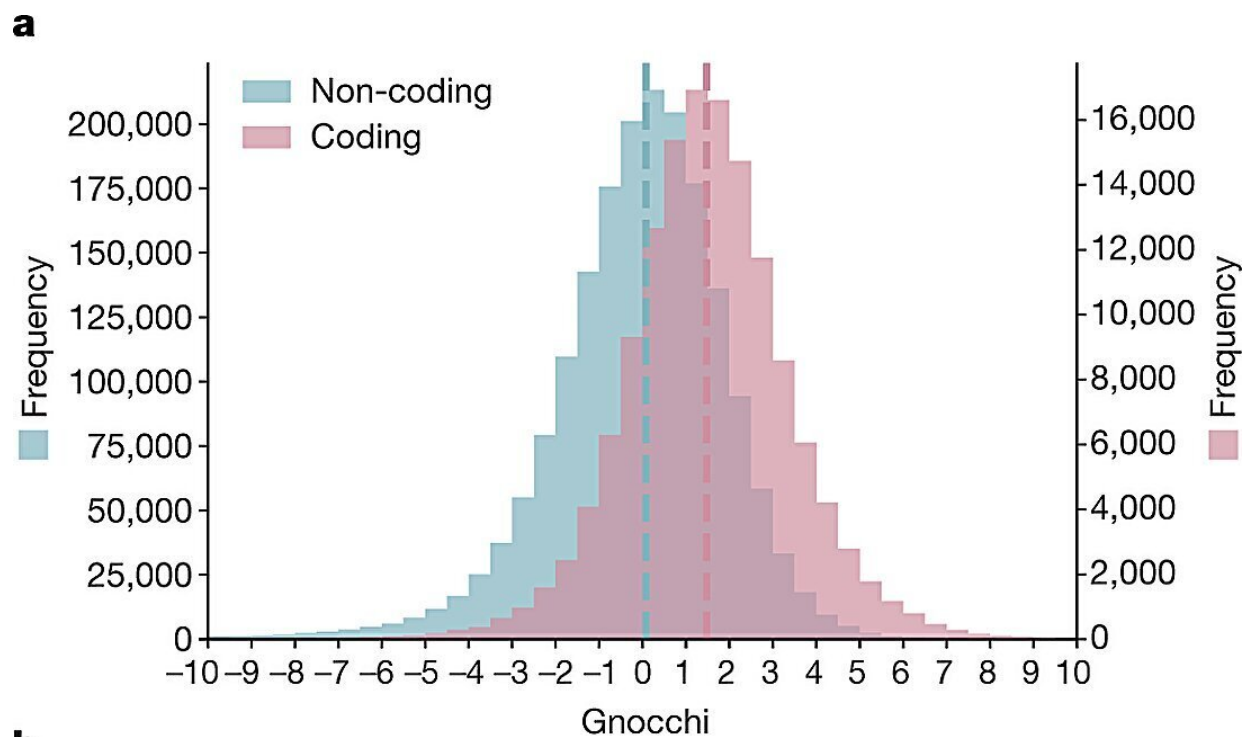


Large-scale DNA sequence resource reveals new regions of the human genome under natural selection

December 7 2023, by Marcela Quintanilla Dieck



b

0 100

Distribution of Gnocchi scores across the genome. Credit: *Nature* (2023). DOI: 10.1038/s41586-023-06045-0

Every human's genome has millions of genetic variants, but most have

little to no effect, making it difficult for clinicians to make medical diagnoses based on genetic differences.

Using patterns of variation from tens of thousands of individuals with whole-genome sequence data, a team led by investigators at Massachusetts General Hospital (MGH) and the Broad Institute of MIT and Harvard recently identified regions of the genome that lack typical variation, indicating that they are important sequences conserved during evolution and [natural selection](#).

The authors of the study, which is published in [Nature](#), note that when a variant arises in one of these regions, it's more likely to have an effect on an individual's health.

"We sought to examine how natural selection shapes patterns of human genetic variation across the whole genome, especially in the non-coding genome, which has been much less characterized than protein-coding regions," says senior author Konrad Karczewski, Ph.D., an Assistant Professor in the Analytic and Translational Genetics Unit in the Department of Medicine at MGH and Associate Member of the Broad Institute of MIT and Harvard.

"While our previous work evaluated the 2% of the genome that encodes genes, our new metrics extend to the [entire genome](#), greatly expanding our knowledge about which functional genomic elements likely harbor variation with potential clinical significance."

Karczewski and his colleagues aggregated and processed information from 76,156 [human genomes](#) into the Genome Aggregation Database (gnomAD), a large international human genome reference resource that they have been expanding and releasing to the public continuously.

The variants in this database have been helping clinical labs worldwide

perform diagnoses of rare diseases, and this release greatly expands the ability to do so in non-coding regions.

The team used the results to build a "genomic constraint map" for the [whole genome](#) (called Gnocchi, for Genomic NOn-Coding Constraint of HaploInsufficient variation). The map indicates which regions of the genome are "constrained," meaning that when variants in the region occur, they are often too damaging and are removed from the population by natural selection.

The team found that constrained regions are enriched for regulatory elements (which control gene expression) and variants implicated in complex human diseases and traits.

The scientists also found that more constrained regulatory elements tend to regulate more constrained protein-coding genes, which in turn suggests that studying non-coding constraint can aid in the identification of constrained genes.

"We anticipate that Gnocchi could be used to prioritize genetic variation discovered in non-coding regions of the [genome](#) in patients with [rare diseases](#), which can potentially provide clues for genetic causes of diseases and starting points for targeted therapeutics," explains Karczewski.

Next, it will be important to add genomic information from other individuals into this newly developed dataset.

"Future efforts towards a larger, more diverse human reference dataset would further improve rare disease diagnoses for all, and create better powered constraint metrics, giving us a better understanding of the distribution and effects of human [genetic variation](#)," says Karczewski.

More information: Siwei Chen et al, A genomic mutational constraint map using variation in 76,156 human genomes, *Nature* (2023). [DOI: 10.1038/s41586-023-06045-0](https://doi.org/10.1038/s41586-023-06045-0)

Provided by Massachusetts General Hospital

Citation: Large-scale DNA sequence resource reveals new regions of the human genome under natural selection (2023, December 7) retrieved 30 April 2024 from <https://medicalxpress.com/news/2023-12-large-scale-dna-sequence-resource-reveals.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.