Publication at a glance

THE COMPLETE SEQUENCE OF A HUMAN GENOME

Why are these papers important?

This landmark study by members of the *Telomere-to-Telomere (T2T) Consortium* is the first fully complete assembly to be produced, 20 years after the initial drafts of the human genome.

The work, as documented in nearly 20 peer-reviewed papers[1], is changing the game in human genetics, "enabling comprehensive studies of genomic variation across the entire human genome, which we expect to drive future discovery in human genomic health and disease.[2,3,4]" Now, researchers have access to millions of new, unexplored bases in the human genome to expand studies of disease association and cellular function.

The tools developed to support the *T2T Consortium* using PacBio HiFi data:

- **Enable** all human genetics researchers to generate reference-quality genomes from diverse backgrounds to better understand human health
- **Remove** barriers of legacy technologies that only provide fractions of the genome
- **Allow** researchers to avoid misidentifying disease-linked mutations by expanding the catalog of non-pathogenic, person-to-person variation

And by comparing the complete human genome to the genomes of our closest primate relatives, researchers can now better understand the differences, i.e., ultimately what makes us "human," even for regions of the genome that previously were inaccessible.

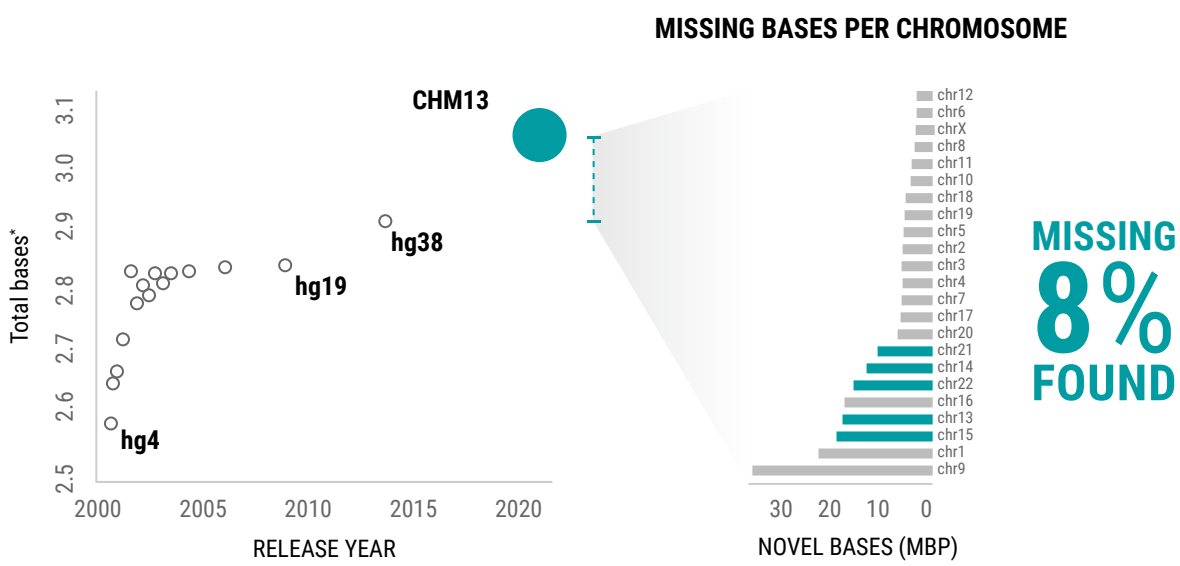
Metric	GRCh38p13	CHM13v1.1	±%
Assembled bases (Gb)	2.92	3.05	+5%
Unplaced bases (Mb)	11.42	0	−100%
Gap bases (Mb)	120.31	0	−100%
# Contigs	949	24	−98%
Ctg NG50 (Mb)	56.41	154.26	+174%
# Issues	230	46	−80%
Issues (Mb)	230.43	8.18	−97%

Key findings

Based on a HiFi backbone, the T2T-CHM13 assembly provides:

- A more complete, contiguous, and correct reference than GRCh38
- 200 million bp of novel sequence
- Gapless assemblies of 22 autosomes, X chromosome, and mitochondrial genome
- 2,226 new paralogous gene copies, including 115 predicted to be protein coding

PacBio



Total non-gap bases in UCSC reference genome releases dating back to September 2000 (hg4) and ending with T2T-CHM13 in 2021

*Excluding mtgenome, chromosome Y, and Ns in previous assemblies

New bases in the CHM13 assembly relative to GRCh38 per chromosome, with the acrocentrics highlighted in blue

"[The previously missing portion] of the genome has not been overlooked because of a lack of importance but rather because of technological limitations. High-accuracy long-read sequencing has finally removed this technological barrier."

Nurk et al. (2022) The complete sequence of a human genome. *Science* Vol 376, Issue 6588, pp. 44-53

What role did PacBio technology play?

The authors utilized multiple sequencing techniques and chose PacBio HiFi reads for building the genome assembly backbone. Their combination of long read lengths (up to 25 kb) and high accuracy (99.9%), enabled the assembly of highly repetitive centromeric satellite arrays and closely related segmental duplications.

What's next?

HiFi sequencing is the essential technology for delivering a gold standard reference genome, generating genome diversity, and opening a world of possibilities. Using PacBio whole genome sequencing for *de novo* assembly application, a 6 Gb diploid genome can be sequenced using three SMRT® Cells targeting 30-fold coverage and can be assembled in less than one day. In addition to the genome assembly, HiFi sequencing will also provide genome-wide detection of 5-methylcytosine (5mC) methylation.

KEY REFERENCES

1. **Completing the human genome.** *Science* Vol 376, Issue 6588
2. Nurk et al. (2022) **The complete sequence of a human genome.** *Science* Vol 376, Issue 6588, pp. 44-53
3. Aganezov et al. (2022) **A complete reference genome improves analysis of human genetic variation.** *Science*, 376:6588
4. Vollger MR, et al. (2022) **Segmental duplications and their variation in a complete human genome.** *Science*, 376:6588

Learn about whole genome sequencing for *de novo* assembly: **pacb.com/wgs-human**

Read more about a **new era of genomics**

Explore the **infographic**

102-193-686 REV02 09JUN2022