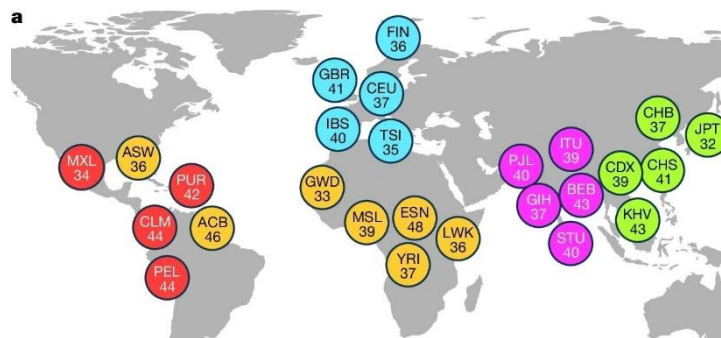


With Advanced Sequencing Scientists Open New Atlas of Genetic Diversity

Researchers investigated large-scale [structural variants](#) (SVs), complex and poorly understood insertions, deletions, and rearrangements in DNA, using next-generation 'long-read' sequencing. Their groundbreaking dataset comprised 1,019 individuals across 26 global populations. The study further leveraged a novel graph-based analytical framework, allowing for the creation of over 107,000 sequence-resolved biallelic SVs, which the authors made open-access.

The high-resolution genomic investigation not only significantly furthers our understanding of the true diversity of human genetics but also progresses our identification and future management of disease-causing [genetic variants](#) in patients.



Study

The present work does just this: A consortium of researchers undertook a massive, multinational project to map SVs using a globally diverse cohort. Study samples were acquired from the 1000 Genomes Project (1kGP) and initially comprised 1,064 samples ([lymphoblastoid cell lines](#)).

Strict quality control (QC) using a combination of [DNA concentration](#) determination (multimode microplate reader), DNA purity evaluation (spectrophotometer), and DNA fragment length verification (Femto Pulse system) reduced the dataset to 1,019. This dataset comprised participants from 26 distinct ancestries across Africa, the Americas, Europe, and East and South Asia.

The long-read sequencing platform used was the [Oxford Nanopore Technologies](#) (ONT) LRS, a cutting-edge technology capable of generating data with a median read length of over 20,000 base pairs.

To analyze this complex dataset, they engineered a novel computational framework called SAGA (SV analysis by [graph augmentation](#)). This process involved four key steps: First, aligning long reads to both linear (GRCh38) and graph-based (HPRC) references; second, SV discovery using Sniffles, DELLY, and the graph-aware SVarp algorithm, including specialized remapping to resolve inversion alignment artifacts; third, augmenting the pangenome graph to incorporate new SVs despite complexities in multiallelic VNTR genotyping; and finally, genotyping the cohort using Giggles software to determine variant carriers (n = 967 samples), noting that multiallelic sites showed higher Mendelian inconsistency (15.1%).

Results

The present study resulted in the production of a richly annotated, publicly available catalog of more than 100,000 sequence-resolved SVs (biallelic), alongside 369,685 multiallelic variable

number tandem repeats (VNTRs) genotyped using the Vamos tool. Identified SVs included inversions, deletions, duplications, and insertions, totalling a greater than tenfold increase in the number of fully resolved insertion sites, filling a critical gap in human [genomic knowledge](#).

Mendelian consistency experiments leveraging family trios (two parents and a child) within the cohort demonstrated the study's high accuracy and extremely low error rate (deletions and insertions at just 3.87% and 4.44%, respectively) for biallelic SVs. Notably, most of the novel SVs identified in this study were found to be extremely rare, with 59.3% having a [minor allele frequency](#) (MAF) of less than 1%. Individuals of African descent demonstrated the highest degree of SV diversity.

Finally, the study provided novel insights into the biological mechanisms that create SVs, detailing how mobile DNA elements, such as L1 and [SVA retrotransposons](#), drive genetic innovation by promoting SV formation and translocation through locus-specific processes, including promoter hijacking (e.g., the 8q21.11 L1 source element).

Conclusion

The present study represents a commendable leap forward in our knowledge and understanding of [human genomics](#). The application of long-read sequencing successfully allowed for the discovery and annotation of more SVs (especially insertions), and the diversity of the sample cohort (26 distinct ancestries across several continents) validates the generalizability and global application of study findings.

Furthermore, the resultant comprehensive and accurate SV atlas, being open access, opens the doors to a new era of genetic medicine, allowing for the identification and early treatment of genetic conditions that we hitherto didn't even know existed. Notably, when applied to rare-disease genomes, the resource filtered 55% of candidate SVs while retaining 94% (35/37) of validated causal variants. This open-access resource will be invaluable for the scientific community, enabling a deeper understanding of human evolution, population genetics, and the functional consequences of [genetic variation](#).

Source:

<https://www.news-medical.net/news/20250724/Scientists-open-new-atlas-of-genetic-diversity-with-advanced-sequencing.aspx>