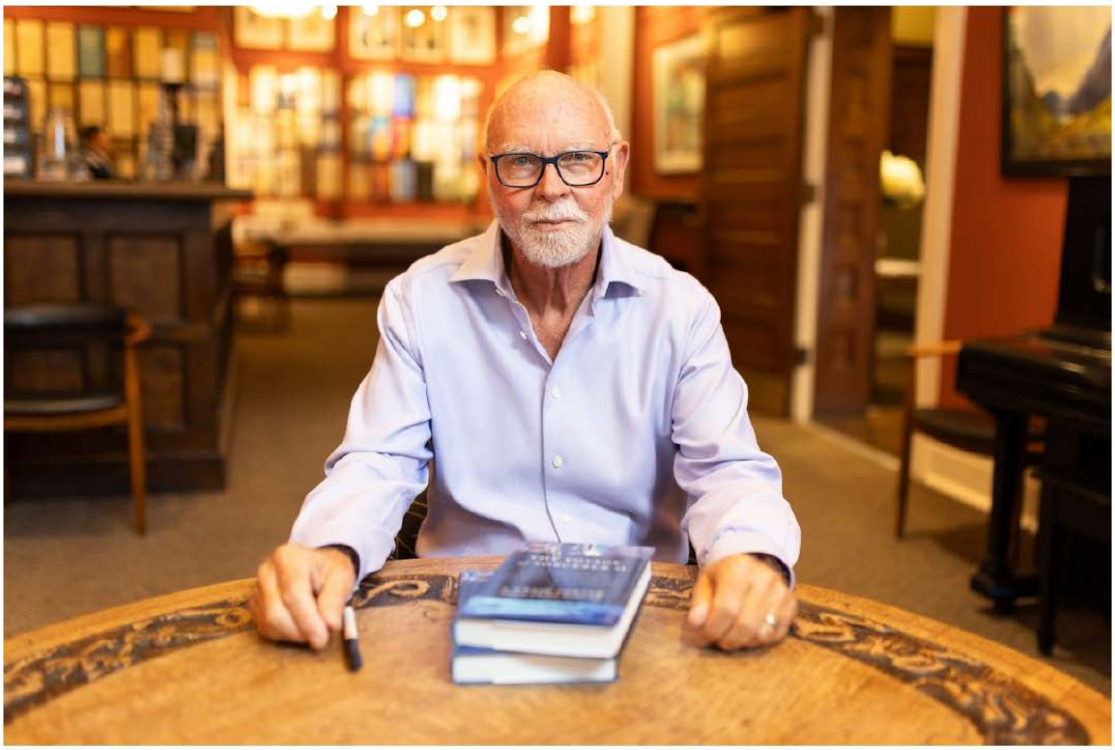

Scientist and author Craig Venter signs copies of his book at the Arts and Letters club in Toronto, in October, 2023.

MELISSA TAIT/THE GLOBE AND MAIL

Genomics pioneer's DNA used in study

Craig Venter donated blood at Toronto's SickKids hospital as his final contribution to scientific research

IVAN SEMENIUK
SCIENCE REPORTER

Two and half years ago, Craig Venter, the scientist-entrepreneur who is best known for accelerating the race to sequence the human genome, donated blood during a visit to Toronto's SickKids hospital in hopes of sparking another leap forward.

Now, Dr. Venter, who died on April 29, has got his wish. He is a posthumous co-author on what is likely to be his final contribution to science: a technique for sequencing the most complete version of a human genome possible for a reasonable cost.

In a study posted online last week and submitted for peer review, researchers at the hospital's Centre for Applied Genomics revealed that they used DNA from blood donated by Dr. Venter for the newly developed method.

Dr. Venter was also involved in preparing the study for publication until shortly before his death at age at 79 from the side effects of cancer treatment.

"He was quite adamant in his last few days that this paper be submitted," said the centre's director, Stephen Scherer.

Known as a telomere-to-telomere assembly (after the structures at the tips of each chromosome) the genome released by SickKids is a near-complete reading of a single individual's DNA. The goal of the effort was to minimize the gaps that can occur in the sequencing process. Such gaps can hide genetic variations between individuals that may be linked to disease risk.

The more of the genome that is sequenced, the higher the likelihood that a potentially dangerous variant will be detected.

"The goal of the field right now is to see how far you can push the clinical detection rate limits," Dr. Scherer said.

Genetic information is encoded in the molecular structure of DNA, which is built like a long twisting ladder in which each rung, called a base pair, represents a single bit of information. Collectively, those bits contain instructions for building proteins and other biological functions.

An entire human genome consists of about three billion base pairs spread across 23 chromosomes. When the Human Genome Project was launched by the U.S. National Institutes of Health in 1990, scientists anticipated it would take 15 years to read the sequence of a single representative individual. (In practice, DNA from several donors was read to stitch the genome together.)

Dr. Venter pioneered his own approach, called "shotgun sequencing," which famously sped up the process and made it possible for his company, Celera Genomics, to give the government-funded project a run for its money. The parallel efforts were declared completed in 2003. By then, about 92 per cent of the human genome had been sequenced.

Dr. Venter was also involved in preparing the study for publication until shortly before his death at age at 79 from the side effects of cancer treatment.

Since then, it has been a challenge to fill in the remaining blanks in the DNA sequence. This is because the sequence is built up from smaller overlapping fragments of DNA. The smaller the fragments, the harder they are to assemble. Areas of the genome where the DNA encodes a repeating pattern are especially tricky.

The situation is analogous to assembling the works of Shakespeare from short groups of sentences. If the groups become larger and include entire pages of text, the assembly gets easier.

In 2022, another U.S.-led project called the Telomere-to-Telomere Consortium published its version of a gapless human genome.

That effort included 90 authors and the genome was assembled using multiple methods at a cost far higher than conventional genome sequencing for clinical purposes. Meanwhile, the team at SickKids has been working on the same problem with a different goal in mind: Building a gapless genome for a cost low enough to make it practical for an extended family or some other sample population to be screened for signs of an inherited disorder.

"Our approach was to make something that's scalable so that we can do hundreds of these things," said Si Lok, a genomicist and technology development lead at the SickKids centre.

Dr. Venter's DNA was ideal for their purpose because it had already been sequenced by Celera in the mid-2000s. This provided a reference for the SickKids team to measure their improvement.

Dr. Venter, who was a member of the research centre's advisory board a the time of his death, was only too happy to support the effort, Dr. Scherer said. One reason may have been the chance to show that a small, focused team could produce something more efficiently than a large group – an echo of his battle with the Human Genome Project.

The result of the SickKids' effort was a genome of 3,077,506,360 base pairs in length, far exceeding in completeness Dr. Venter's original reference genome created nearly 20 years earlier.

"I'm sure Craig was happy about that," said Steven Salzberg, a computational biologist at Johns Hopkins University and a member of the Telomere-to-Teselomere Consortium.

Dr. Salzberg said he was interested to see the SickKids' result but noted that a genome recently mapped by the consortium went further by including the fully sequenced chromosomes inherited by both parents of an individual.

Among the more interesting details in Dr. Venter's genome is a variant that has been associated with certain types of cancer, an association that only became apparent as more genomes were available for comparison. The new result uncovered thousands of additional variants not seen in the initial reference.

"I believe the take away is what Craig was pushing all along," said Dr. Scherer. "We need the genomes of entire populations, including those people with diseases," to fully realize the predictive power of the human genome.