

International gnomAD Consortium releases its first major studies of human genetic variation

May 28 2020, by Tom Ulrich



Credit: CC0 Public Domain

For the last eight years, the Genome Aggregation Database (gnomAD) Consortium (and its predecessor, the Exome Aggregation Consortium, or ExAC), has been working with geneticists around the world to compile and study more than 125,000 exomes and 15,000 whole genomes from populations around the world.

Now, in seven papers published in *Nature*, *Nature Communications*, and *Nature Medicine*, gnomAD Consortium scientists describe their first set of discoveries from the database, showing the power of this vast collection of data. Together the studies:

1. present a more complete catalog and understanding of a class of rare genetic variation called loss-of-function (LoF) variants, which are thought to disrupt [genes](#)' encoded proteins;
2. introduce the largest comprehensive reference map of an understudied yet important class of genetic variation called [structural variants](#);
3. show how tools that account for unique forms of variation and variants' biological context can help clinical geneticists when trying to diagnose patients with [rare genetic disease](#); and
4. illustrate how population-scale datasets like gnomAD can help evaluate proposed drug targets.

Researchers at the Broad Institute of MIT and Harvard and Massachusetts General Hospital (MGH) served as co-first or co-senior

authors on all of the studies, with scientists from Imperial College London in the United Kingdom, the direct-to-consumer genetics company 23andMe, and other institutions contributing to individual papers. More than 100 scientists and groups internationally have provided data and/or analytical effort to the consortium.

"These studies represent the first significant wave of discovery to come out of the gnomAD Consortium," said Daniel MacArthur, scientific lead of the gnomAD project, a senior author on six of the studies, an institute member in the Program in Medical and Population Genetics at Broad Institute, and now director of Centre for Population Genomics at the Garvan Institute of Medical Research and Murdoch Children's Research Institute in Australia. "The power of this database comes from its sheer size and population diversity, which we were able to reach thanks to the generosity of the investigators who contributed data to it, and of the research participants in those contributing studies."

"In a sense, gnomAD is the product of a consortium of consortia, in that the underlying data represents the work and contributions of many groups who have been collecting exome and genome sequences as a way of understanding human biology," said Konrad Karczewski, first author on the collection's flagship paper in *Nature* and a computational biologist at Broad and MGH's Analytic and Translational Genetics Unit. "Each of these papers represents someone bringing a new angle to the dataset, saying, 'I have an idea on how we can put all of this to work,' and creating a new resource for the genetics community. It was amazing to see it unfold."

GNOMAD LOOKBACK

MacArthur and his colleagues at Broad and MGH built ExAC and then gnomAD to expand on the work of the 1000 Genomes Project, the first large-scale international effort to catalog human genetic variation, and

other projects.

"In 2012, my lab was sequencing the genomes of patients with rare disease, and found that existing catalogs of normal variation weren't large or diverse enough to help us interpret the genetic changes we were seeing," MacArthur recalled. "At the same time, our colleagues around the world had sequenced tens of thousands of people for studies of common, complex disorders. So we set about bringing these datasets together to create a reference dataset for rare disease research."

The ExAC consortium released its first collection of whole exome data in October 2014. It then started gathering whole genome data, evolving into the gnomAD Consortium and releasing gnomAD v1.0 in February 2017.

Subsequent gnomAD releases focused on increasing the numbers of exomes and genomes, the volume of variants highlighted in the data, and the diversity of the dataset.

The new papers are based on the gnomAD v2.1.1 dataset, which includes genomes and exomes from more than 25,000 people of East and South Asian descent, nearly 18,000 of Latino descent, and 12,000 of African or African-American descent.

COMPREHENSIVE CATALOG

Two of the seven papers show how large genomic datasets can help researchers learn more about rare or understudied types of genetic variants.

The flagship study, led by Karczewski and MacArthur and published in *Nature*, describes gnomAD and maps loss-of-function (LoF) variants: genetic changes that are thought to completely disrupt the function of

protein-coding genes. The authors identified more than 443,000 LoF variants in the gnomAD dataset, dramatically exceeding all previous catalogs. By comparing the number of these rare variants in each gene with the predictions of a new model of the human genome's mutation rate, the authors were also able to classify all protein-coding genes according to how tolerant they are to disruptive mutations—that is, how likely genes are to cause significant disease when disrupted by genetic changes. This new classification scheme pinpoints genes that are more likely to be involved in severe diseases such as intellectual disability.

"The gnomAD catalog gives us our best look so far at the spectrum of genes' sensitivity to variation, and provides a resource to support gene discovery in common and rare disease," Karczewski explained.

While Karczewski and MacArthur's study focused on small variants (point mutations, small insertions or deletions, etc.), graduate student Ryan Collins, Broad associated scientist Harrison Brand, institute member Michael Talkowski, and colleagues used gnomAD to explore structural variants. This class of genomic variation includes duplications, deletions, inversions, and other changes involving larger DNA segments (generally greater than 50-100 bases long). Their study, also published in *Nature*, presents gnomAD-SV, a catalog of more than 433,000 structural variants identified within nearly 15,000 of the gnomAD genomes. The variants in gnomAD-SV represent most of the major known classes of structural variation and collectively form the largest map of structural variation to date.

"Structural variants are notoriously challenging to identify within whole genome data, and have not previously been surveyed at this scale," noted Talkowski, who is also a faculty member in the Center for Genomic Medicine at MGH. "But they alter more individual bases in the genome than any other form of variation, and are well established drivers of human evolution and disease."

Several surprising findings came out of their survey. For instance, the authors found that at least 25 percent of all rare LoF variants in the average individual genome are actually structural variants, and that many people carry what should be deleterious or harmful structural alterations, but without the phenotypes or clinical outcomes that would be expected.

They also noted that many genes were just as sensitive to duplication as to deletion; that is, from an evolutionary perspective, gaining one or more copies of a gene can be just as undesirable as losing one.

"We learned a great deal by building this catalog in gnomAD, but we've clearly only scratched the surface of understanding the influence of genome structure on biology and disease," Talkowski said.

TOOLS FOR BETTER DIAGNOSIS

Three of the papers reveal how gnomAD's deep catalogs of different types of genetic variation and the cellular context in which variants arise can help clinical geneticists more accurately determine whether a given variant might be protective, neutral, or harmful in patients.

In a *Nature* paper, Beryl Cummings, a former Broad/MGH graduate student now at Maze Therapeutics, MacArthur, and colleagues found that tissue-based differences in how segments of a given gene are expressed can change the downstream effects of variants within those segments on biology and disease risk. The team combined data from gnomAD and the Genotype Tissue Expression (GTEx) project to develop a method that uses these differences to assess the clinical significance of variants.

In *Nature Communications*, MacArthur, graduate student Qingbo Wang, and collaborators surveyed multinucleotide variants—ones consisting of two or more nearby base pair changes that are inherited together. Such

variants can have complex effects, and this study represents the first attempt to systematically catalog these variants, examine their distribution throughout the genome, and predict their effects on gene structure and function.

And in a separate *Nature Communications* study, MacArthur, Nicola Whiffin and James Ware of Imperial College London, and colleagues explored the impact of DNA variants arising in the 5-prime untranslated regions of genes, which are located just ahead of where the cell's transcriptional machinery starts reading a gene's protein code. Variants in these regions can trick a cell to start reading a gene in the wrong place, but haven't previously been well-documented.

"Clinical laboratories use gnomAD every day," said Heidi Rehm, a clinical geneticist; an institute member in Broad's MPG and medical director of the Clinical Research Sequencing Platform at Broad; chief genomics officer in the MGH Department of Medicine; and co-chair with Broad institute member Mark Daly of the gnomAD steering committee. "The methods in these studies are already helping us better interpret a patient's genetic test results."

GUIDING DRUG DEVELOPMENT

The remaining two gnomAD studies describe how diverse, population-scale genetic data can help researchers assess and pick the best drug targets.

In 2018, Broad associated scientist Eric Minikel mused on his research blog about whether genes with naturally-occurring predicted LoF variants could be used to assess the safety of targeting those genes with drugs. He wrote that if a gene that's naturally inactivated doesn't seem to have harmful effects, perhaps that gene could be safely inhibited with a drug. That blog post became the basis of a *Nature* paper in which Minikel,

MacArthur, and colleagues applied the gnomAD dataset to probe this question. They suggest ways to incorporate insights about LoF variants into the drug development process.

Leveraging the expertise at Broad, The Michael J. Fox Foundation initiated a collaboration between Imperial College's Whiffin, MacArthur, Broad postdoctoral fellow Irina Armean, 23andMe's Aaron Kleinman and Paul Cannon, and others to use LoF variants cataloged in gnomAD, UK Biobank, and 23andMe to study the potential safety liabilities of reducing the expression of a gene called LRRK2, which is associated with risk of Parkinson's disease. In *Nature Medicine*, they use these data to predict that drugs that reduce LRRK2 protein levels or partially block the gene's activity are unlikely to have severe side effects.

"We've cataloged large amounts of gene-disrupting variation in gnomAD," MacArthur said. "And with these two studies we've shown how you can then leverage those variants to illuminate and assess potential drug targets."

GROWING IMPACT

Public sharing of all data has been a core principle of the gnomAD project from its inception. The data behind these seven papers were publicly released via the gnomAD browser without usage or publication restrictions in 2016.

"The wide-ranging impact this resource has already had on medical research and clinical practice is a testament to the incredible value of genomic data sharing and aggregation," MacArthur said. "More than 350 independent studies have already made use of gnomAD for research on cancer predisposition, cardiovascular disease, rare genetic disorders, and more since we made the data available."

"But we are very far from saturating discoveries or solving variant interpretation," he added. "The next steps for the consortium will be focused on increasing the size and population diversity of these resources, and linking the resulting massive-scale genetic data sets with clinical information."

More information: Karczewski K, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. Online May 27, 2020. [DOI: 10.1038/s41586-020-2308-7](https://doi.org/10.1038/s41586-020-2308-7).

Collins R, Brand H, et al. An open resource of structural variation for medical and population genetics. *Nature*. Online May 27, 2020. [DOI: 10.1038/s41586-020-2287-8](https://doi.org/10.1038/s41586-020-2287-8).

Cummings B, et al. Transcript expression-aware annotation improves rare variant discovery and interpretation. *Nature*. Online May 27, 2020. [DOI: 10.1038/s41586-020-2329-2](https://doi.org/10.1038/s41586-020-2329-2).

Wang Q, et al. Landscape of multi-nucleotide variants in 125,748 human exomes and 15,708 genomes. *Nature Communications*. Online May 27, 2020. [DOI: 10.1038/s41467-019-12438-5](https://doi.org/10.1038/s41467-019-12438-5).

Whiffin N, et al. Characterising the loss-of-function impact of 5' untranslated region variants in 15,708 individuals. *Nature Communications*. Online May 27, 2020. [DOI: 10.1038/s41467-019-10717-9](https://doi.org/10.1038/s41467-019-10717-9).

Minikel E, et al. Evaluating potential drug targets through human loss-of-function genetic variation. *Nature*. Online May 27, 2020. [DOI: 10.1038/s41586-020-2267-z](https://doi.org/10.1038/s41586-020-2267-z).

Whiffin N, Armean IM, Kleinman A, et al. The effect of LRRK2 loss-of-function variants in humans. *Nature Medicine*. Online May 27, 2020.

[DOI: 10.1038/s41591-020-0893-5](https://doi.org/10.1038/s41591-020-0893-5).

Provided by Broad Institute of MIT and Harvard

Citation: International gnomAD Consortium releases its first major studies of human genetic variation (2020, May 28) retrieved 30 April 2024 from

<https://phys.org/news/2020-05-international-gnomad-consortium-major-human.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.