

GENOMICS

'Platinum' genome shapes up

Disease sites targeted in assembly of more-complete version of the human genome sequence.

BY EWEN CALLAWAY

Geneticists have a dirty little secret. More than a decade after the official completion of the Human Genome Project, and despite the publication of multiple updates, the sequence still has hundreds of gaps — many in regions linked to disease. Now, several research efforts are closing in on a truly complete human genome sequence, called the platinum genome.

"It's like mapping Europe and somebody says, 'Oh, there's Norway. I really don't want to have to do the fjords,'" says Ewan Birney, a computational biologist at the European Bioinformatics Institute near Cambridge, UK, who was involved in the Human Genome Project. "Now somebody's in there and mapping the fjords."

The efforts, which rely on the DNA from peculiar cellular growths, are uncovering DNA sequences not found in the official human genome sequence that have potential links to conditions such as autism and the neurodegenerative disease amyotrophic lateral sclerosis (ALS).

In 2000, then US President Bill Clinton joined leading scientists to unveil a draft human genome. Three years later, the project was declared finished. But there were caveats: that human 'reference' genome was more than 99% complete, but researchers could not get to 100% because of method limitations.

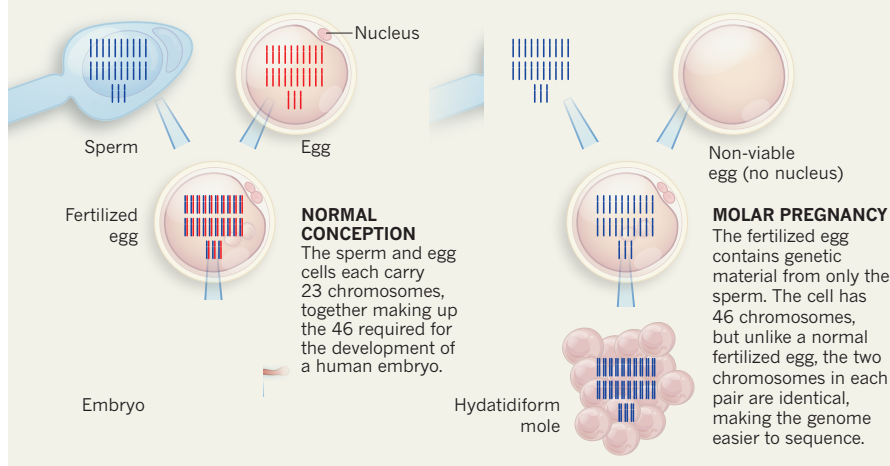
Sequencing machines cannot process entire chromosomes, so scientists must first make many identical copies of the DNA and cut them into short stretches, with the breaks in different places. After sequencing, a computer program looks for overlapping patterns to 'stitch' the resulting segments back together.

This approach worked for most of the genome, because DNA sequences are almost identical across its three billion 'letters' (the As, Cs, Ts and Gs). But in some parts, big differences exist between the versions of chromosomes that an individual inherits from the mother and father. Attempts to stitch together these regions to sequence the DNA led to gaps when the differing sequences gave conflicting solutions.

The problem can be likened to assembling a single jigsaw puzzle from the mixed-up pieces of similar, but not identical, puzzles. If one puzzle piece is identical across the sets, any copy of it will do. But if one set contains a much larger version of the matching piece, or if a piece is missing, the puzzle will not fit together. In particular, long, repetitive stretches near genes vexed the computer algorithms used to analyse the data. And the problem was made worse because DNA

TO SIMPLIFY A SEQUENCE

To produce uninterrupted DNA sequences of human chromosomes, geneticists are turning to hydatidiform moles. These are formed when a sperm cell enters an egg that has lost its nucleus, making it non-viable.



from multiple people was used, adding to the variation between the genomes.

As a result, when a person's genome is sequenced — for instance, to look for the cause of a disease — crucial bits of DNA may be overlooked because they do not have counterparts in the published genome. "There's a whole level of genetic variation that we're missing," says Evan Eichler, a genome scientist at the University of Washington in Seattle, a leading proponent of the platinum-genome efforts. To plug the gaps, researchers need a supply of human cells with just a single version of each chromosome, to remove the possibility of conflicting solutions — a single set of puzzle pieces, in other words.

Sperm and egg cells contain a single copy of each chromosome, but these cells cannot divide and produce copies of themselves. So in recent years, geneticists have turned to cells from growths called hydatidiform moles, created when a sperm fertilizes an egg that is missing its own genetic material (see "To simplify a sequence"). The fertilized cell copies its genome and starts dividing, just as the cells in a normal fertilized egg would. The resulting ball of cells, which is usually removed in the first trimester of pregnancy, contains identical copies of each human chromosome.

Cells taken from one such mole were used in the early 1990s to create a cell line called CHM1. In a *Nature* paper published on 10 November, Eichler and his colleagues describe how they

used sections of the CHM1 genome to fill about 50 especially troublesome holes in the official human genome sequence. They also shortened many more gaps, including in genes linked to ALS and Fragile X syndrome, a neurodevelopmental disease with autism-like symptoms (M. J. P. Chaisson *et al. Nature* <http://doi.org/w69>; 2014). In total, the team mapped around 1 million DNA letters that were missing in the original reference genome.

A true platinum sequence will be assembled from just one genome, however, because only then can scientists be sure there are no remaining gaps. To this end, a team led by Richard Wilson at Washington University in St. Louis, Missouri, reported a draft sequence of the entire CHM1 genome earlier this month (K. M. Steinberg *et al. Genome Res.* <http://doi.org/w7b>; 2014). Researchers at the firm Pacific Biosciences in Menlo Park, California, are similarly working on the whole CHM1 genome, but are using sequencers that work with longer stretches of uninterrupted DNA, and so produce fewer gaps than typical sequencers. The firm released a draft genome assembly in February. The hope is that the method will speed up the platinum genome's arrival.

"The chances of actually achieving this, for one genome, are looking much better," says Deanna Church, a genome scientist at the firm Personalis in Menlo Park. Still, Birney says that the human reference genome is more about "constant improvement" than completion. "For sure, somebody's going to be fiddling around with this in 10–20 years' time." ■