



(2025/12/10)

From dots to lines: new database catalogs human gene types using 'ACTG' rules

The Joint Open Genome and Omics Platform 1.0 (JoGo 1.0) catalogs 19,194 human genes with a novel naming system. Researchers expect the database to benefit medicine and genomics, and to provide a common language for human gene types.

Fukuoka, Japan—Whether you turn red when drinking alcohol, dislike certain smells, or metabolize drugs differently from others, the explanation often lies in your DNA, or more precisely, your gene types.

People share the same genes but not the exact same gene types. These types are unique combinations of multiple DNA sequence differences that together shape our biological traits. Researchers have long investigated these genetic variations, but traditional tools analyze only 150-300 bases at a time, providing isolated “dots” of information. Advances in long-read sequencing, which can read tens of thousands of bases at once, are now connecting these dots into “lines,” showing how variations work together as functional gene types.

Yet without a standard naming system, researchers remain stuck describing each variant in fragmented and redundant ways.

“It is like explaining a cup by only listing the shape of its handle, its color, or other separate features. It creates barriers to cross-study comparison and slows translation into healthcare,” says Professor [Masao Nagasaki](#) of [Kyushu University's Medical Institute of Bioregulation](#). “For example, fields like transplant matching or drug metabolism have their own naming schemes, but none are widely adopted.”

To address this gap, Nagasaki's team introduced the ACTG hierarchical nomenclature and built a global database called the [Joint Open Genome and Omics Platform 1.0](#) (JoGo 1.0) —a project spanning nearly five years including data acquisition, with about two and a half years devoted to constructing the database itself. The work was recently published in [Nucleic Acids Research](#) and was selected as one of the journal's Breakthrough Articles.

Inspired by the four fundamental DNA bases, the ACTG naming system organizes human gene types into four progressively expanding levels: A for the amino acid sequence, C for the coding sequence, T for the transcript level covering untranslated regions, and G for the complete gene body including introns.

“One key feature is that we rank gene types based on global frequency,” Nagasaki explains.

For example, the most common variant of the gene Aldehyde Dehydrogenase 2 (ALDH2)—the key enzyme that breaks down acetaldehyde—is designated as ALDH2:a1c1t1g1. A variant with reduced enzymatic activity, often found in East Asian populations and responsible for the flushed red face people experience when consuming alcohol, is categorized in the system as ALDH2:a2. This variant represents a change at the amino acid level. The numbering system indicates global frequency: a lower number signals a more common variant, while a higher

number points to greater rarity, and may be associated with a higher risk for certain diseases.

The database draws on DNA data from 258 genomes sampled across five continents—150 sourced from public resources and 108 newly sequenced from cell samples contributed by volunteers in the 1000 Genomes Project.

JoGo 1.0 offers both an interactive online viewer and a privacy-preserving local viewer, enabling secure integration of sensitive datasets.

Fittingly, “JoGo” means “funnel” in Japanese, reflecting the database’s role in compressing massive genomic information into meaningful, usable knowledge. It catalogs 4.7 million gene types (haplotype) across more than 19,000 genes, and can link each gene type to public resources such as [ClinVar](#), the [GWAS Catalog](#), and [GTEx](#). This allows researchers to interpret clinical variants, trait associations, and tissue-specific gene expression. Moreover, with data representing all five inhabited continents, JoGo 1.0’s visualizations can highlight geographically distinct patterns, aiding population-specific genetic screening and informing drug development.

Nagasaki and his team are continuously expanding the database, increasing both sample size and population diversity, and expect to release JoGo 2.0 within two years. As more genomes are added, the frequency-based numbering will be refined to better reflect global patterns.

“Having consistent names for whole genes means we can finally speak a common language,” says Nagasaki. “Just as there is active research and discussion around blood types today, I hope this new nomenclature will lead to a deeper understanding of, and public dialogue around, human gene types.”

###

For more information about this research, see “JoGo 1.0: the ACTG hierarchical nomenclature and database covering 4.7 million haplotypes across 19,194 human genes,” Masao Nagasaki, Toshiaki Katayama, Yuki Moriya, Yayoi Sekiya, Shuichi Kawashima, Ryo Teraoka, Shuto Machida, Taichi Matsubara, Hiroki Hashimoto, Akihiro Asakura, Akio Nagano, Riu Yamashita, Toyoyuki Takada, Nobutaka Mitsuhashi, Mayumi Kamada, Yasuyuki Ohkawa, Katsushi Tokunaga, Yosuke Kawai, Variant Information Standardization Collegium, *Nucleic Acids Research*, <https://doi.org/10.1093/nar/gkaf1232>

About Kyushu University

Founded in 1911, [Kyushu University](#) is one of Japan's leading research-oriented institutions of higher education, consistently ranking as one of the top ten Japanese universities in the Times Higher Education World University Rankings and the QS World Rankings. Located in Fukuoka, on the island of Kyushu—the most southwestern of Japan’s four main islands—Kyushu U sits in a coastal metropolis frequently ranked among the world’s most livable cities and historically known as Japan’s gateway to Asia. Its multiple campuses are home to around 19,000 students and 8,000 faculty and staff. Through its [VISION 2030](#), Kyushu U will “drive social change with integrative knowledge.” By fusing the spectrum of knowledge, from the humanities and arts to engineering and medical sciences, Kyushu U will strengthen its research in the key areas of decarbonization, medicine and health, and environment and food, to tackle society’s most pressing issues.

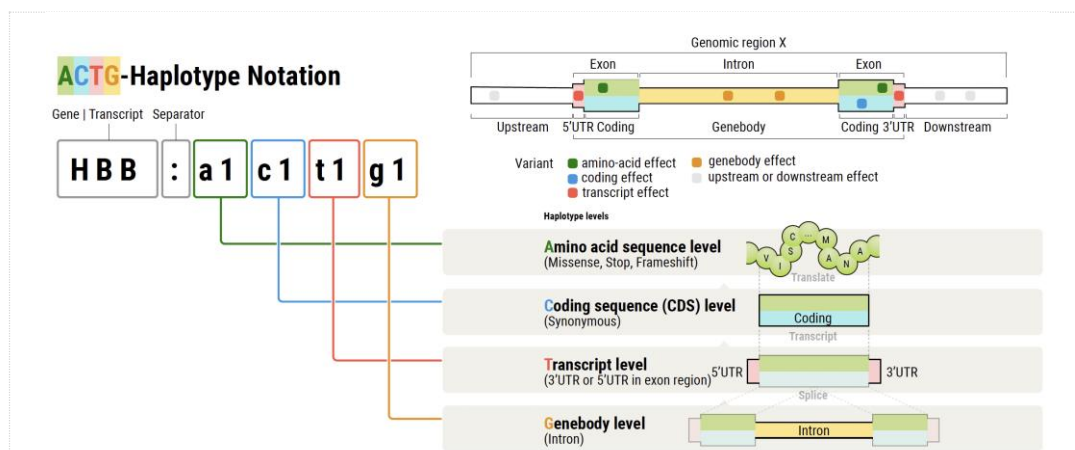


Fig. 1. Graphical abstract: the ACTG hierarchical naming system

The ACTG nomenclature organizes gene variants into four levels, each capturing progressively broader genomic regions. Variants are ranked by global frequency, with lower numbers indicating more common variants worldwide.

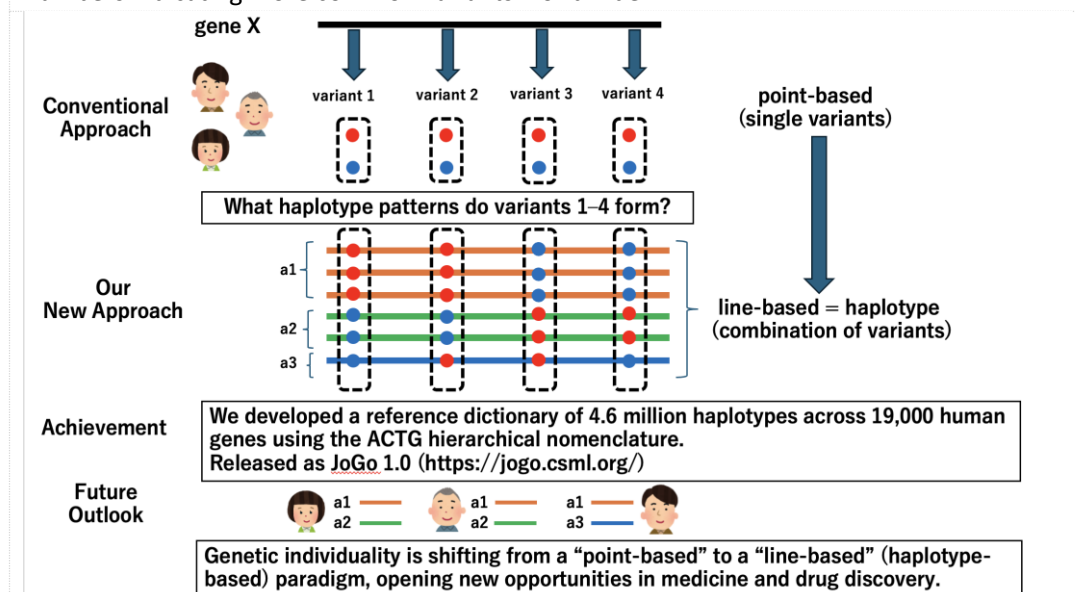


Fig. 2. A new nomenclature connects dots to lines

The ACTG naming system shifts genetic interpretation from isolated "point-based" variants to full "line-based" haplotypes—groups of genetic variations on the same chromosome that are inherited together from a single parent. Using this framework, researchers built JoGo 1.0, cataloging 4.7 million variants across more than 19,000 human genes.

[Contact]

Professor name and title: Masao Nagasaki, Professor

Professor Affiliation: Medical Institute of Bioregulation, Kyushu University

Tel: +81 092-642-4815

E-mail: nagasaki.masao.822@m.kyushu-u.ac.jp