

EXPLAINED

Relevant for: Science & Technology | Topic: Science and Technology- developments and their applications and effects in everyday life

The story so far: On May 27, a preprint titled [The complete sequence of the human genome](#) was posted in the online repository *bioRxiv*. In this preprint, scientists from the Telomere-to-Telomere (T2T) Consortium, an international collaboration of around 30 institutions, reported the most complete sequencing of the human genome until now. They have, in the process, discovered over a hundred new genes that code for proteins. The total size of the genome they have sequenced is close to 3.05 billion base pairs. This adds 200 million base pairs to the last draft of the human genome that was published in 2013. The results come with the caveat that about 0.3% may still have errors, and that among the sex chromosomes, only the X chromosome has been sequenced.

The [Human Genome Project](#) that began in 1990 gave the first results of the complete human genome sequence in 2003. For the first time, we were able to read the blueprint of human life. However, though it was announced as the complete human genome, about 15% of it was incomplete. Due to limitations of technology, scientists were not able to piece together some repetitive parts of the human genome.

Plumbing the 'dark' genome for new genes

Solving some of the problems, an [updated "complete" version was released in 2013](#), which still missed out on 8% of the genome. Now, the [researchers have nearly completed the job](#), adding 200 million base pairs and 115 new protein-coding genes to the list.

The human genome is the entire set of deoxyribonucleic acid (DNA) belonging to a human. This resides in the nucleus of every cell of the human being. The DNA consists of a double-stranded molecule, each of which is built up by four bases – adenine (A), cytosine (C), guanine (G) and thymine (T). Every base on one strand pairs with a complementary base on the other strand (A pairs only with T, and C only with G). In all, the genome is made up of 3.05 billion such base pairs, approximately.

Of these, there are long stretches that do not seem to have a particular function. On the other hand, protein-coding sequences or protein-coding genes are DNA sequences that get transcribed on ribonucleic acid (RNA) as an intermediate step. These in turn make the proteins responsible for various functions such as keeping the body healthy or determining the colour of the eye — proteins carry out the instructions encoded in the genes.

The DNA used did not belong to any person. According to a report in *Nature*, it was a cell line derived from a tissue known as a complete hydatidiform mole. This is the tissue that forms when a sperm inseminates an egg that has no nucleus. Hence, this tissue has the chromosomes of just the father.

For one thing, it has no information about the Y chromosome.

We know that all chromosomes in an arbitrary cell's nucleus are found in pairs – we have 23 pairs of chromosomes in each cell. However, the sex cells such as sperm and egg cells contain only one of each pair of chromosomes (haploid cells). So, while egg cells always carry a copy of the X chromosome, sperms can carry either an X chromosome or a Y chromosome. The cell line that the researchers studied had an X chromosome only and no Y chromosome. Therefore,

information about the Y chromosome is missing in this release.

It is also not 100% complete. The researchers say that about 0.3% of the genome may have errors.

One of the most important uses of this release will be that it forms a standard for comparison in future sequencing attempts, according to Dr. Satyajit Rath, a visiting faculty member at Indian Institutes of Science Education and Research (IISER), Pune, and an expert on immunology. Just as the standard of time is given by the beats of a caesium clock placed in the International Bureau of Standards, this sequence of the human genome will be a gold standard of reference for future attempts.

The level of accuracy is unprecedented and while earlier, people were trying to piece together strands of DNA that were a few hundred base pairs long, the technology used by the Telomere-to-Telomere Consortium used sequencing technology that could scan 20,000 base pairs at one go. This is a significant technological feat.

[Our code of editorial values](#)

This study was completed before the emergence of the Delta variant of SARS-CoV-2 now dominating in the U.K.

END

Downloaded from crackIAS.com

© **Zuccess App** by crackIAS.com

Crackin