

ヒトゲノムのハプロタイプデータベース



JoGo Portal
<https://jogo.csml.org/>



Joint Open Genome and Omics Platform (JoGo) Platform (<https://jogo.csml.org/>) は、ヒト遺伝子を網羅するハプロタイプカタログを収載したデータベースです。

これまでの遺伝子変異解析では、個々の遺伝子変異や多型が与える影響を調べる研究が中心でした。しかしながら、体質など遺伝子の変異や多型が与える影響には、複数の遺伝子の変異や多型の組み合わせ（ハプロタイプ）によって決まるものがあります。

その身近な例が、ABO血液型です。ABO式血液型は、赤血球細胞の表面の2種類の糖鎖抗原の組み合わせによって決まります。A型糖鎖には末端にN-アセチルガラクトサミンが、B型糖鎖にはガラクトースが結合しています。9番染色体上にある両親から受け継いだA型糖鎖とB型糖鎖を作る2箇所の糖転移酵素の遺伝子座の配列の違いの組み合わせによって、AA、AO、BB、BO、AB、OO（「O」はその型の糖鎖が付加しない）の配列の違いが生じ、その結果として、A型（AA、AO）、B型（BB、BO）、AB型（AB）、O型（OO）の血液型が決まります。

ハプロタイプ解析の重要性は以前から認識され、ヒト白血球抗原（HLA）、キラー免疫グロブリン様受容体（KIR）、シトクロムP450（CYP）遺伝子座など、限られた数の複雑な遺伝子ファミリーに対してのみ行われてきました。しかしながら、短鎖型シーケンス法で決められた多くのゲノム配列では、長い領域にわたる正確なハプロタイプを決めることが困難だったため、ゲノム全体にわたって解析可能なハプロタイプデータベースはこれまで存在しませんでした。

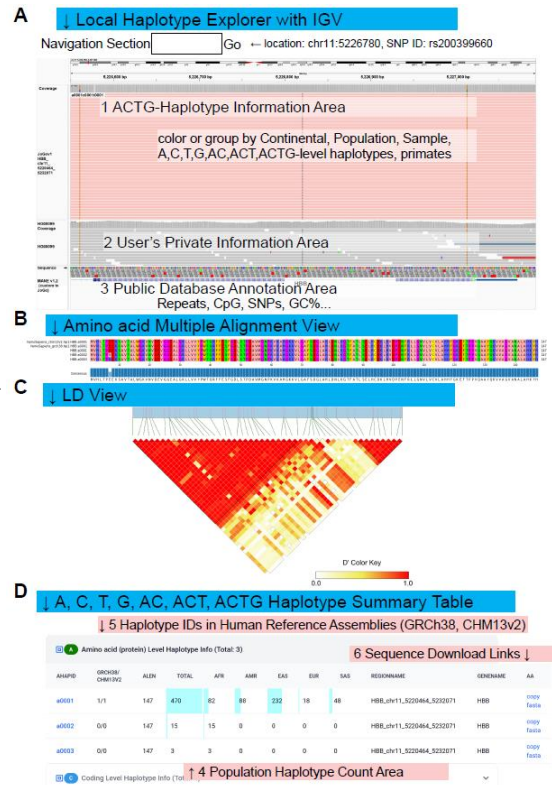
JoGoでは、日本人105人に由来するサンプルを含む長鎖型シーケンス法で決定された5大陸の258人分のゲノム配列をもとに、ゲノム全域にわたるハプロタイプをカタログ化し、データベースに収載しました。そして、各ハプロタイプを臨床アノテーション（ClinVar）、表現型情報（GWASカタログ）複数組織間の発現相関情報（GTEx）にリンクさせるとともに、3つの研究プロジェクトで解析された1,280人分の不死化B細胞における遺伝子発現情報をもとに、同一検体におけるハプロタイプと遺伝子発現量の関係についても調べることができるようになっています。

ヒトのゲノムは99.9%同じと言われていますが、残り0.1%の違いによって、特定の病気への罹りやすさとか、特定の薬への副作用の出やすさといった個人の体質の違いが決まるものと考えられています。これまで、一つ一つの遺伝子の変異や多型（＝“点”）の違いでは説明できなかった個人の体質の違いを、配列の組合せ（＝“線”）である「ハプロタイプ」として捉えることができたようになったことで、人間の遺伝学的多様性を広く説明できるようになるものと期待されます。

JoGo Portalは、JST統合化推進プログラムの研究開発課題「日本人塩基配列情報の公開可能なゲノム・オミクス情報基盤による双方向型研究教育データベース開発と国際連携」（研究代表者 長崎 正朗 九州大学 生体防御医学研究所 教授）で支援しています。

参考文献

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).



(A) IGVを用いた配列閲覧画面。(B) Aレベルでのハプロタイプ配列アライメント。(C) ACTGレベルの連鎖不平衡(LD)ヒートマップ。(D) 民族集団別のハプロタイプの集団別カウント。

ACTG階層命名法



JoGo Portal
<https://jogo.csml.org/>



これまでゲノム全体にわたって適用可能なハプロタイプの命名法はありませんでした。そこで、長崎教授らは、ゲノム全体で一意的ハプロタイプIDを付与できる「ACTG階層命名法」を提案しています。

ACTG階層命名法では、ハプロタイプを、

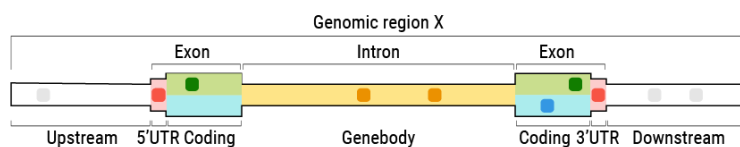
- A:** タンパク質コード領域内のアミノ酸置換を伴うバリエント
- C:** タンパク質コード領域内のアミノ酸置換を伴わないバリエント
- T:** 転写産物の非コード領域 (5'-UTRと3'-UTR) のバリエント
- G:** ゲノム遺伝子領域のイントロン内のバリエント

の4つの階層で定義しています。

ACTG-Haplotype Notation

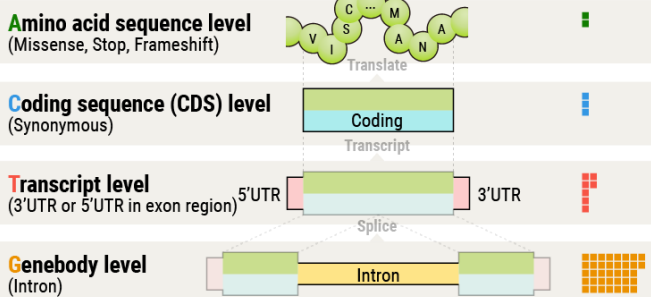
Gene | Transcript Separator

HBB : a1 c1 t1 g1



Variant
● amino acid effect ● genebody effect
● coding effect ● upstream or downstream effect
● transcript effect

Haplotype levels



A
C
T
G

JoGo Platform v.1.0の収載データ数 (2025年12月2日現在)

解析対象サンプル数	258	サンプル
うち日本人由来サンプル	108	サンプル
遺伝子数 (MANE標準タンパク質コード遺伝子)	19,194	遺伝子
ハプロタイプの数	4,656,478	件
A: (タンパク質コード領域内のアミノ酸置換を伴うバリエント)	174,376	件
C: (タンパク質コード領域内のアミノ酸置換を伴わないバリエント)	300,610	件
T: (転写産物領域の非コード領域(5'-UTRと3'-UTR)のバリエント)	486,288	件
G: (遺伝子領域のイントロン内のバリエント)	3,695,204	件

参考論文

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).