

JoGo-Next：国際標準に向けたヒトハプロタイプデータベース構築とAI活用

研究代表者　九州大学　長﨑 正朗

分担研究者　国立健康危機管理研究機構　河合 洋介

分担研究者　国立成育医療研究センター　黒木 陽子



トーゴーの日シンポジウム2026

品川ザ・グランドホール

2026年10月5日（月） 14:25-14:35



①研究開発の構想（背景）

2023–2025年度　DICP開発課題（当時）
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

【目標】塩基配列を公開可能な100検体の日本人のゲノム・オミクス情報に加え、ヒトゲノム長鎖型の構造多型情報解析規格の統一を進めているCoLoRSとの国際連携を通じ国外の約2,000検体の統計情報を統合する。また、得られた構造多型はVISCで定められるRDFを活用予定である。これにより、希少疾患から多因子疾患まで構造多型の理解に役立つ日本人ゲノム情報基盤の役割を果たす。

JoGo1.0ポータルの開発と一般公開

ACTG-Haplotype 命名法の開発

https://jogo.csml.org/　2025.8一般公開

➡**世界初のコード遺伝子のハプロタイプ（型）の統一記法の開発と辞書化**

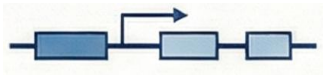
➡訪問者は約2,800人（8月）→6,700人（9月）→55,000人（10月）、230,000（11月）、**ページビューは、115万→273万→ 849万件と急増**し、その後も同様のアクセスが続いている。※AIによる情報登録のためのアクセスも含む

1. ﾋﾄﾀﾝﾊﾟｸ質ｺｰﾄﾞ領域の収載



約2万遺伝子を辞書化

2. 遺伝子毎の構造を対象



460万ﾊﾌﾟﾛﾀｲﾌﾟを収載

3.高精度長鎖型参照ﾃﾞｰﾀ



国内105検体含む5大陸258検体

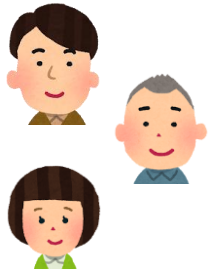
※40万以上の50塩基以上の挿入欠失含む

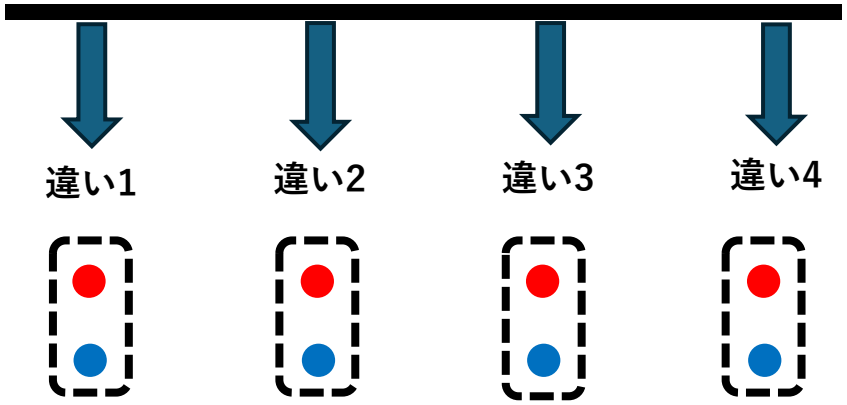
NAR DB Issue 2026.1掲載（雑誌社により**上位1%以内が選ばれる**Breakthrough Articleの称号付与を受けた）

①研究開発の構想（背景）

ハプロタイプ（型）とは何なのか？

遺伝子X





点

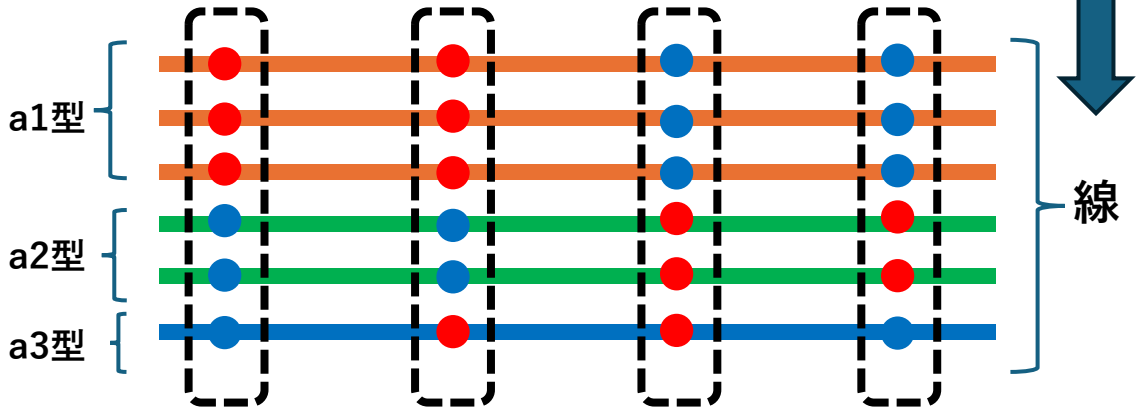
2^3 = 8パターン？

※2本の染色体の入替えを同一視するため、n個のヘテロ接合部位の位相パターンは2^(n−1)通り

従来

遺伝子の違い1-4はどのようなパターン（ハプロタイプ＝型）か？

提案



線

実際は3パターン

10SNPあれば
2^9=512

20SNPあれば
2^19=524,288

ACTG-Haplotype Notation

達成

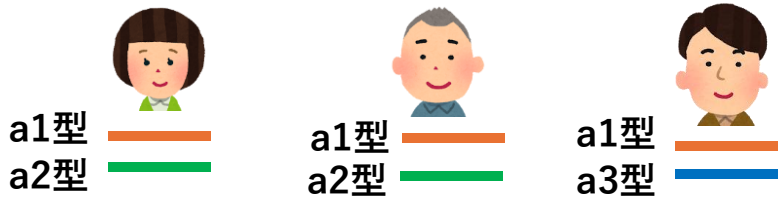
ACTG命名法による約2万ヒト遺伝子の460万の遺伝子型辞書の整備
JoGo1.0（https://jogo.csml.org）での公開

JoGo1.0

実際はそこまで多くない

タンパクレベルだと各遺伝子**通常5-15パターンでヒト集団の99%をカバー可**

今後



辞書化することの意義

遺伝子の個人差は「点」から「線」の時代へ、医療・創薬の活用へ

Nagasaki *et al. NAR* 2026

① 研究開発構想（概要・開発項目・収載予定のデータ）

開発・公開済

【題目】JoGo-Next：国際標準に向けたヒトハプロタイプデータベース構築とAI活用

【目標】申請者は、世界で唯一の2万のヒトタンパク質コード遺伝子を対象とした460万種からなるハプロタイプ辞書を構築2025年8月にJoGo1.0として公開をした。DBの国際誌NAR DB Issue 2026においてBreakthrough Articleとしてtop1%のDBとしての称号が付与された。**本研究提案「JoGo-Next」では、非コード遺伝子・制御領域を対象とし、さらに対象サンプル数を大幅に拡大することで、唯一のヒトハプロタイプ辞書としてのde facto standardとして国際的な地位を確立する。**

JoGo-Nextの新機軸の６つの【開発項目】と【収載予定のデータ】

本研究期間の開発内容

1. ﾋﾄ非ｺｰﾄﾞ・制御領域の収載



3.5万lncRNA・10万のﾌﾟﾛﾓｰﾀ・210万ｴﾝﾊﾝｻを辞書化

2.複雑な遺伝子群構造の収載



KIR/LILR/RCCX領域※1などの免疫関連複合領域を体系的に辞書化

3. 参照人数とT2Tｹﾞﾉﾑ※2の拡大

韓国100人
中国500人
他



※2 T2T=ﾃﾛﾒｱからﾃﾛﾒｱまで

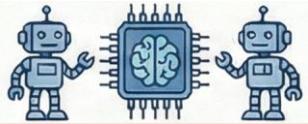
国内105検体含む500～1000人規模のT2T/長鎖ｼｰｸｴﾝｽ情報に基づく

4. 短鎖WGSからの型推定と公開



3千人の個人型情報（海外）
1万人の型頻度（国内）

5. AIｴｰｼﾞｪﾝﾄによる半自動化



LLM活用 AIｴｰｼﾞｪﾝﾄでDB更新・解析・注釈を半自動化

6. 国際連携と標準化の取り組み

HUMAN PANGENOME PROJECT

国際ﾊﾟﾝｹﾞﾉﾑﾌﾟﾛｼﾞｪｸﾄや国際免疫学会等と連携 ➡ de facto standardへ

※1 RCCX領域=STK19-C4A-C4B-CYP21A2-TNXB

将来展望：ｹﾞﾉﾑｻｲｴﾝｽのﾊﾌﾟﾛﾀｲﾌﾟ言語のde facto standardとしての地位を確立

研究者に加えAIも活用できる「ｹﾞﾉﾑﾊﾌﾟﾛﾀｲﾌﾟ言語」数百万～数千万ｽｹｰﾙのﾊﾌﾟﾛﾀｲﾌﾟ情報が創薬や診断の仮説生成を加速

創薬・医療現場への貢献
疾患ﾘｽｸや個人差をより正確に解釈

③ 研究開発計画（実施計画） JoGo-NextにおけるACTG-ハプロタイプ記法について

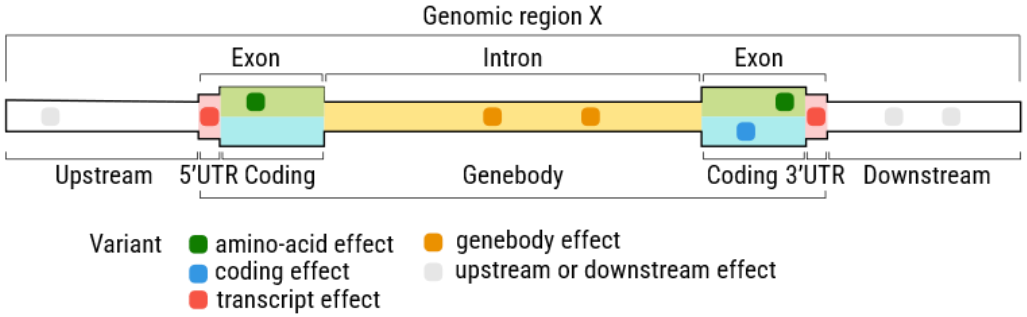
JoGo1.0

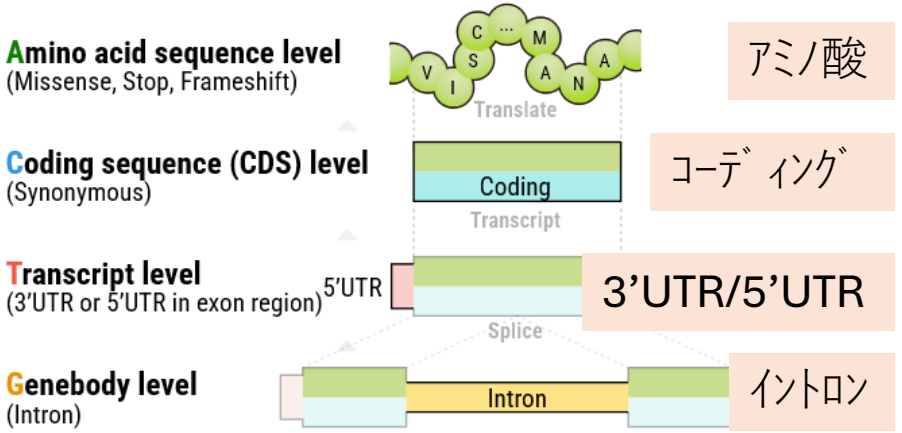
数字はサンプルを行った世界集団の多い順に1から並ぶ

コード遺伝子

HBB: a1 c1 t1 g1

NM_000518: a1 c1 t1 g1





JoGo-Nextで新たに対象とする領域の記法

Nagasaki *et al. NAR* 2026論文中において
非コード遺伝子/制御領域の記載法は定義済

非コード遺伝子

Gencode(v49:2026.1)
3.5万lncRNA対象

1. intronがないlncRNA ※短縮記法(=フル記法)

NR_002819 : t1 g0 (=NR_002819:a0c0t1g0)

2. intronがあるlncRNA

NR_144568 : t1 g1 (=NR_144568:a0c0t1g1)

制御領域（エンハンサ・プロモータ）

1. Encodeの場合

EH38E3212459 : r1

2. fanta.bioの場合

FCHS_26 : r1

3. 指定領域の場合

GRCh38_chr17_17973344−17973515 : r2

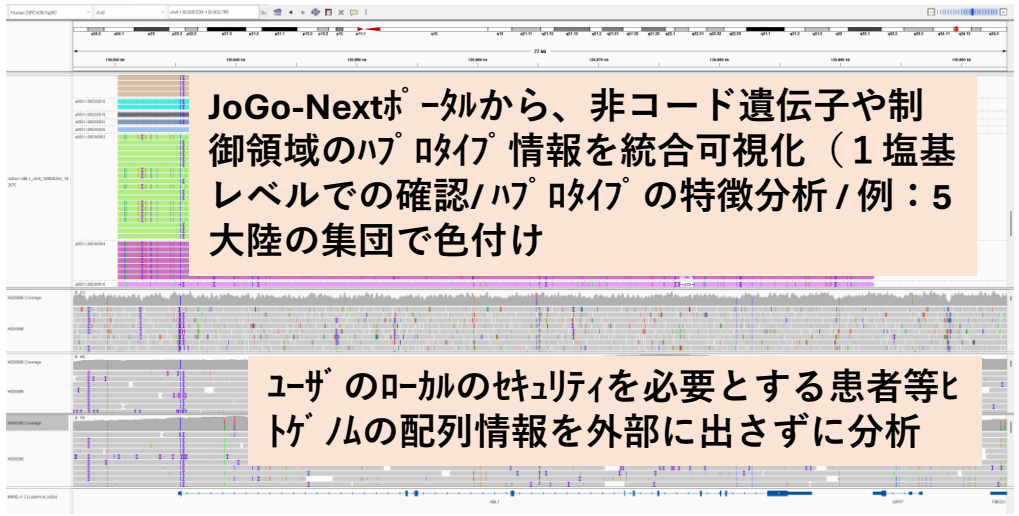
Encode及び
fanta.bioの
10万のﾌﾟﾛﾓｰﾀ・
210万ｴﾝﾊﾝｻ対象

③ 研究開発計画（実施計画）　JoGo-Nextにおける可視化とアノテーションについて

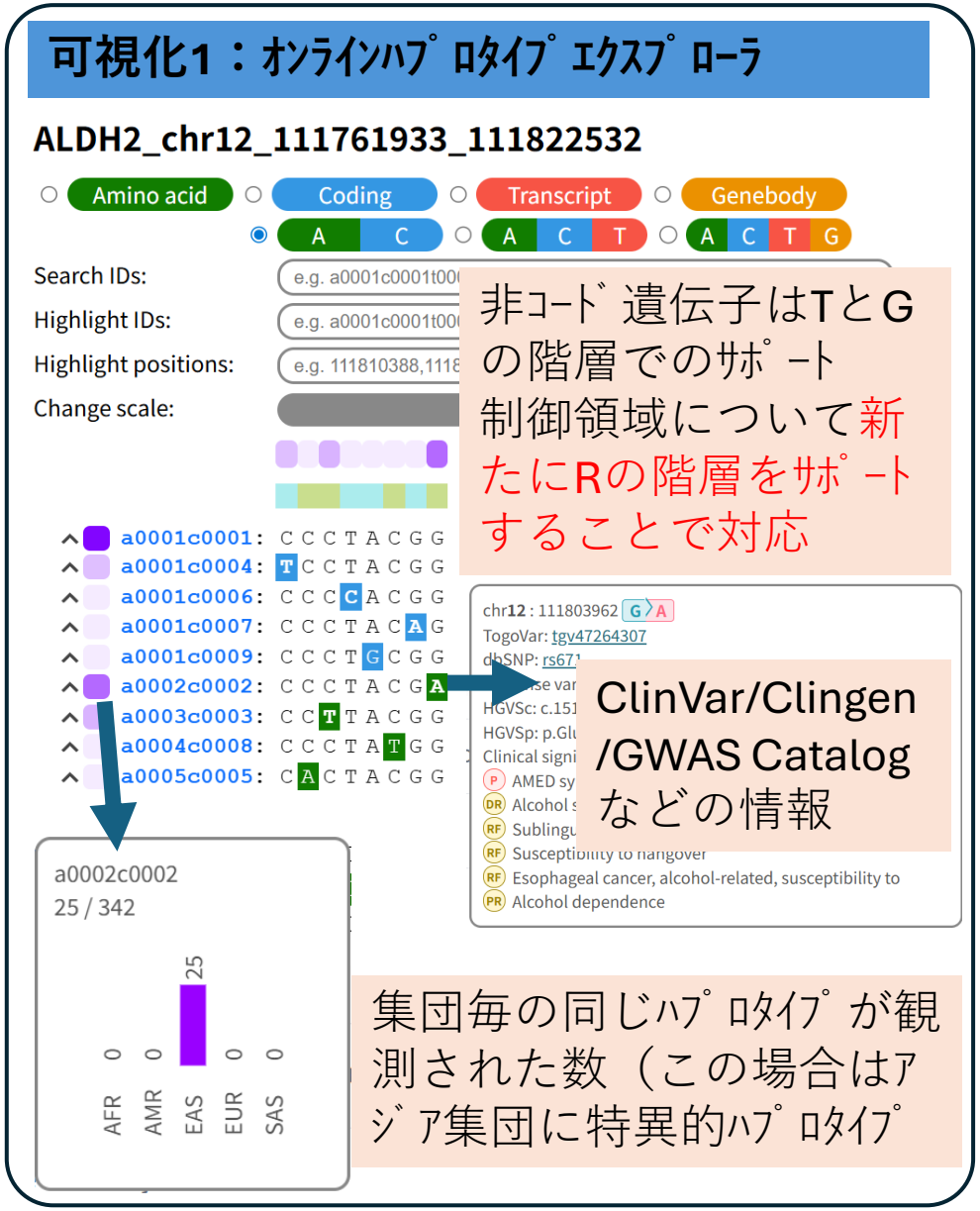
「JoGo1.0」
（可視化1）ウェブポータル上での可視化ツール（右図→）
（可視化2）ローカルシステム上での可視化ツール（下図↓）
を開発し2万コード遺伝子460万ハプロタイプを情報提供
さらに、変異とハプロタイプの間の外部データベースとのアノテーションを行うことができる（右図→）

「JoGo-Next」
この可視化ツールを活用しつつ、非コード遺伝子、制御領域に対応することで情報整備・公開を進める。

可視化2：ローカルハプロタイプエクスプローラ



可視化1：オンラインハプロタイプエクスプローラ



③研究開発計画（6つの実施計画と研究開発体制）

体制図

①研究代表者
長﨑・九大



・研究統括
・項目1-5の主開発担当
・標準化取り組み

組織メンバ

国際免疫学会年会 組織代表

6. 国際連携と標準化の取り組み

国際ﾊﾟﾝｹﾞﾉﾑﾌﾟﾛｼﾞｪｸﾄや国際免疫学会等と連携 ➡ de facto standardへ

The 19th International HLA & Immunogenetics Workshop
May 19–24, 2026 | Numazu, Japan

参加

HUMAN PANGENOME PROJECT
Asian Pangenome Consortium

2026/5/19-26 IHIWS2026
2026/6/9-12 TBRN2026

・標準化取組支援
GA4GH / RDF支援

参加

参加



研究協力者 特任教授 片山 俊明

②研究分担者
河合・JIHS



・T2T情報解析
・105検体の日本人T2T検体
項目6の主担当

・T2Tでの国際連携
・標準化取り組み

#JaSaPaGe
日本人10検体のT2T構築

情報解析基盤
OASIS (九大)



提供

3. 参照人数とT2Tｹﾞﾉﾑの拡大

(3-1) T2Tの国内外検体の整備・解析

(3-2) 長鎖型の国内外検体の整備・解析

1. ﾋﾄ非ｺｰﾄﾞ・制御領域の解析・収載

(1-1) 3.5万lncRNA辞書

(1-2) 10万のﾌﾟﾛﾓｰﾀ・210万ｴﾝﾊﾝｻ辞書

4. 短鎖WGSからの型推定と公開

(4-1) ﾊﾌﾟﾛﾀｲﾌﾟ推定技術開発

適用

(4-2) 3千人の個人型情報（海外）

適用

(4-3) 1万人の型頻度（国内）



National Center Biobank Network

中央バイオバンク長
徳永 勝士

研究協力者

提供

中央ﾊﾞｲｵﾊﾞﾝｸの1万人の短鎖型ｼｰｸｴﾝｽ情報提供

5. AIｴｰｼﾞｪﾝﾄによる半自動化

(5-1) AI解析ﾜｰｸﾌﾛｰ構築

(5-2) AIｱﾉﾃｰｼｮﾝﾜｰｸﾌﾛｰ構築

性能評価
ﾌｨｰﾄﾞﾊﾞｯｸ

専門家

③研究分担者
黒木・NCCHD
ゲノム医療研究部

項目2のRCCX部分、項目5の評価担当

・希少疾患ｹﾞﾉﾑ診断での実績と活用

・希少疾患発症者における*denovo*構造の同定

LILR専門家

2.複雑な遺伝子群構造の収載

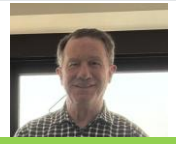
(2-1) LILR領域解読・辞書化

(2-2) RCCX領域※解読・辞書化

(2-3) KIR領域解読・辞書化

RCCX専門家





Fred Hutch Cancer Center

教授
Daniel Geraghty

研究協力者

KIR専門家

KIR情報提供

2,504人の独自
KIR配列情報の提供

※RCCX領域=STK19-C4A-C4B-CYP21A2-TNXB

④ 利用者像

ヒト

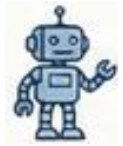


ーヒトの変異とハプロタイプの関係、ハプロタイプに関する知識ドメインに関する情報取得。また、そのハプロタイプを得るためにつかった元のデータの情報取得。
ー中学、高校、大学、大学院の教育、研究、医療においてもハプロタイプ辞書として使うことを想定。

【使用者の主な研究領域としては以下を想定】
ヒトを主に対象としたゲノムサイエンス研究者全般、希少疾患、多因子疾患、単一遺伝子疾患など幅広い疾患に関連する研究者、バイオインフォマティクス、システム生物学、集団遺伝学など

副作用診断キット(NUDT15)やリスク診断キット(APOE)とJoGoハプロタイプ分析の例は参考資料p25-p28参照

AI



赤字の部分がJoGo1.0に加えてJoGo-Nextを開発することで追加できる情報

ChatGPT、Claude、Geminiとの連携、また、これらAI情報に取り込まれることでAIの中でハプロタイプ辞書の単語やアノテーションが活用される。local LLMの機能拡充にともないlocalでの利用も視野に入る。

【例1】MCM6のintronにあるrs4988235がLCT（ラクターゼ）の転写制御にエンハンサとして関連することが知られています。このrs4988235をもつハプロタイプ型はJoGo-Nextでは、r2型で、MCM6遺伝子のa1:c1:t1型にみられ、その約半数がこのパターンを持ちます。

【例2】APOE遺伝子は抗アミロイドベータ（Aβ）抗体薬投与の副作用がハプロタイプ型によって違うことが知られています。rs429358がリスクになることが知られていますが、このハプロタイプ型はa02です。さらに、このa02型は、JoGo-Nextでは、EH9999:r02の場合、発現が高値になることがわかっています。EH999:r02近傍のAPOC1:a1c1とAPOC4:a2c1の組み合わせが日本では80%、ヨーロッパでは50%、アフリカでは確認されません（解析が進むことでこういうアノテーションが可能となる）。

⑤ 学術的・社会的に期待されるインパクト

「ヒトハプロタイプ辞書によるヒトとAIへの網羅的な情報基盤」としての貢献

ヒトとAIの双方にとって、膨大な**塩基配列の差分**を意味のある**機能単位（ﾊﾌﾟﾛﾀｲﾌﾟ）**へ縮約できることは、**新しい“言語”を獲得することに等しい。これにより、**①タンパク質全長の配列差に基づく機能差に加え、②ｴﾝﾊﾝｻ/ﾌﾟﾛﾓｰﾀ等の調節要素の組み合わせに基づく発現差、③遺伝子本体と調節領域の組み合わせに基づく表現型差といった、従来扱いにくかったﾊﾌﾟﾛﾀｲﾌﾟ単位の関係性**を、ｹﾞﾉﾑ･ｵﾐｸｽ測定情報の統合から導出する基盤となる。

JoGo1.0で構築したハプロタイプ辞書の対象領域

Exon　Intron

約２万

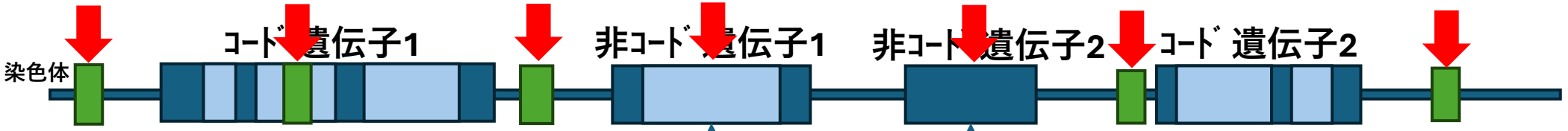


JoGo-Nextで構築予定のハプロタイプ辞書の新規対象領域

制御領域のﾊﾌﾟﾛﾀｲﾌﾟ構造を整備する意義は参考資料p29-p31に記載

約200万

制御領域 (ｴﾝﾊﾝｻ・ﾌﾟﾛﾓｰﾀ等) (ENCODE ID:EH38E3212459 / fanta.bio ID: FCHS_26)に基づく



E1:r01
E1:r02
E1:r03
E1:r04

E2:r01
E2:r02
E2:r03

E3:r01
E3:r02

NR_144568:a0c0t1g1

NR_002819:a0c0t1g0

E4:r01
E4:r02
E4:r03
E4:r04
E4:r05

E5:r01
E5:r02
E5:r03

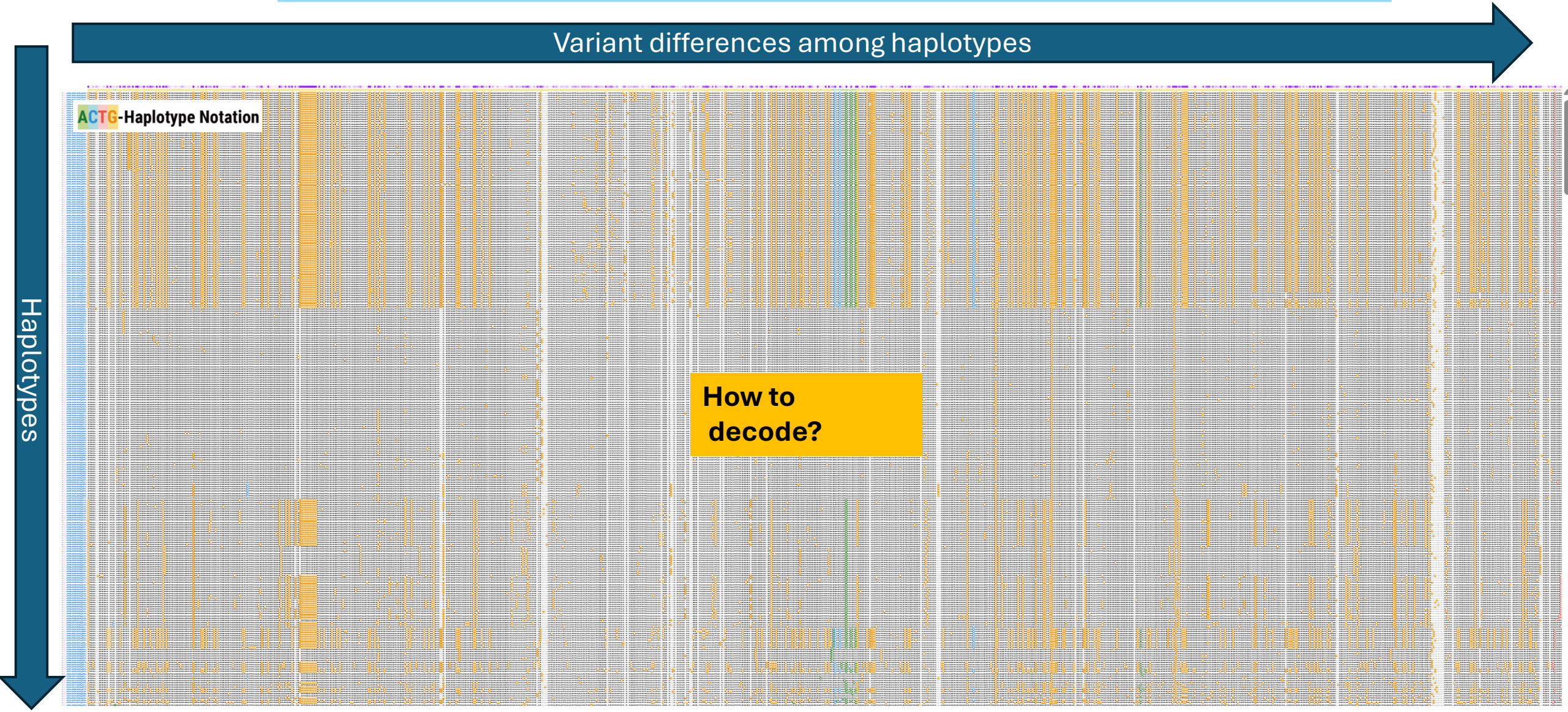
Encode DB
fanta.bio
等との連携

#前述のMCM6の例はintronがenhancerとなるこの例

Application 2

JoGo 1.0: BRCA1 ACTG-Haplotype Diversity Map

Overview of full ACTG-haplotype diversity of BRCA1 (breast cancer related)



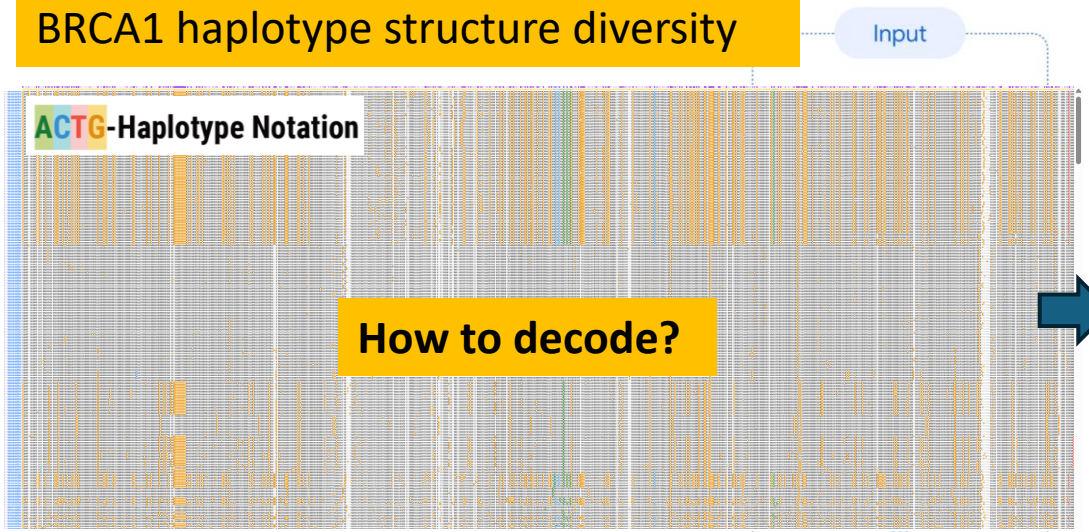
https://jogo.csml.org/gene?genename=BRCA1

想定される活用事例

Decoding Haplotype Diversity Across ACTG Levels into Molecular Functional Effects

multi-level molecular AI: sequence, RNA, protein structure, regulation and quantum ML

BRCA1 haplotype structure diversity

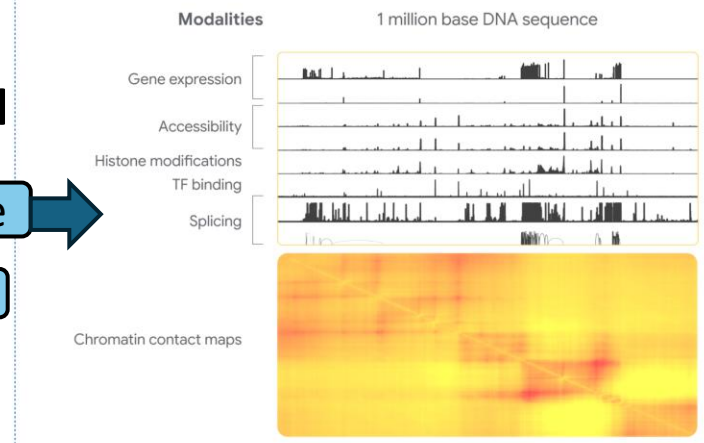


ACTG-level

AlphaGenome

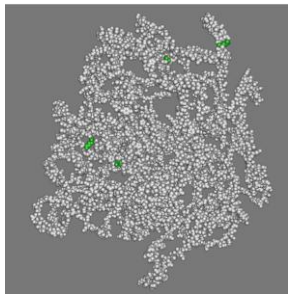
QTFPred

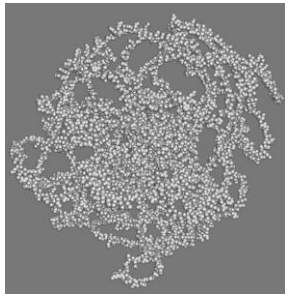
Prediction of functional diversity



A-level

AlphaFold3





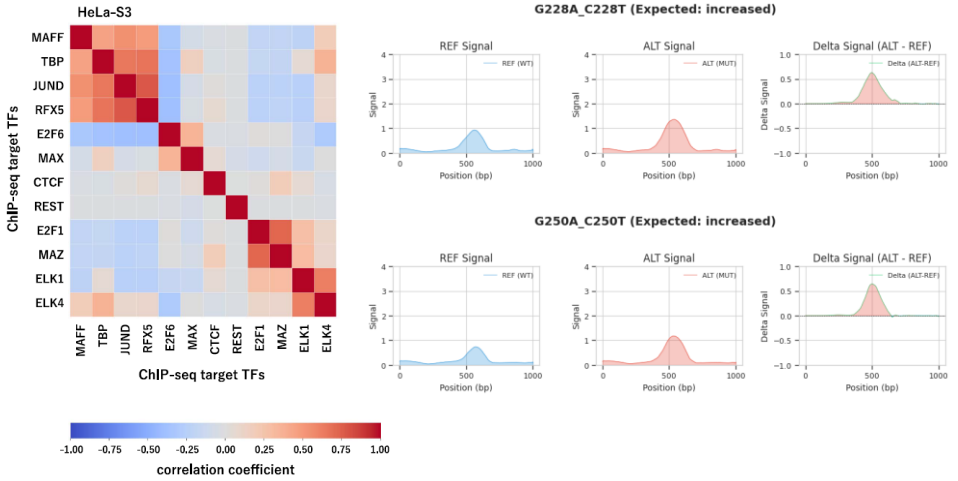
ACT-level

mRNABERT

UTR-LM

CodonBERT

3UTRBERT



疾患・形質・薬剤応答の解釈を高解像度化し、精密医療・創薬研究を支える基盤へ

Advancing regulatory variant effect prediction with AlphaGenome Avsec *et al* 2026 Nature

QTFPred: robust high-performance quantum machine learning modeling that predicts main and cooperative transcription factor bindings with base resolution Matsubara *et al* 2025

まとめ 研究開発構想（概要・開発項目・収載予定のデータ）再掲

開発・公開済

【題目】JoGo-Next：国際標準に向けたヒトハプロタイプデータベース構築とAI活用

【目標】申請者は、世界で唯一の2万のヒトコーディング遺伝子を対象とした460万からなるハプロタイプ辞書を構築2025年8月にJoGo1.0として公開をした。DBの国際誌NAR DB Issue 2026においてBreakthrough Articleとしてtop1%のDBとしての称号が付与された。**本研究提案「JoGo-Next」では、非コード遺伝子・制御領域を対象とし、さらに対象サンプル数を大幅に拡大することで、唯一のヒトハプロタイプ辞書としてのde facto standardとして国際的な地位を確立する。**

JoGo-Nextの新機軸の６つの【開発項目】と【収載予定のデータ】

本研究期間の開発内容

1. ﾋﾄ非ｺｰﾄﾞ・制御領域の収載



3.5万lncRNA・10万のﾌﾟﾛﾓｰﾀ・210万ｴﾝﾊﾝｻを辞書化

2.複雑な遺伝子群構造の収載



KIR/LILR/RCCX領域などの免疫関連複合領域を体系的に辞書化

3. 参照人数とT2Tｹﾞﾉﾑ(※)の拡大

韓国100人
中国500人
他





※T2T=ﾃﾛﾒｱからﾃﾛﾒｱまで

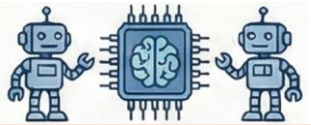
国内105検体含む500～1000人規模のT2T/長鎖ｼｰｸｴﾝｽ情報に基づく

4. 短鎖WGSからの型推定と公開



3千人の個人型情報（海外）
1万人の型頻度（国内）

5. AIｴｰｼﾞｪﾝﾄによる半自動化



LLM活用 AIｴｰｼﾞｪﾝﾄでDB更新・解析・注釈を半自動化

6. 国際連携と標準化の取り組み

HUMAN PANGENOME PROJECT



国際ﾊﾟﾝｹﾞﾉﾑﾌﾟﾛｼﾞｪｸﾄや国際免疫学会等と連携 ➡ de facto standardへ

※RCCX領域=STK19-C4A-C4B-CYP21A2-TNXB

将来展望：ｹﾞﾉﾑｻｲｴﾝｽのﾊﾌﾟﾛﾀｲﾌﾟ言語のde facto standardとしての地位を確立

研究者に加えAIも活用できる「ｹﾞﾉﾑﾊﾌﾟﾛﾀｲﾌﾟ言語」数百万～数千万ｽｹｰﾙのﾊﾌﾟﾛﾀｲﾌﾟ情報が創薬や診断の仮説生成を加速

創薬・医療現場への貢献
疾患ﾘｽｸや個人差をより正確に解釈

おしまい