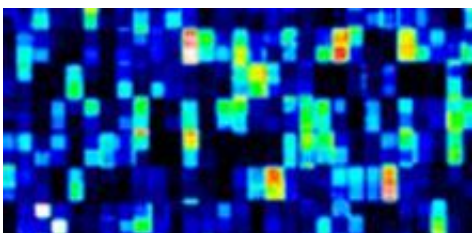


Scientists develop more accurate whole genome variant discovery and interpretation

June 29 2015



Scientists from the Icahn School of Medicine at Mount Sinai have developed a new approach to build nearly complete genomes by combining high-throughput DNA sequencing with genome mapping. The methodology enabled researchers to detect complex forms of genomic variation, critically important for their association with human disease, but previously difficult to detect. The study was published today in *Nature Methods*, and is a collaboration with scientists at European Molecular Biology Lab, Weill Cornell Medical College, Cold Spring Harbor Laboratory, Rockefeller University, University of California, San Francisco, Pacific Biosciences, and BioNano Genomics.

Conventional next-generation sequencing (NGS) techniques are able to accurately detect certain types of variation, such as single nucleotide variants and small insertions or deletions, but miss many large or complex forms of genomic variation that are associated with [human](#)

[disease](#). Further, these previous approaches are poorly suited for completely de novo analysis of genomes and for phasing the maternal and paternal haplotypes of an individual.

"We created a high-throughput strategy that builds highly contiguous de novo genomes without the need for complex jumping libraries or targeted approaches. This strategy, in some cases, automatically resolved complete arms of chromosomes," said Ali Bashir, PhD, Assistant Professor of Genetics and Genomics at the Icahn School of Medicine and senior author of the study. "While we focused this study on a human [genome](#), the method can be applied to any new genome, including those with high genomic complexity, such as plants, that have been extremely challenging to study."

To overcome limitations with existing NGS methods, the study authors combined two single molecule approaches: long read sequencing from Pacific Biosciences and Nanochannel Array technology from BioNano Genomics. Pacific Biosciences sequencing enables reads exceeding 10kb in length, which can directly resolve and phase complex forms of variation. The NanoChannel Array from BioNano confines and linearizes DNA molecules up to megabases in length to provide high-resolution sequence motif physical maps, termed 'genome maps'.

The researchers studied the NA12878 diploid genome, a well-sequenced sample that is part of the 1000 Genomes project and often used for benchmarking new techniques. The study authors mapped variation and built assemblies with both technologies, then combined the two to create a "hybrid" assembly that dramatically improved the contiguity of each. The resulting hybrid assembly N50s, the length such that 50% of all base pairs are contained in scaffolds of the given length or longer, approach 30Mb - on par with the best assemblies to date at a fraction of the cost and labor.

"The study revealed an unprecedented view of genomic complexity, in many cases identifying regions overlooked by conventional sequencing or further refining previously known genetic variant classes," said study co-author Jan Korbelt, PhD, Group Leader at the European Molecular Biology Laboratory. "We had notable success in challenging regions such as inversions and tandem repeats," added co-author Robert Sebra, PhD, Assistant Professor of Genetics and Genomic Sciences at the Icahn School of Medicine. "For example, a systematic underrepresentation of tandem repeat sizes was observed in the human reference genomes. Such expansions, as we observed within the LPA gene which has been associated with plasma lipid levels, are increasingly being identified as important markers for disease."

"By using a powerful combination of new technologies, we can finally begin to circumvent biases induced by overreliance on a single reference genome" said co-author Eric Schadt, PhD, Founding Director of the Icahn Institute, and Professor of Genomics at the Icahn School of Medicine. "Fully de novo approaches will increasingly become standard practice to enable direct and comprehensive characterization of genome variation. This will accelerate our understanding of the links to human diseases that such variations induce."

More information: Matthew Pendleton, Robert Sebra, Andy Wing Chun Pang, Ajay Ummat, Oscar Franzen, Tobias Rausch, Adrian M Stütz, William Stedman, Thomas Anantharaman, Alex Hastie, Heng Dai, Markus Hsi-Yang Fritz, Han Cao, Ariella Cohain, Gintaras Deikus, Russell E Durrett, Scott C Blanchard, Roger Altman, Chen-Shan Chin, Yan Guo, Ellen E Paxinos, Jan O Korbelt, Robert B Darnell, W Richard McCombie, Pui-Yan Kwok, Christopher E Mason, Eric E Schadt & Ali Bashir. "Assembly and Diploid Architecture of an Individual Human Genome via Single Molecule Technologies." *Nature Methods*. [DOI: 10.1038/nmeth.3454](https://doi.org/10.1038/nmeth.3454)

Provided by The Mount Sinai Hospital

Citation: Scientists develop more accurate whole genome variant discovery and interpretation (2015, June 29) retrieved 15 May 2024 from <https://phys.org/news/2015-06-scientists-accurate-genome-variant-discovery.html>

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.