➔ GENETIC VARIATION

Exploring human genomic diversity with gnomAD

In a coordinated set of seven publications, the Genome Aggregation Database (gnomAD) consortium presents a catalogue of human genetic variation on an unprecedented scale. As the successor of the Exome Aggregation Consortium (ExAC), which focused on exome sequence data for 60,706 individuals and identified 7.4 million small genetic variants in coding regions, gnomAD includes genetic variation for 15,708 whole genomes in addition to 125,748 exomes. The increase in sample size and inclusion of non-coding regions yielded more than 240 million small genetic variants as well as structural variation.

In the flagship publication, Karczewski et al. describe how they reached the latest release of gnomAD by aggregating, reprocessing and jointly variant-calling raw data from a total of 141,456 unrelated individuals, who were sequenced as part of various population genetic and complex disease-specific studies. The team identified 14.9 million and 229.9 million high-quality variants in the exome and genome data sets, respectively, after filtering. They then systematically catalogued predicted loss-of-function (pLoF) variants in whole-exome data, identifying 443,769 high-confidence pLoF variants in protein-coding genes. Moreover, the team classified all protein-coding genes according to their sensitivity to genetic disruption, along a spectrum ranging from tolerance to inactivation.

A list of genes that do not tolerate pLoF variants, which are often deleterious, supports the identification of essential genes, which could serve as candidate drug targets. Minikel et al. emphasize the value of the pLoF catalogue for therapeutic drug target discovery and validation and describe the general principles underlying such a pLoF-guided approach.

A case study in *Nature Medicine* further showcases the usefulness of using pLoF variant data to explore the safety profile of candidate drug targets. Previous model organism studies investigating LRRK2 inhibition in Parkinson disease raised concerns about potential on-target toxicity; analysis of large genomic data sets, including gnomAD, showed that heterozygous *LRRK2* loss-of-function mutations were not strongly associated with disease phenotypes, suggesting that therapeutic inhibitors targeting this gene remain a viable strategy.

Complementing the analysis of small genetic variants, Collins et al. set out to characterize structural variation (deletions, duplications, insertions or other rearrangements of DNA >50 bp in size) from 14,891 genomes across diverse global populations. The resulting sequence-resolved reference database, termed gnomAD-SV, more than doubles previously projected numbers, ultimately identifying 335,470 high-quality structural variants, corresponding to a median of 7,439 structural variants per genome.

In another publication, Cummings et al. demonstrate the utility of gnomAD data combined with RNA sequencing data from the Genotype-Tissue Expression (GTEx) project by developing a novel transcript-level annotation metric that improves the interpretation of rare disease variants. The two final publications investigated particular subsets of genetic variation in gnomAD: Whiffin et al. focused on variants that generate or destroy open reading frames in the 5′ untranslated region of protein-coding genes, whereas Wang et al. characterized the functional impact of multi-nucleotide variants (that is, clusters of variants on the same haplotype).

Since its initial release in October 2016, gnomAD has already proved itself as an invaluable clinical genetics resource, supporting the identification and interpretation of disease-causing variants and potential therapeutic targets. Moreover, comprehensive knowledge of the natural genetic variation across diverse human populations supports research into basic human biology. However, although the gnomAD data span six global and eight sub-continental ancestries, the diversity of sampled populations remains insufficient to comprehensively capture the extent of global human genetic variation as well as more population-specific variation. To address this issue, future iterations of gnomAD will increase sample size and diversity to drive further exploration into patterns of variation across the coding and non-coding genome.

Linda Koch

> " gnomAD has already proved itself as an invaluable clinical genetics resource "

Credit: P. Morgan/Springer Nature Limited

ORIGINAL ARTICLES Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020) | Minikel, E. V. et al. Evaluating drug targets through human loss-of-function genetic variation. *Nature* **581**, 459–464 (2020) | Whiffin, N. et al. The effect of LRRK2 loss-of-function variants in humans. *Nat. Med.* https://doi.org/10.1038/s41591-020-0893-5 (2020) | Collins, R. L. et al. A structural variation reference for medical and population genetics. *Nature* **581**, 444–451 (2020) | Cummings, B. B. et al. Transcript expression-aware annotation improves rare variant interpretation. *Nature* **581**, 452–458 (2020) | Whiffin, N. et al. Characterising the loss-of-function impact of 5′ untranslated region variants in 15,708 individuals. *Nat. Commun.* **11**, 2523 (2020) | Wang, Q. et al. Landscape of multi-nucleotide variants in 125,748 human exomes and 15,708 genomes. *Nat. Commun.* **11**, 2539 (2020)

RELATED LINKS
gnomAD: https://gnomad.broadinstitute.org/
gnomAD collection: https://www.nature.com/collections/afbgiddede