

# High-throughput, sensitive approach helps reveal what's 'real' in genome-wide association data

June 2 2016

---



From left to right, Satich Nandakumar, Jacob Ulirsch, Vijay Sankaran, Pardis Sabeti and Ryan Tewhey. Credit: Sabeti Lab, Broad Institute

Tackling one of the key challenges facing current human genetics, a pair

of multi-institutional research teams have demonstrated a tool that should help untangle which genetic variants actually create risk for heart disease, diabetes and a host of other diseases.

The research—conducted by scientists from Broad Institute of MIT and Harvard, Harvard University, and Dana-Farber/Boston Children's Cancer and Blood Disorders Center, and unveiled in two papers published June 2 in *Cell*—employs an experimental technique called the "massively parallel reporter assay." The technique lets researchers probe thousands of DNA variations to identify ones that affect gene regulation—how genes are turned on and turned off.

The problem that geneticists face with disease-causing variants is an overabundance of candidates. Over the last decade, researchers worldwide have identified numerous stretches of human DNA associated with risk for a wide range of diseases as well as with other important physical traits, in an approach known as genome-wide association studies (GWAS). The catch is that each region can harbor hundreds of genetic variants, and it is very hard to tell which one actually makes people more likely to get sick.

"With GWAS, you get a set of signals, which can tell you which regions of the genome are associated with a particular disease or trait," said Vijay Sankaran, a Broad Institute associate member, a pediatric hematologist/oncologist at Dana-Farber/Boston Children's, and senior author on one of the two *Cell* papers. "But it is hard to know which hits are causal hits, and which are just going along for the ride."

The picture gets particularly complicated when talking about variants in non-coding DNA, including the vast stretches of DNA containing sequences that control gene expression. By some estimates, between 85 and 90 percent of the variants picked up by GWAS lie in such regions. Thus, scientists are seeking ways to connect the dots between non-coding

GWAS variants, human biology, and, ultimately, human illness.

"We want to move from understanding the component pieces of the genome to understanding what changes in those components do," said Pardis Sabeti, an institute member of the Broad Institute, Harvard computational geneticist and evolutionary biologist, and senior author on the second study, whose lab probes the role genetic variation, writ large, plays in human and microbial evolution. "We need very sensitive technology to be able to identify these functional changes, particularly if they're subtle."

## Going massive

Reporter assays, a staple of the genomics toolkit for decades, help scientists sift through GWAS data to find variants that truly affect gene expression or function. A researcher takes a DNA fragment from what may be an enhancer, couples it within a plasmid to a "reporter" gene that provides a readout (e.g., the luciferase gene), and inserts the plasmid into cells. If the readout materializes (e.g., if the cells glow), the enhancer sequence drove expression of the reporter. By running the assay with different variations of the same fragment, a pattern can emerge suggesting whether certain variants affect expression.

Such assays, however, have one major disadvantage: They don't scale to the level needed to investigate the thousands to tens of thousands of variants that might turn up in a GWAS.

Broad alum Tarjei Mikkelsen (now with the biotechnology company 10X Genomics), and Broad research scientist Alexandre Melnikov worked out the principles of one flavor of MPRA while working in the lab of Broad founding director and president Eric Lander. In a 2012 *Nature Biotechnology* paper, they noted that tagging each plasmid with a short, unique DNA barcode provided a second readout. By sequencing

and counting the mRNAs produced from each plasmid, they could easily identify the variant(s) with the greatest influence on gene expression and quantify the magnitude of that influence.

And because each barcode was unique to each plasmid, Mikkelsen and Melnikov's team could pool and assay thousands of variants simultaneously.

## **Homing in on blood cell traits**

Sankaran's lab used Mikkelsen and Melnikov's MPRA system to scrutinize more than 2,750 non-coding variants in 75 GWAS hits linked to red blood cell traits. And as he, Mikkelsen, and co-first authors Jacob Ulirsch and Satish Nandakumar reported in their *Cell* paper, MPRA data uncovered 32 hits that actually impacted on [gene expression](#). Using additional computational and functional assays to further probe the effects of a subset of these variants on [red blood cell](#) traits, the team found that several known genes may have heretofore-unrecognized roles in blood cell development.

"One of the unexpected lessons we learned was that many of the variants tweaked a master blood development regulator, GATA1," said Ulirsch, a staff scientist in Sankaran's lab. "There was a common pattern. Going one by one, variant by variant, we would never have been able to see this."

## **Building MPRA 2.0**

While Mikkelsen and Melnikov's original method is quite powerful, Sabeti's lab wanted to see if they could make it even more robust.

"The original version of MPRA is limited in how many variants you can

test," said Ryan Tewhey, a postdoctoral fellow in Sabeti's lab and first author on the second *Cell* paper. "We wanted to know, can you expand this technology out? Can you test tens of thousands of variants at once? And can you make it more sensitive?"

Tewhey, Sabeti, and their team doubled the length of each DNA barcode and upped the number of barcodes to as many as 350 per variant. They then used their enhanced assay to study more than 32,000 possible B cell regulatory variants identified by the 1000 Genomes Project, deeply characterizing one associated with risk of ankylosing spondylitis (an autoimmune disease). They also highlighted another 842 candidate variants, including 53 particularly promising ones associated with human traits and diseases.

As they discussed in their *Cell* paper, the added barcodes reduced the noise in their data and increased the assay's overall sensitivity.

"With more barcodes you can start to detect more subtle changes in expression, including changes that might arise from differences between alleles," Tewhey added.

## **Another view into regulation**

MPRA isn't the only approach for pulling causal needles out of GWAS haystacks, and Tewhey is realistic that it won't be a panacea for studying all of the cell's mechanisms for regulating expression.

"For promoters and enhancers, we know it works well," he said. "For things related to long distance connectivity or the genome's shape, we're not as confident. "

Sankaran points out that MPRA really shines in its ability to find themes in [genetic variation](#) that researchers can marry to other genetic,

structural, or functional data.

"When you start to get all these independent pieces together, you get a real fine view of what's important," he said.

**More information:** Melnikov A, Murugan A, et al. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nature Biotechnology*. February 26, 2012. [DOI: 10.1038/nbt.2137](https://doi.org/10.1038/nbt.2137)

Ulirsch JC, Nandakumar SK, et al. Systematic functional dissection of common genetic variation affecting red blood cell traits. *Cell*. June 2, 2016.

Tewhey R, Kotliar D, et al. Direct identification of hundreds of expression-modulating variants using a multiplexed reporter assay. *Cell*. June 2, 2016.

Provided by Broad Institute of MIT and Harvard

Citation: High-throughput, sensitive approach helps reveal what's 'real' in genome-wide association data (2016, June 2) retrieved 26 April 2024 from <https://medicalxpress.com/news/2016-06-high-throughput-sensitive-approach-reveal-real.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.