copying and cell divisions leading to sperm are ongoing from puberty throughout adulthood). Overall, this means that **if the hypothesis were true, we’d expect to see that species with younger genetic fathers (and a shorter interval between puberty and fatherhood) should have less sperm bias in new mutations than species with older genetic fathers (and a longer interval between puberty and fatherhood)**.



**Data**: The team decided to compare mutations in humans to mutations in olive baboons (*Papio anubis*). The two primates are closely related, but male baboons go through puberty at age 6 and reproduce at age 10 on average, while humans that produce sperm typically start puberty around age 13 and don’t reproduce until age 32 on average. This means that the cells that produce baboon sperm go through about 4 years’ worth of DNA replication and cell division before a sperm leads to offspring, while the cells that produce human sperm go through about 19 years’ worth of DNA replication and cell division before a sperm leads to a child! That’s a big difference, providing lots more opportunity for mutations caused by DNA copying errors to accumulate in the sperm-producing cells of humans compared to those of baboons. To get data on mutations, the team sequenced the genomes of three generations of three different human families (26 people total) and three generations of two baboon families (29 baboons total).

They focused on tallying up selected spots in the genome where the DNA sequences of the parents were the same, but their child’s genetic sequence was different from that of both parents.[[3]](#footnote-3) The only way to explain this outcome is if one of the parents contributed a brand new mutation to the child. Then the researchers used additional sequence information to determine which parent (mother or father) had contributed each mutation.[[4]](#footnote-4) This is what they found:

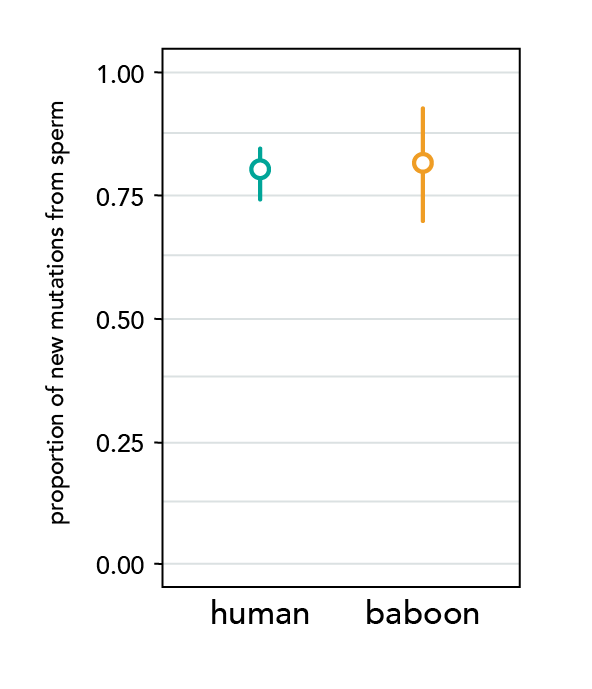
Chart

Description automatically generated

*Graph 1 Graph 2*

In these graphs, each offspring in which new mutations were identified is represented by a circle and a triangle. The circle indicates the number of mutations contributed by the sperm and the triangle indicates the number contributed by the egg. The lines show the overall relationship between age at conception (on the *x*-axis) and mutations (on the *y*-axis). The shading around these regression lines indicates the 95% confidence intervals. Notice that the 95% confidence interval around the male baboon regression line is much wider than the interval around the other regression lines.

The researchers used their data to calculate the proportion of all new mutations that came from the father (i.e., were paternal), as shown in the graph below. The vertical lines represent the 95% confidence intervals for each point. A chi-squared test showed that the two values are not significantly different (p = 0.91).



*Graph 3*

**Stepping into science**: The research team was led by Felix Wu, a graduate student at Columbia University in New York City, and Molly Przeworski, a professor there. Felix is not a native city boy. He spent a large part of his childhood running around the woods of Vermont. When stuck indoors, he would work on origami, mess around with the Rubik’s cube, and read fantasy/sci-fi novels. He still does some of those things (though not as often as he would like), but being an evolutionary biologist allows him to bring together that love of nature, solving puzzles, and storytelling into a single job. Molly did grow up in a large city and spent most of her leisure time reading novels. She was utterly uninterested in math or science—her least favorite topic in school was biology. But a great teacher drew her into math in college and after a series of false starts, she discovered genetics and evolutionary biology.

**Reference**: Wu, F.L., Strand, A.I., Cox, L. A., Ober, C., Wall, J. D., Moorjani, P., and Przeworski, M. (2020) A comparison of humans and baboons suggests germline mutation rates do not track cell divisions. *PLoS Biol*. 18(8): e3000838. https://doi.org/10.1371/journal.pbio.3000838

**Comprehension questions:**

1. Based on the first paragraph, what is the source of all genetic variation?
2. Based on the first paragraph, why is genetic variation important to evolution?
3. From which gamete (egg or sperm) do most de novo mutations in human babies come?

The researchers tested a hypothesis to explain observed patterns in de novo mutations.

What was that hypothesis?

* + 1. Explain how the cause of mutation described by the hypothesis above could lead to a different number of mutations carried by sperm compared to eggs.
    2. If the hypothesis above were true, what would you expect to observe regarding the effect of paternal age on both the number of mutations carried by sperm and on sperm bias (the difference between the number of mutations contributed by sperm compared to eggs)?

1. In general terms, how did the team decide to test their hypothesis?
2. If their hypothesis were true, what would the team expect to observe regarding sperm bias in humans and baboons and why?

**Data interpretation questions:**

1. Compare the data on Graphs 1 and 2.
   * 1. What similarities do you see?
     2. Based on these data, do you think there is sperm bias in new mutations in baboons and humans? What aspects of the data lead you to this interpretation?
     3. Based on these data, how does parental age affect the number of mutations contributed by eggs and sperm? What aspects of the data lead you to this interpretation?
     4. The regression line for the baboon fathers in Graph 2 does not slope upwards and also has a wider confidence interval than the other three regression lines. What aspect of the data might explain the wider confidence interval on that regression line? How might this affect our ability to discern patterns in the data? Refer to specific parts of Graph 2 in your response.
2. After looking at Graph 3, a student concludes that most mutations do not come from copying errors. Do you agree with the student? Why or why not? Refer to Graph 3 in your explanation.
3. An alternative hypothesis about the source of de novo mutations is that they might mostly occur when DNA is damaged and incorrectly repaired.
   * 1. If DNA damage and faulty repair occur at a constant rate over time in the cells that ultimately produce eggs and sperm, and this were the cause of most de novo mutations, what relationship would you expect to observe between the age of parents and the number of de novo mutations contributed?
     2. Do any of the data in the graphs support or contradict the expectation of the alternative hypothesis described above? If so, describe the data and explain how they relate to the alternative hypothesis.
     3. Based on the alternative hypothesis, would you expect to see a difference between the number of mutations contributed by sperm and eggs? Explain your reasoning.

1. The initial stages of oogenesis (egg production) occur during fetal development before birth. During this stage, the DNA that will wind up in eggs is copied in preparation for cell division – and it is during this stage that mutations caused by copying errors might occur. However, this process is then halted before any cell division takes place. The final cell divisions that produce eggs will actually occur during the reproductive years, during the menstrual cycle and fertilization – but no DNA replication occurs at this time. [↑](#footnote-ref-1)
2. Technically, we are focused on DNA replication in the stem cells that produce eggs and sperm. The stem cells that eventually lead to sperm go through more cycles of DNA replication than do the stem cells that produce eggs, providing more opportunities for DNA replication errors to occur. [↑](#footnote-ref-2)
3. The researchers focused on single nucleotide polymorphisms (SNPs) – single base pair differences. To be extra certain that they were detecting brand new mutations, they focused on genome positions where the child was a heterozygote and both parents were homozygotes. They then checked their data using several different methods. [↑](#footnote-ref-3)
4. This is trickier to do than it might sound. After all, primates are diploid; if a primate offspring shows up with a mutation on one chromosome out of a pair, how do we tell which chromosome, the one from the egg or the one from the sperm, contributed that mutation? Figuring it out requires looking at nearby positions on the chromosome – i.e., figuring out the genetic background on which the mutation most likely occurred. If the new mutation is located on a chromosome near a signature sequence present only in the genetic father, we know that person must have contributed the mutation. [↑](#footnote-ref-4)