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, 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." 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.

### 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

### Publication at a glance

**THE COMPLETE SEQUENCE OF A HUMAN GENOME**

<table>
<thead>
<tr>
<th>Metric</th>
<th>GRCh38p13</th>
<th>CHM13v1.1</th>
<th>±%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assembled bases (Gb)</td>
<td>2.92</td>
<td>3.05</td>
<td>+5%</td>
</tr>
<tr>
<td>Unplaced bases (Mb)</td>
<td>11.42</td>
<td>0</td>
<td>-100%</td>
</tr>
<tr>
<td>Gap bases (Mb)</td>
<td>120.31</td>
<td>0</td>
<td>-100%</td>
</tr>
<tr>
<td># Contigs</td>
<td>949</td>
<td>24</td>
<td>-98%</td>
</tr>
<tr>
<td>Ctg NG50 (Mb)</td>
<td>56.41</td>
<td>154.26</td>
<td>+174%</td>
</tr>
<tr>
<td># Issues</td>
<td>230</td>
<td>46</td>
<td>-80%</td>
</tr>
<tr>
<td>Issues (Mb)</td>
<td>230.43</td>
<td>8.18</td>
<td>-97%</td>
</tr>
</tbody>
</table>
“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.”


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

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

Read more about a new era of genomics

Explore the infographic

Research use only. Not for use in diagnostic procedures. © 2022 Pacific Biosciences of California, Inc. ("PacBio"). All rights reserved. Information in this document is subject to change without notice. PacBio assumes no responsibility for any errors or omissions in this document. Certain notices, terms, conditions and/or use restrictions may pertain to your use of PacBio products and/or third-party products. Refer to the applicable PacBio terms and conditions of sale and to the applicable license terms at pacb.com/license. Pacific Biosciences, the PacBio logo, PacBio, Circulomics, Omniome, SMRT, SMRTbell, Iso-Seq, Sequel, Nanobind, and SBB are trademarks of PacBio. All other trademarks are the sole property of their respective owners.