Human genome haplotype database

JoGo Portal

https://jogo.csml.org/

The Joint Open Genome and Omics Platform (JoGo) Platform (https://jogo.csml.org/) is a database containing a comprehensive haplotype catalog of human genes.

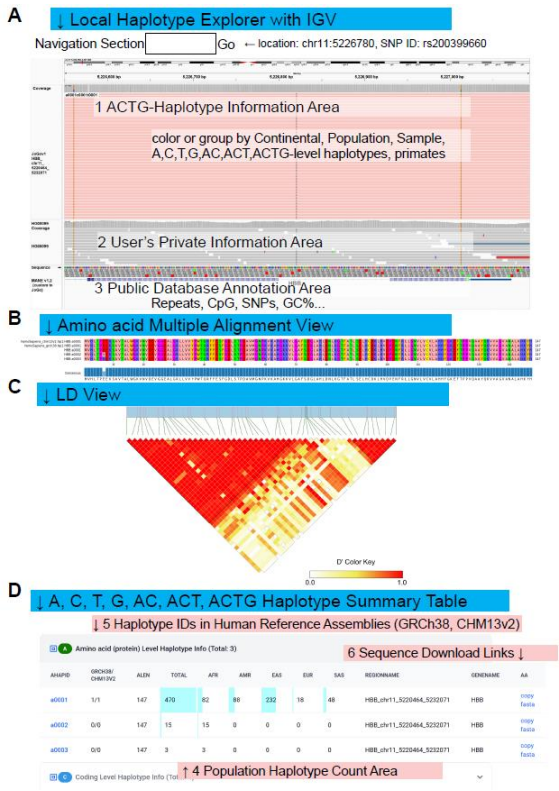
Previous gene mutation analysis has primarily focused on investigating the effects of individual gene mutations or polymorphisms. However, some effects of gene mutations and polymorphisms, such as physical characteristics, are determined by combinations of multiple gene mutations and polymorphisms (haplotypes).

A familiar example of this is the ABO blood type system. The ABO blood type is determined by the combination of two types of sugar chain antigens on the surface of red blood cells. Type A sugar chains have N-acetylgalactosamine attached to their ends, while Type B sugar chains have galactose attached. Depending on the combination of sequences at the two glycosyltransferase loci on chromosome 9—inherited from both parents—that produce the A and B sugar chains, different sequences emerge: AA, AO, BB, BO, AB, and OO (where “O” indicates that no sugar chains of that type are attached). Consequently, this determines the blood types: Type A (AA, AO), Type B (BB, BO), Type AB (AB), and Type O (OO).

The importance of haplotype analysis has long been recognized, but it has been applied only to a limited number of complex gene families, such as human leukocyte antigen (HLA), killer immunoglobulin-like receptors (KIR), and cytochrome P450 (CYP) loci. However, because it was difficult to determine accurate haplotypes over long regions using the short-read sequencing methods that had been established for many genomic sequences, no haplotype database capable of analyzing the entire genome had existed until now.

JoGo cataloged and included in its database haplotypes across the entire genome, based on the genome sequences of 258 individuals from five continents, determined using long-read sequencing methods, including samples from 105 Japanese individuals. Each haplotype is linked to clinical annotation (ClinVar), phenotypic information (GWAS catalog), and cross-tissue expression correlation information (GTEx). Furthermore, based on gene expression information from 1,280 immortalized B cells analyzed in three research projects, the relationship between haplotype and gene expression levels in the same sample can also be investigated.

It is said that 99.9% of the human genome is the same, but it is believed that the remaining 0.1% difference determines individual differences in constitution, such as susceptibility to certain diseases or likelihood of experiencing side effects from certain medications. Until now, these differences in individual constitution could not be explained by differences in individual gene mutations or polymorphisms (= "points"), but now it is possible to understand them as "haplotypes," which are combinations of sequences (= "lines"), and it is expected that this will allow us to explain human genetic diversity more broadly.



(A) Sequence browsing screen using IGV. (B) Haplotype sequence alignment at the A level. (C) Linkage disequilibrium (LD) heatmap at the ACTG level. (D) Haplotype counts by ethnic group.

JoGo Portal is supported under JST Database Integration Coordination Program (DICP), "Development of research and educational platform with open human genomic and omics international database (Principal Investigator: NAGASAKI Masao, Professor, Medical Institute of Bioregulation, Kyushu University)".

Reference

Masao Nagasaki *et al.*, JoGo 1.0: the ACTG hierarchical nomenclature and database covering 4.7 million haplotypes across 19,194 human genes. Nucleic Acid Research 2026. (DOI:10.1093/nar/gkaf1232).

Japan Science and Technology Agency（JST）
Department of Information Services, Office of NBDC Program
Tel: 03-5214-8491　E-mail: nbdc-dicp@jst.go.jp

(Apr 2026)

ACTG hierarchical haplotype nomenclature

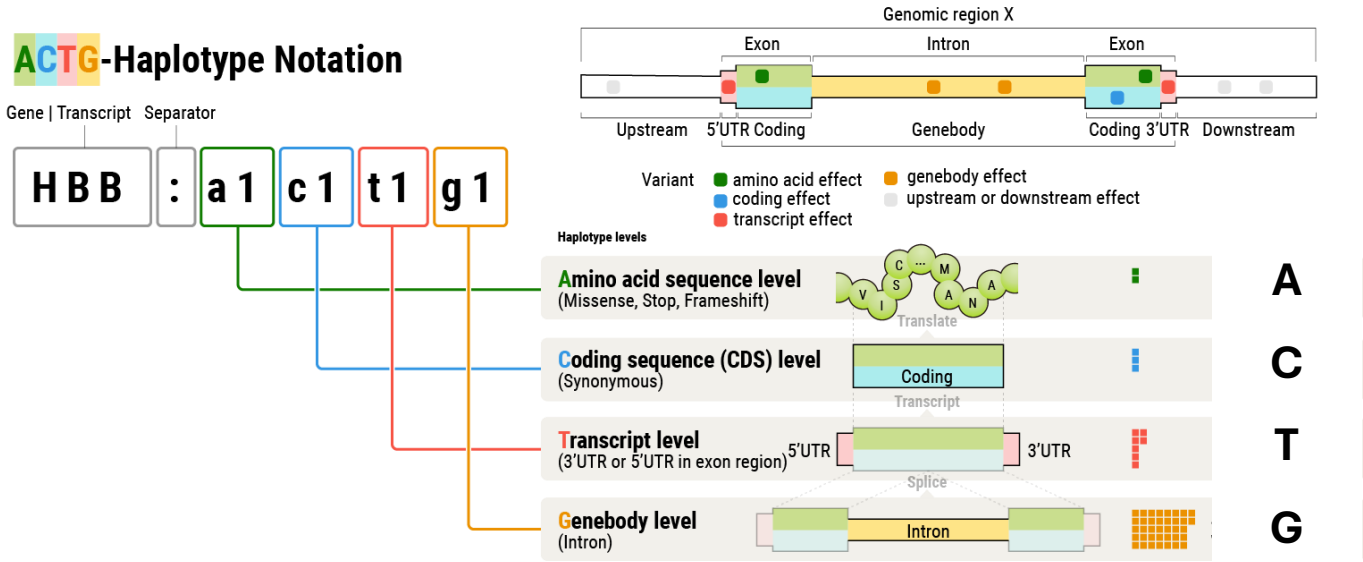
JoGo Portal

https://jogo.csml.org/

Until now, there has been no haplotype nomenclature that can be applied across the entire genome. Therefore, Professor Nagasaki and his colleagues have proposed the "ACTG hierarchical nomenclature," which can assign a unique haplotype ID across the entire genome.

In the ACTG hierarchical haplotype nomenclature, haplotypes are defined in four hierarchical levels:

A: Variants with amino acid substitutions within the protein-coding region
C: Variants without amino acid substitutions within the protein-coding region
T: Variants in the non-coding region of the transcript (5'-UTR and 3'-UTR)
G: Variants within introns of genomic gene regions



Number of data points included in JoGo Platform v.1.0 (as of December 2, 2025)

Number of samples analyzed: 258 samples

Of which, 108 samples are of Japanese origin

Number of genes (MANE standard protein-coding genes): 19,194 genes

Number of haplotypes:	4,656,478
A: (Variants with amino acid substitutions within the protein-coding region):	174,376
C: (Variants without amino acid substitutions within the protein-coding region):	300,610
T: (Variants in the non-coding region (5'-UTR and 3'-UTR) of the transcript region):	486,288
G: (Variants within introns of the gene region):	3,695,204

Reference

Masao Nagasaki *et al.*, JoGo 1.0: the ACTG hierarchical nomenclature and database covering 4.7 million haplotypes across 19,194 human genes. Nucleic Acid Research 2026. (DOI:10.1093/nar/gkaf1232).

Japan Science and Technology Agency (JST)
Department of Information Services, Office of NBDC Program
Tel: 03-5214-8491 E-mail: nbdc-dicp@jst.go.jp

(Apr 2026)