

## Thousands of never-before-seen human genome variations uncovered

November 10 2014



Dr. Evan Eichler and his colleagues are exploring the use of new technologies for uncovering genetic variations in humans and for closing gaps in the mapping of the human genome. Eichler is shown in his office at the University of Washington in Seattle. Credit: Clare McLean

Thousands of never-before-seen genetic variants in the human genome have been uncovered using a new genome sequencing technology. These



discoveries close many human genome mapping gaps that have long resisted sequencing.

The technique, called single-molecule, real-time DNA sequencing (SMRT), may now make it possible for researchers to identify potential genetic mutations behind many conditions whose genetic causes have long eluded scientists, said Evan Eichler, professor of genome sciences at the University of Washington, who led the team that conducted the study.

"We now have access to a whole new realm of <u>genetic variation</u> that was opaque to us before," Eichler said.

Eichler and his colleague report their findings Nov. 10 in the journal *Nature*.

To date, scientists have been able to identify the genetic causes of only about half of inherited conditions. This puzzle has been called the "missing heritability problem." One reason for this problem may be that standard <u>genome sequencing</u> technologies cannot map many parts of the genome precisely. These approaches map genomes by aligning hundreds of millions of small, overlapping snippets of DNA, typically about 100 bases long, and then analyzing their DNA sequences to construct a map of the genome.

This approach has successfully pinpointed millions of small variations in the <u>human genome</u>. These variations arise from substitution of a single nucleotide base, called a single-nucleotide polymorphisms or SNP. The standard approach also made it possible to identify very large variations, typically involving segments of DNA that are 5,000 bases long or longer. But for technical reasons, scientists had previously not been able to reliably detect variations whose lengths are in between—those ranging from about 50 to 5,000 bases in length.



The SMRT technology used in the new study makes it possible to sequence and read DNA segments longer than 5,000 bases, far longer than standard gene sequencing technology.

This "long-read" technique, developed by Pacific Biosciences of California, Inc. of Menlo Park, Calif., allowed the researchers to create a much higher resolution structural variation map of the genome than has previously been achieved. Mark Chaisson, a postdoctoral fellow in Eichler's lab and lead author on the study, developed the method that made it possible to detect structural variants at the base pair resolution using this data.

To simplify their analysis, the researchers used the genome from a hydatidiform mole, an abnormal growth caused when a sperm fertilizes an egg that lacks the DNA from the mother. The fact that mole genome contains only one copy of each gene, instead of the two copies that exist in a normal cell. simplifies the search for genetic variation.

Using the new approach in the hydatidiform genome, the researchers were able to identify and sequence 26,079 segments that were different from a standard human reference genome used in genome research. Most of these variants, about 22,000, have never been reported before, Eichler said.

"These findings suggest that there is a lot of variation we are missing," he said.

The technique also allowed Eichler and his colleagues to map some of the more than 160 segments of the genome, called euchromatic gaps, that have defied previous sequencing attempts. Their efforts closed 50 of the gaps and narrowed 40 others.

The gaps include some important sequences, Eichler said, including parts



of genes and regulatory elements that help control gene expression. Some of the DNA segments within the gaps show signatures that are known to be toxic to Escherichia coli, the bacteria that is commonly used in some genome sequencing processes.

Eichler said, "It is likely that if a sequence of this DNA were put into an E. coli, the bacteria would delete the DNA." This may explain why it could not be sequenced using standard approaches. He added that the gaps also carry complex sequences that are not well reproduced by standard <u>sequencing technologies</u>.

"The sequences vary extensively between people and are likely hotspots of genetic instability," he explained.

For now, SMRT technology will remain a research tool because of its high cost, about \$100,000 per <u>genome</u>.

Eichler predicted, "In five years there might be a long-read sequence technology that will allow clinical laboratories to sequence a patient's chromosomes from tip to tip and say, 'Yes, you have about three to four million SNPs and insertions deletions but you also have approximately 30,000-40,000 structural variants. Of these, a few structural variants and a few SNPs are the reason why you're susceptible to this disease.' Knowing all the variation is going to be a game changer."

**More information:** "Resolving the complexity of the human genome using single-molecule sequencing." <u>DOI: 10.1038/nature13907</u>

Provided by University of Washington

Citation: Thousands of never-before-seen human genome variations uncovered (2014, November



10) retrieved 28 April 2024 from <u>https://medicalxpress.com/news/2014-11-thousands-never-before-seen-human-genome-variations.html</u>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.