

GENOMICS

The Backstory of the First Complete Sequence of the Human Genome

A Q&A with BBI researchers reveals why the merger of two sequencing systems was crucial to landmark study

By: BBI Communications on April 29, 2022



Left to right: Drs. Evan Eichler, Glennis Logsdon, and Danny Miller

By BBI Communications

[The following is an edited interview with BBI's Drs. Evan Eichler and Danny Miller, as well as postdoctoral fellow Dr. Glennis Logsdon, on the first complete sequence of the human genome, announced in late

March in a [series of papers published in Science.](#)]

BBI: What is the backstory behind this breakthrough paper of the completion of sequencing the human genome?

Eichler: When we were finishing the human genome back in 2003, we realized a lot of the errors had to do with trying to sort out different complex regions. We had inadvertently combined the mother's and the father's structures of that part of the genome that were incompatible with each other. It created a lot of gaps in repetitive regions, but especially in the duplicated regions that contained genes.

In 2003, I submitted a white paper to the National Institutes of Health to start building genomic resources from a human source that only contained paternal information (i.e. a complete hydatidiform mole). That led to research over several years resolving many of those gaps of the original human genome.

Fast forward 15 years. Long-read sequencing had been used for several years and we started sequencing that same source in an effort to completely resolve the complex regions and gaps. Separately, Adam Phillippy at the National Institutes of Health and Karen Miga at the University of California, Santa Cruz also had started sequencing it using a different long read sequencing technology. Adam called me early in 2018 and said, "Wait a minute, we're competing against each other. Let's come together and work on this."

So, the Telomere-to-telomere (T2T) consortium that ultimately completed the human genome was led by Adam, Karen, and me – and, of course, the hard work of many colleagues and associates from our labs and others' labs.

Logsdon: When I joined the Eichler lab in 2018, I had an interest in sequencing some of the most difficult regions of the genome. I proposed to use long-read sequencing to completely sequence the centromeres, which are constricted regions of each chromosome essential for inheritance and genome stability. At the time, it had not been done.

There were two main modes of sequencing, and our two groups – Evan's lab and the partnership between Drs. Phillippy and Miga – were using different modes. We were using primarily HiFi sequencing from

Pacific Biosciences (PacBio), and they were using primarily Oxford Nanopore Technologies (ONT).

Miller: My role was focused on the clinical utility of this study, of the completion of the human genome. How do we use this to identify disease-causing variants that may have been missed because they were in regions not fully resolved in the existing version of the genome? I joined the effort in 2019 with about 50 other scientists.

Eichler: We were early adopters of PacBio HiFi sequencing. Adam and Karen were pretty committed to using ONT in terms of driving the project. There was quite a bit of back-and-forth in the beginning on which was the best technology for the complex regions, but Adam, Karen, and I quickly came around to the idea that two were better than one. In fact, the success, the completion of this project, depended on merging both technologies to complete the whole genome. Merging of the two technologies was key.

Logsdon: The project really took off in the beginning of June of 2020. Before then, we (people in the three labs) had talked over Zoom calls. We worked regularly that summer, primarily through Slack channels, and had an initial draft of the human genome in September.

Miller: For that initial draft, my contribution was to help guide which regions should be focused on to improve the genome assembly. Subsequently, with the variants paper (a related paper published in Science) it was identifying and discussing variants in clinically relevant regions and whether the genome we assembled had any rare disease-causing variants of its own that could be misleading when using it as a reference genome.

BBI: With this milestone achievement, what's the future of the human genome for clinical applications?

Miller: As we showed in the variants paper, using this improved assembly will likely lead to improved outcomes if used for clinical testing. This is because you can resolve more medically relevant genes and identify disease-causing variants you may have missed when using other reference genomes. The challenging question is: When will clinical labs adopt this new assembly or a similar assembly? That will take time. It's hard to say with clinical labs. They require a careful

evaluation process if they work on an older version of the reference – and there's a big cost to switching. I think it's five to 10 years before they are using the T2T assembly.

Eichler: That timing is not too far off because of all the issues – related to clinical testing and CLIA certification that need to happen. There is disease-causing variation that we are seeing for the first time. What's really important is that it is a fundamental change of how variants are discovered. Both at the research level and in the clinic. In fact, most of the long-read stuff to this day that people are using in the clinic is still based on alignment to a reference, whether that reference is T2T or one of the older references.

That is not the right way forward. The reason we're not solving more genetic cases is that we're missing genetic variation that's important to disease because it is just too complex to map to a reference. Ultimately, I think we're going to get individuals coming into a clinic and we're going to generate a T2T assembly first for each patient's genome. So, you basically have the mother's and the father's complement of that child laid out in front of you and you're going to discover variations second.

Right now, people are taking sequence reads and aligning them and trying to interpret variants from a reference. I think every patient is going to be their own reference genome. They will have a T2T assembly, so we should live so long that this will be part of our medical records. At the present time, in principle, this can be generated, it's just that it's too expensive and too time-consuming. Think about this for cancer. If you could completely sequence and assemble cancer genomes – or the pre-cancerous genomes – you could, in fact, identify missed disease-causing mutations and develop therapies much earlier on. That may be further away, but this is all going to happen. It needs to happen if human genetics is going to continue to improve human health.

Share: 

CATEGORIES

BBI Admin & Development

BBI Grants

BBI Members

Cell Free DNA

Genomics

Infectious Diseases

Living Biobank

Mutational Scanning

Pediatric Sequencing

Promising Young Scientists

Single Cell Sequencing

Therapeutics

Related



BBI's Danny Miller Leading New Effort Offering Sequencing Services

[April 29, 2022](#) | [By BBI Communications](#)

BBI's Dr. Danny Miller and the UW's Nanopore Sequencing Core are providing both targeted and whole genome long read sequencing on the Oxford Nanopore platform to assist other researchers with their projects.

[Read More](#)

ABOUT

RESEARCH

RECENT TWEETS

Follow us

RESOURCES

NEWS

EVENTS

CONTACT

CONTACT

206-543-9660

info@brotmanbaty.org

Box 357657 | Seattle, WA

98195-8047

Health Sciences Building H-564