

PRESS RELEASE (2025/12/03)

ヒト遺伝子の“型”を網羅 データベース「JoGo」一般公開

～1.9 万遺伝子に対する 470 万の配列型を収載、病気や遺伝解析を後押し～

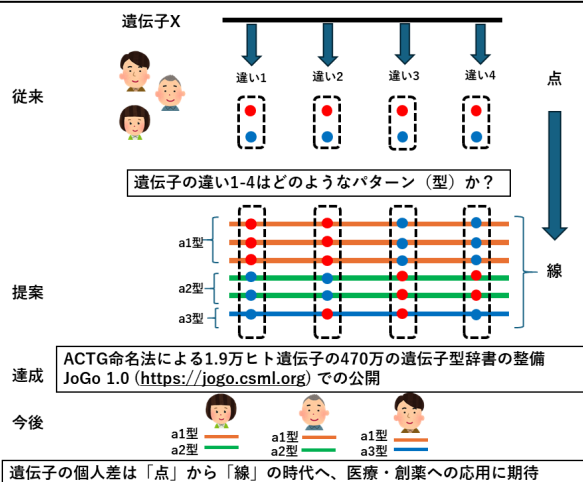
ポイント

- ① 全遺伝子の“型”を統一規則で記述する共通言語が求められていた
- ② 世界初、ヒト 1.9 万の遺伝子の“型”を ACTG 階層で体系化した辞書を構築
- ③ 遺伝子の個人差は「点」から「線」の時代へ、医療・創薬への応用に期待

概要

遺伝子の型は身近な例に、血液型が挙げられます。遺伝子は塩基配列から構成されており、その塩基配列の型（ハプロタイプ（※1））は個人ごとに（または一人一人、人によって）少しずつ違います。その違いが遺伝子の発現量や薬の代謝機能といった遺伝子機能の差を生み出すことが知られています。しかし、全遺伝子を網羅する共通の“型”の命名法や辞書は、これまでありませんでした。九州大学生体防御医学研究所の長崎正朗教授らは、汎用的な新規遺伝子型表記法（ACTG 階層命名法（※2））を確立し、ヒトの 19,194 遺伝子について、4,656,478 個の遺伝子型を整理しました。本データベースは、JoGo (Joint Open Genome and Omics: ジョーゴ; <https://jogo.csml.org>)（※3）からアクセスできます。

JoGo は、世界集団（※4）を対象に、長鎖型シーケンサ（※5）の情報に基づき、遺伝子型を統一規則でカタログ化した世界初のデータベースです。今回公開した同データベースの 1.0 版である JoGo 1.0 では、5 大陸の 258 人から得られた全ゲノム長鎖型シーケンサ由来の塩基配列情報を解析しました（このうち 108 人は本プロジェクトチームが測定）。ACTG 階層命名法は、A（アミノ酸）／C（コーディング）／T（転写：UTR）／G（遺伝子本体）の遺伝子型を定義し、世界集団における頻度順で ID を付番します。また、国際参照配列（※6）（GRCh38 と CHM13v2）由来の遺伝子型も収載しており、同一の番号空間で直接比較できるのが特徴です。さらに、臨床、形質、遺伝子発現に関連する公共データベース（ClinVar（※7）、GWAS Catalog（※8）、GTEx（※9）など）の情報を、遺伝子型に重ねて閲覧できます。本 ACTG 階層命名法により、ゲノム・メディカルサイエンスの遺伝子を表現するための辞書に、いままで遺伝子解析は、個々の変異という「点」の情報が中心でしたが、ACTG 階層命名法により、遺伝子全体の配列型という「線」の情報を扱えるようになります。この辞書を整備することで、病気との関連や薬の反応をより高い解像度で把握できるようになります。JoGo 1.0 では、オンラインでの遺伝子型の閲覧機能に加え、個人データを外部に出さないローカル閲覧にも対応しています。本成果は Nucleic Acids Research に Breakthrough Article（※10）として 2025 年 11 月 29 日に掲載されました。



研究者からひとこと：遺伝子型の身近な例に、血液型が挙げられます。例えば、ABO 遺伝子（※11）の配列型が血液型を決めます。JoGo の ACTG ハプロタイプ命名規則を用いることで、ヒトの持つ約 2 万の遺伝子についても、遺伝子型に基づく共通言語で議論できるようになります。近い将来、「私は〇〇遺伝子型が××型なのでこの匂いが好きだ」といった表現が、一般社会にも普及していくことを期待しています。（長崎 正朗）

【研究の背景と経緯】

ヒトの遺伝子は、同じ遺伝子であっても、個人間を比較すると塩基配列の型（ハプロタイプ）が一般的に異なります。ヒト集団全体で考えた場合、遺伝子型の違いは、通常、1カ所の塩基配列の違いではなく、遺伝子上の複数カ所の塩基配列の違いによります（図1）。生体内において遺伝子単位で機能しており、これら複数カ所の差がその遺伝子の機能の差（例えば、遺伝子自身の発現量、薬の代謝機能、ウイルスの免疫反応の差など）を生み出します。

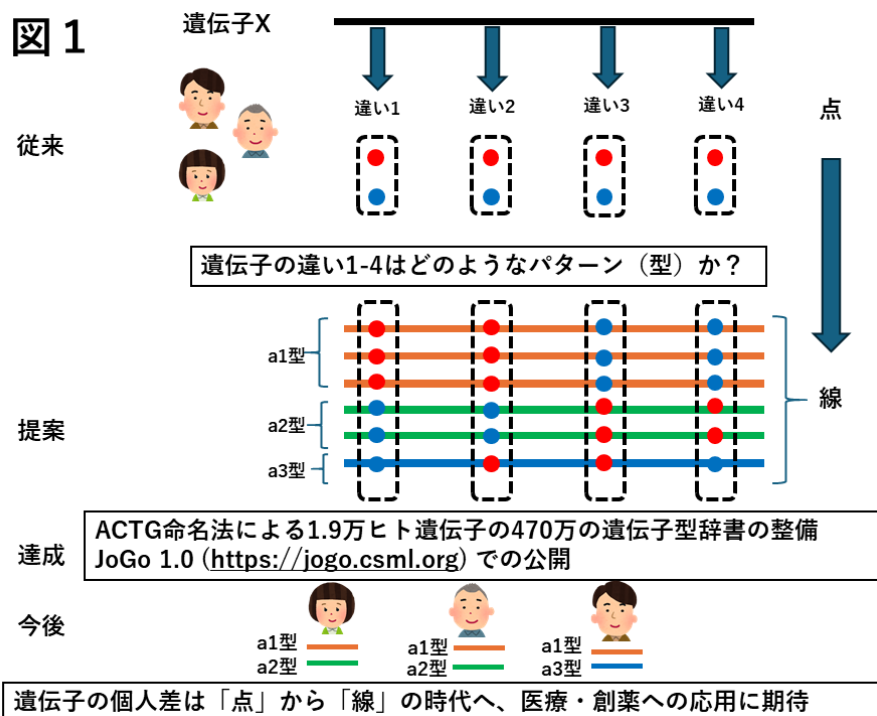


図1 遺伝子の“点”の違いと“型”の違いの模式図

個々の変異（点）の違いだけでなく、複数の変異が組み合わさって生じる遺伝子全体の配列型（ハプロタイプ）が、個人間の遺伝子差を形成することを示す概念図。ACTG 階層命名法により、19,194 遺伝子・約 470 万の遺伝子型を体系的に整理できることを示している。

このように、遺伝子型が遺伝子単位で機能することは知られており、一部の免疫系の遺伝子（HLA や KIR（※12））、薬剤代謝の遺伝子（CYP（※13））では、各研究コミュニティで、対象とする遺伝子型の命名規則とデータベース構築が進められてきました。これらの取り組みにより、免疫系や薬剤代謝の各領域において個別の命名規則とデータベースが普及し、基礎研究から臨床応用においても、日々活用されています。

一方、遺伝子型を対象として情報整備することの重要性は認識されていましたが、これまで、ヒト全遺伝子を対象にした型の共通命名法、また、その共通命名法に基づく包括的な遺伝子型のデータベースは存在しませんでした。

この理由の1つとして、一般的な塩基配列を測定する機器は、あらかじめ決まった位置の塩基配列情報を分離する測定機器（SNP アレイ（※14））や数百塩基の DNA 断片を測定する機器（短鎖型シーケンサ（※15））に依存してきたことが挙げられます。ヒトゲノムの遺伝子の全長が数千から数十万塩基配列であること、異なる遺伝子の間で、配列パターンが局所的によく似た領域を持つ遺伝子があるため、遺伝子全体の塩基配列型を正確に決定することが難しい状況でした。

一方、近年の新たな測定機器として、数万塩基の DNA 断片を測定する機器（長鎖型シーケンサ）の1塩基あたりのエラー率が改善されたことで、遺伝子全体の塩基配列型を正確に決めることが可能になりつつあります。

【研究の内容と成果】

そこで、本研究では、共通した命名規則により、ヒトの遺伝子型を生物学的な意味に基づいて階層的に整理する ACTG 階層命名法（A（アミノ酸）／C（コーディング）／T（転写：UTR）／G（遺伝子本体、オプションとして U：遺伝子領域の上流／D：下流の配列情報を付与））を確立しました。例えば、アルコール代謝産物であるアセトアルデヒド（※16）を分解する重要な酵素（ALDH2）の最もよく見られる遺伝子型は、ALDH2:a1c1t1g1 と記述します。この各 A、C、T、G 階層の番号は、参照集団における出現頻度順（a1≥a2…）で付番されるのが特徴です。東アジア集団でみられるアセトアルデヒドの分解能が a1 に比べて相対的に低い遺伝子型は、アミノ酸階層の違いであり、また、頻度が世界集団で見た場合に低いため、ALDH2:a2 型に分類されます。

Joint Open Genome and Omics Platform (JoGo) の 1.0 版 (JoGo 1.0) では、5 大陸、258 人の全ゲノム長鎖型シーケンサ由来の情報（本プロジェクトチームで 108 検体のシーケンス情報を測定）とヒト国際参照配列（GRCh38 と chr13v2）の情報をヒトの遺伝子型をサンプリングする参照集団として選定しました。基礎研究、臨床において用いられている国際参照配列も含めることで、国際参照配列に含まれる各遺伝子型の世界集団における位置づけを ACTG 階層命名法で付番された番号を確認することで比較可能にすることを目的としています。

ACTG 階層命名法に基づき、これらの大量の塩基配列情報を、九州大学のオミクスサイエンスセキュア情報解析システムを中心とし、九州大学、東京大学、京都大学の大型計算機システムを用いて計算をすることで情報解析を進めました。

JoGo 1.0 では、米国の NCBI（国立生物工学情報センター）とヨーロッパの EMBL-EBI（欧州バイオインフォマティクス研究所）が共同で進めたヒト遺伝子の標準遺伝子とその標準転写物を統一する規格（MANE（※17））によって定義された遺伝子のうち 19,194 個のヒト遺伝子を対象として、遺伝子型を収集し、合計 174,376 の A 階層 300,610 の C 階層、486,288 の T 階層、3,695,204 の G 階層の遺伝子型（総計 4,656,478）をカタログ化することに成功しました。

これにより、世界の研究者が遺伝子を 1 つ 1 つの変異という「点」の情報に加え、新たに遺伝子型という「線」の共通言語を用いて情報を整理、共有できる基盤を整備することができました。

これらすべての遺伝子型情報は、JoGo 1.0 (<https://jogo.csml.org>) から検索でき、オンラインのハプロタイプビューアで可視化できます（図 2）。

図 2 HBB 遺伝子型をオンラインハプロタイプビューアで表示した結果

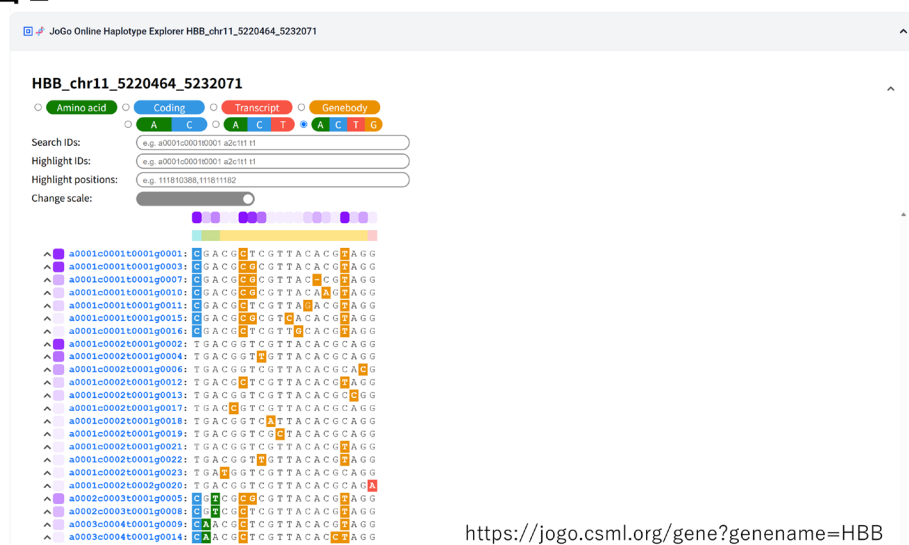


図 2 HBB 遺伝子の ACTG ハプロタイプをオンラインハプロタイプビューアで表示した例

JoGo 1.0 のオンラインビューアにより、HBB 遺伝子の A/C/T/G 各階層における差分や塩基置換の位置関係を視覚的に確認できることを示した画面例。

※HBB（β グロビン）遺伝子：赤血球中ヘモグロビンの β 鎖を作る遺伝子で、酸素運搬に重要。変異はサラセミア（遺伝性貧血）などの原因となることが知られている。

（JoGo：<https://jogo.csml.org/gene?genename=HBB>）

さらに、JoGo 1.0 では、IGV ソフトウェア（※18）を活用し、利用者の個別データとの安全な統合を実現する、プライバシー保護型ローカルハプロタイプビューアを提供しています。具体的には、JoGo 1.0 において、ACTG 階層命名法に基づく遺伝子型情報を付与した bam 形式（※19）のファイルをユーザーのローカル環境にダウンロードできる機能を実装しました。この機能により、一般公開できない患者のシーケンス情報を利用者のローカル環境に保持したまま、JoGo 1.0 で提供する遺伝子型情報を統合して可視化、分析することが可能となります（図 3）。

JoGo 1.0 は、258 検体の遺伝子型に含まれる 13,613,642 個の一塩基多型、1,669,786 個、1,364,238 個の 50 塩基未満の挿入・欠失配列、332,421 個、69,409 個の 50 塩基以上の挿入・欠失配列を収載しています。これらのすべての情報は JoGo 1.0 の JoGo Genomics > Variant のページから検索を行うことができます。

**図 3 プライバシー保護ローカルハプロタイプビューア
～HBB遺伝子型と個別検体の遺伝子型との比較～**

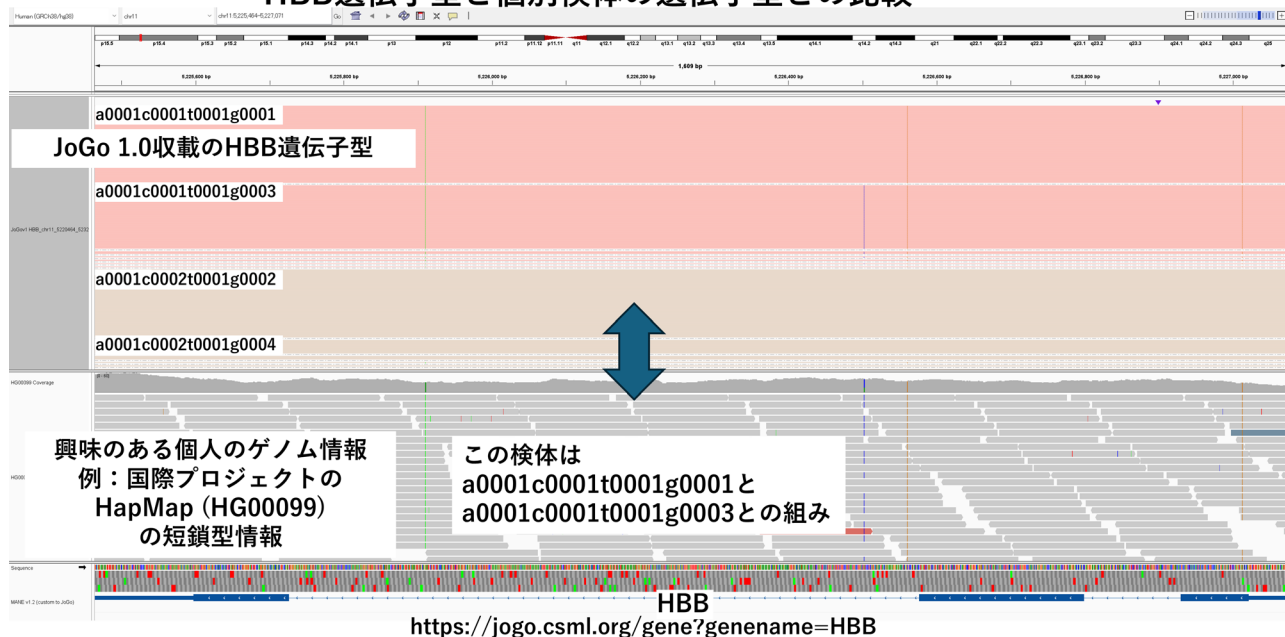
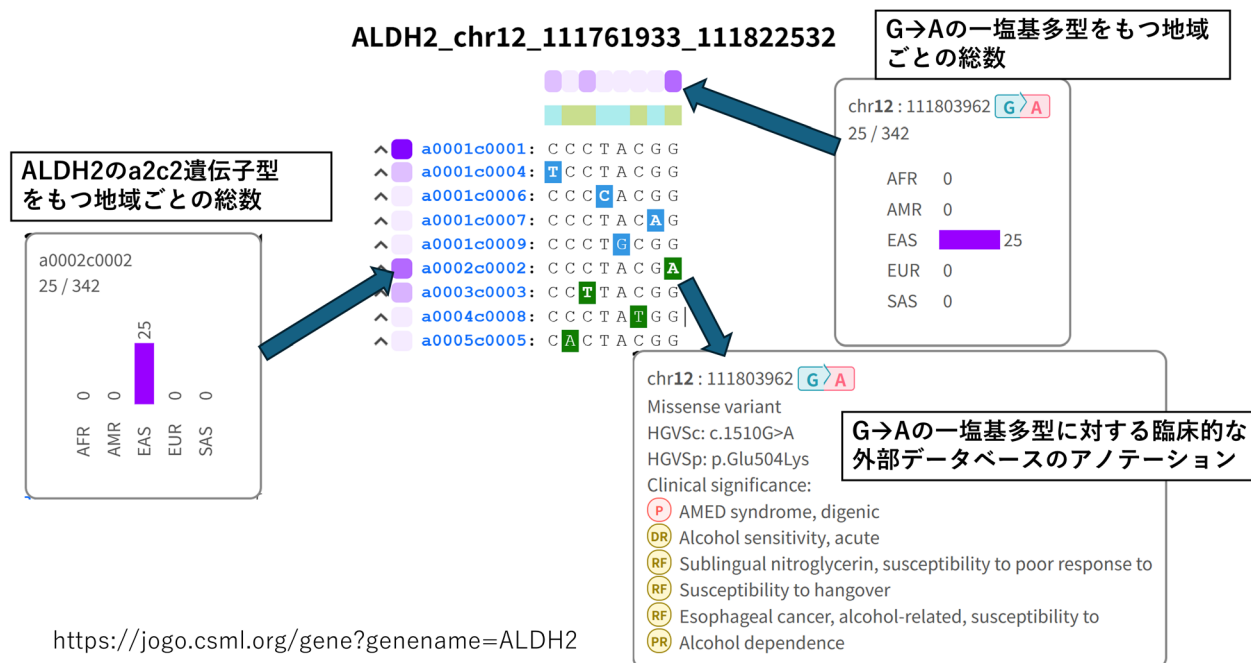


図 3 プライバシー保護型ローカルハプロタイプビューアによる HBB 遺伝子型の比較

JoGo 1.0 が提供する ACTG 情報付き BAM をローカル環境にダウンロードし、IGV 上で「JoGo に収載された HBB 遺伝子型」と「個別検体のゲノムデータ」を安全に統合・比較できる仕組みを示す。個人データを外部に出さずに解析可能。

また、JoGo 1.0 では、収載した遺伝子型の各塩基の集団間の違いに対する国際データベースに基づくアノテーションの拡充にも注力しています。具体的には、臨床情報の関連として、ClinVar、疾患や形質情報との関係として GWAS Catalog、組織ごとの遺伝子発現情報の関連として GTEx などの公共データベースに収載されている変異情報を ACTG 階層の文脈で対応付ける（アノテーション）機能を提供しています（図 4）。

図 4 ALDH2遺伝子型の地域差・関連変異・外部臨床情報へのアクセス



<https://jogo.csml.org/gene?genename=ALDH2>

図 4 ALDH2 遺伝子型の地域差と外部データベースへのアノテーション表示例

ACTG 階層に基づく ALDH2 の a1/a2 ハプロタイプの違いと、AFR/EUR/EAS など世界集団別の頻度を示す。

ClinVar などの外部データベースの臨床的知見にリンクし、ハプロタイプを文脈に統合して参照できる。

(JoGo : <https://jogo.csml.org/gene?genename=ALDH2>)

遺伝子を“型”という共通言語で整理し共有することの有用性を示す例として、JoGo 1.0 では長期間培養できるよう処理された不死化 B 細胞（※20）を用いた 3 つの研究、合計 1,280 検体について、各遺伝子の遺伝型と遺伝子発現量の相関解析（ハプロタイプ QTL（※21））を行っています。結果、3 研究で共通してみられる 65 万件の有意な相関が得られており、これらの情報についても JoGo 1.0 上（JoGo Genomics > HaplotypeQTL）で提供を行っています。

この相関解析の結果をさらに分析することで、どの遺伝子型を持つ場合に発現量が高くなるのか、低くなるか、変わらないかを、ワンストップで評価できるようになっています（図 5）。

図 5 METTL27遺伝子型と遺伝子発現との関連解析の結果

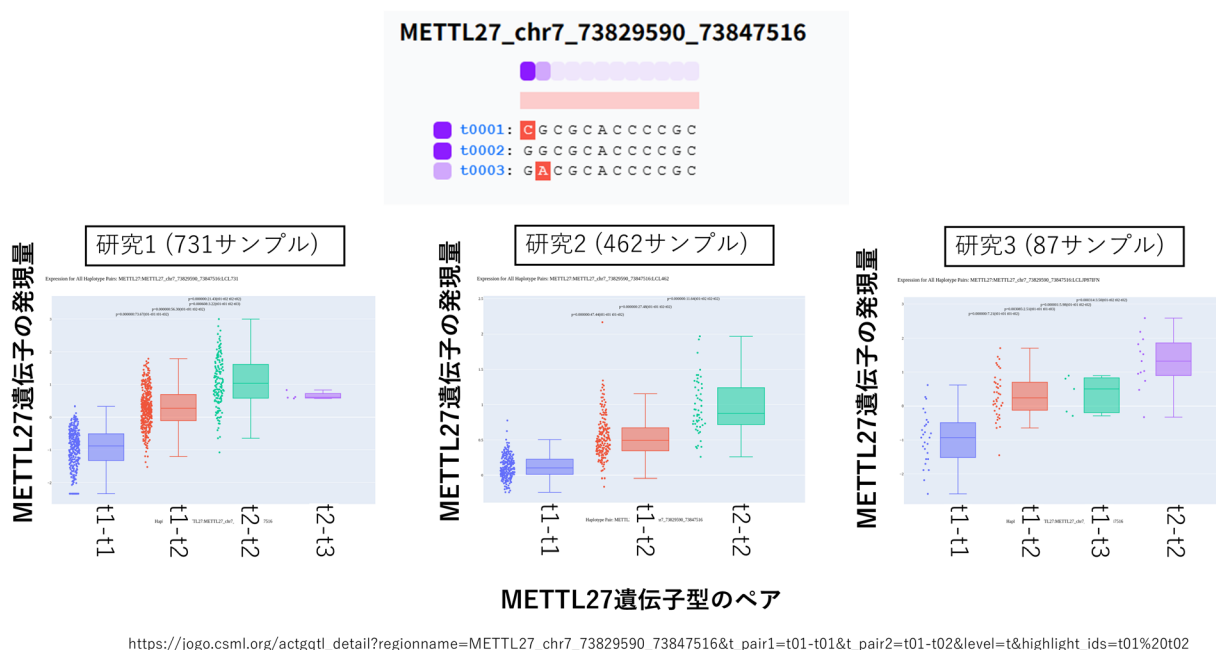


図 5 METTL27 遺伝子型と遺伝子発現量の関連解析（ハプロタイプ QTL）の例

不死化 B 細胞を用いた 3 つの独立研究（731・462・87 検体）における、METTL27 のハプロタイプ別の発現量分布を示す。JoGo 1.0 に収載されたハプロタイプ QTL 情報により、どの遺伝子型で発現が高い／低い／変わらないかを一元的に評価できることを示している。

※METTL27：メチルトランスフェラーゼ関連の機能をもつとされ、RNA・タンパク質修飾に関わる可能性がある。

(https://jogo.csml.org/actgqtl_detail?regionname=METTL27_chr7_73829590_73847516)

【今後の展開】

JoGo 1.0 では、遺伝子を“型”という共通言語で情報整理することの有用性を示す例として不死化 B 細胞に対する遺伝子型と遺伝子発現との関係の情報を整備してきましたが、さらに、臓器ごとの遺伝子発現、また、疾患発症リスクと遺伝子型との関連解析（ハプロタイプ QTL 解析）を拡充することで、ACTG 階層と疾患・形質・発現とのアノテーション拡充を進めていきます。

これにより、免疫系に関わる HLA 遺伝子や薬剤代謝に関わる CYP 遺伝子と同様に、遺伝子型情報を軸とした精密医療や創薬研究への応用を強化し、将来的には、全遺伝子を対象とした「遺伝子型に基づく個人特性の理解と医療応用」を支える基盤を確立することを目指します。

本研究で公開した JoGo 1.0 は初版カタログであり、約 90%の遺伝子の A、C、T 階層で高いカバー率を達成していますが、G 階層についてのカバー率は十分とは言えません。今後も、参照集団の多様性の向上と人数拡大を行うことで、JoGo 2.0 版の構築を進めていく予定です。

【用語解説】

（※1）ハプロタイプ

同じ染色体上で一緒に受け継がれる複数の変化（変異）の組み合わせ。単独の変異よりも、まとまりとして機能に影響することがある。

（※2）ACTG 階層命名法

本研究で確立された新規遺伝子型命名方式。遺伝子配列型を A（アミノ酸）／C（コーディング＝タンパク質領域）／T（転写産物＝UTR 含む）／G（遺伝子本体＝イントロン含む）の階層で表し、頻度順に ID を付ける。

（※3）JoGo（Joint Open Genome and Omics:ジョーゴ）

長鎖型シーケンサの情報を基盤に、人類集団における遺伝子型（ハプロタイプ）をカタログ化した本研究で確立・実装された公開データベース。<https://jogo.csml.org/> からアクセスできる。

（※4）世界集団

アフリカ、ヨーロッパ、東アジア、南アジア、アメリカの 5 大陸の多様な個体群を指す。

（※5）長鎖型シーケンサ

長い DNA 断片（数万塩基）を読み取れる測定装置。短鎖型シーケンサ（※15）に比べ長い DNA 断片を読み取れるため、遺伝子全体の配列型が把握しやすい。

（※6）国際参照配列（リファレンス配列）

国際的に普及しているヒトの代表的な参照配列の名称。いくつかのバージョンがあり、GRCh38 や CHM13v2 などが挙げられる。

（※7）ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>)

臨床的に意味のある変異情報を集めた公開データベース。

（※8）GWAS Catalog (<https://www.ebi.ac.uk/gwas/>)

病気や体質（形質）と関連が報告された変異のカタログ。

（※9）GTEx (<https://gtexportal.org/>)

ヒトの臓器ごとの遺伝子発現と遺伝的違いの関係を集めたプロジェクト／データベース。

（※10）Nucleic Acid Research Breakthrough Article

国際誌 Nucleic Acids Research において、独創性・重要性・科学的完成度が特に高いと編集部が認定した論文にのみ与えられる特別区分です。全掲載論文の上位 2～3% 程度に選ばれる極めて希少なカテゴリであり、研究分野に新たな方向性や大きな進展をもたらす成果として評価されたことを示します。

（※11）ABO 遺伝子

ABO 式血液型は、赤血球表面に存在する糖鎖抗原の違いによって決まる。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付加される。これらの糖を付加する糖転移酵素をコードする ABO 遺伝子は 9 番染色体上にあり、この遺伝子内には A 型を規定する多型、B 型を規定する多型、さらに酵素活性が欠損して糖鎖を付加できない O 型を規定する多型など、複数の変異箇所が存在する。

父母から 1 本ずつ受け継ぐ 2 本の ABO 遺伝子には、それぞれ複数の変異の組み合わせ（ハプロタイプ）が存在し、その 2 本のハプロタイプの組み合わせによって A・B・O の型が決まる。これらの組み合わせとして AA、AO、BB、BO、AB、OO などの遺伝子型が生じ、A 型（AA・AO）、B 型（BB・BO）、AB 型（AB）、O 型（OO）として分類される。

（※12）HLA／KIR

免疫・免疫受容体に関わる代表的な遺伝子群。

（※13）CYP

薬物代謝に関わる代表的な遺伝子群。

（※14）SNP アレイ

あらかじめ決まった位置の塩基配列情報を分離する測定機器

（※15）短鎖型シーケンサ

短い DNA 断片（数百塩基）を高速で読む装置。点状の変異の検出に強いが、遺伝子の全体の配列型の復元が難しい場合がある。

（※16）アセトアルデヒド

アルコール代謝で生じる物質。ALDH2 という酵素で分解される。

（※17）MANE

NCBI と欧州 EMBL-EBI が標準化した遺伝子と標準転写物のセット。

（※18）IGV ソフトウェア（Integrative Genomics Viewer）

ゲノムデータを手元で可視化できるソフト。

（※19）bam 形式（BAM）

塩基配列データの標準的な保存形式。

（※20）不死化 B 細胞

長期間培養できるよう処理された B 細胞のこと。

（※21）QTL／ハプロタイプ QTL

遺伝的な違いと形質（例えば身長や体重）の相関（Quantitative Trait Loci）。ハプロタイプ単位で調べる QTL 解析。

【謝辞】

本研究の JoGo 1.0 の構築は主に、国立研究開発法人科学技術振興機構（JST）ライフサイエンスデータベース統合推進事業（統合化推進プログラム）（JPMJND2302）の支援を受けました。また、本研究に関連する計算の一部のリソースは、学際大規模情報基盤共同利用・共同研究拠点（JHPCN）および高性能計算基盤（HPCI）の支援を一部受けました（プロジェクト ID：jh200047-NWH、jh210018-NWH、jh220014、jh230016、jh240015、jh250010）。とりわけ 深沢 圭一郎、松田 文彦、村田 健史、関谷 勇司、塙 敏博、南里 豪志の各氏に深謝いたします。さらに、本研究に関連する計算の一部のリソースは、九州大学生体防御医学研究所オミクスサイエンスセキュア情報解析システム（OASIS; <https://sis.bioreg.kyushu-u.ac.jp/>）を用いて得られました。

本研究は、他に、国立研究開発法人 日本医療研究開発機構（AMED）の支援（課題番号：JP20ek0109492、JP21wm0425009、JP20ek0109485、JP21ek0109548、JP22fk0210111、JP23fk0210138、JP23ek0210194、JP24gm2010001、JP23ek0109675、JP23ek0109672、JP22tm0424222、JP25fk0310536、JP25fk0310535、JP25wm0625519）、日本学術振興会（JSPS）科研費（JP21H02681）を一部受け、九州大学生体防御医学研究所（MIB）における文部科学省共同利用・共同研究システム形成事業～学際領域展開ハブ形成プログラム～JPMXP1323015486、高深度オミクス医学研究拠点整備事業、および多階層生体防御システム研究拠点の助成を一部受けました。

本 JoGo データベースの一部の情報は、Human Pangenome Reference Consortium（BioProject ID: PRJNA730823）およびその資金提供機関である米国国立ヒトゲノム研究所（National Human Genome Research Institute: NHGRI）に基づいています。

【論文情報】

掲載誌：Nucleic Acids Research

タイトル：JoGo 1.0: the ACTG hierarchical nomenclature and database covering 4.7 million haplotypes across 19,194 human genes（JoGo 1.0：ACTG 階層的命名法と、19,194 ヒト遺伝子にわたる 470 万ハプロタイプを収載するデータベース）

著者名：長崎 正朗^{1,2,*}, 片山 俊明⁵, 守屋 勇樹⁵, Yayoi Sekiya¹, 川島 秀一⁵, 寺岡 凌¹, 町田 宗聡¹, 松原 太一¹, 橋本 洋希¹, 浅倉 章宏¹, 永野 朗夫⁶, 山下 理宇^{8,9}, 高田 豊行¹⁰, 三橋 信孝⁵, 鎌田 真由美⁷, 大川 恭行¹, 徳永 勝士³, 河合 洋介^{3,4}, バリエント情報標準化研究会#

(*:筆頭著者、責任著者)

本論文の共著者として扱うバリエント情報標準化研究会 (<https://dbcls.github.io/visc/>) メンバーは以下です。

長崎 正朗^{1,2}, 片山 俊明⁵, 守屋 勇樹⁵, 川島 秀一⁵, 町田 宗聡¹, 松原 太一¹, 永野 朗夫⁶, 山下 理宇^{8,9}, 高田 豊行¹⁰, 三橋 信孝⁵, 藤澤 貴智⁴, 鎌田 真由美⁷, 佐藤 大輔¹¹, 八谷 剛史¹², 大平 正貴¹³, 黒木 陽子¹⁴, Patricia Severino¹, Samuel Papa Kwesi Owusu¹, Yifan Huangfu¹, 土井 淳¹, 原 雄一郎⁷, 川嶋 実苗⁵, 河合 洋介^{3,4}

1. 九州大学 生体防御医学研究所
2. 京都大学 大学院医学研究科附属ゲノム医学センター
3. 国立健康危機管理研究機構 国立国際医療研究所 ゲノム医科学プロジェクト
4. 情報・システム研究機構 国立遺伝学研究所 生命情報・DDBJ センター
5. 情報・システム研究機構 データサイエンス共同利用基盤施設 ライフサイエンス統合データベースセンター
6. 株式会社 PENQ

7. 北里大学 未来工学部 データサイエンス学科
8. 国立がん研究センター 先端医療開発センター トランスレーショナルインフォマティクス分野
9. 東京大学大学院 新領域創成科学研究科 メディカル情報生命専攻
10. 理化学研究所 バイオリソース研究センター 統合情報開発室
11. 株式会社カモノハシ
12. 株式会社ゲノムアナリティクスジャパン
13. 東京大学 大学院新領域創成科学研究科 先端生命科学専攻
14. 国立成育医療研究センター ゲノム医療研究部

DOI : [10.1093/nar/gkaf1232](https://doi.org/10.1093/nar/gkaf1232)

【お問合せ先】

<研究に関すること>

九州大学 生体防御医学研究所 教授 長崎 正朗（ナガサキ マサオ）

TEL : 092-642-4815

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

<報道に関すること>

九州大学 広報課

TEL : 092-802-2130 FAX : 092-802-2139

Mail : koho@jimu.kyushu-u.ac.jp

国立健康危機管理研究機構 危機管理・運営局 広報管理部

TEL : 03-3202-7181

Mail : press@jihs.go.jp

情報・システム研究機構 データサイエンス共同利用基盤施設

ライフサイエンス統合データベースセンター 広報担当

TEL : 04-7135-5508

Mail : public_relations@dbcls.rois.ac.jp

北里研究所 広報室

TEL : 03-5791-6422 FAX : 03-3444-2530

Mail : kohoh@kitasato-u.ac.jp

理化学研究所 広報部 報道担当

TEL : 050-3495-0247

Mail : ex-press@ml.riken.jp

科学技術振興機構 広報課

TEL : 03-5214-8404 FAX : 03-5214-8432

Mail : jstkoho@jst.go.jp

<JST 事業に関すること>

科学技術振興機構 情報基盤事業部 NBDC 事業推進室

川口 貴史（カワグチ タカフミ）

TEL：03-5214-8491

Mail：nbdc-funding@jst.go.jp

Kyushu
University **VISION 2030**
総合知で社会変革を牽引する大学へ