



非コード領域を対象としたハプロタイプカタログ構築と異なるクラウド拠点間での横断的バッチジョブシステム実装

【課題番号】jh260014

【研究代表者】長崎 正明¹ 【共同研究者】大川 恭行²、南里 豪志³、塙 敏博⁴、関谷 勇司⁵、深沢 圭一郎⁶、松田 文彦⁷、村田 健史⁸、竹房 あつ子⁹、合田 憲人⁹、大江 和一¹⁰、丹生 智也¹⁰、土井 淳¹、関谷 弥生¹、浅倉 章宏¹、男澤 良子¹、寺岡 凌¹、橋本 洋希¹、町田 宗聡¹、松原 太一¹、南 めぐみ¹、星 甫逸凡¹、陳 健東¹、Samuel Papa Kwesi Owusu¹、唐 悦儀¹

¹九州大学 生体防御医学研究所 バイオメディカル情報解析分野 ²九州大学 生体防御医学研究所 トランスクリプトミクス分野 ³九州大学 情報基盤研究開発センター ⁴東京大学 情報基盤センター ⁵東京大学 大学院情報理工学系研究科 ⁶総合地球環境学研究所 基盤研究部 ⁷京都大学 大学院医学研究科 情報通信研究機構 総合テストベッド研究開発推進センター ⁸国立情報学研究所 アーキテクチャ科学系 ⁹国立情報学研究所 クラウド基盤研究開発センター

研究目的

これまで一般公開可能な国内外258検体の中・高深度長鎖型ヒトシーケンス情報を起点として、大規模情報解析資源により、世界で初めてのヒト遺伝子を網羅する、1.9万遺伝子、総計460万のACTGハプロタイプ辞書を構築するとともに、JoGoポータルサイト(https://jogo.cslm.org/)において一般公開を行った。

本研究課題では、解析対象を遺伝子領域から非コード領域へ拡張し、国際プロジェクトENCODEに登録されたプロモーターやエンハンサー等の制御領域について、国内外拠点における高精度のハプロタイプ辞書を構築することを目的とする。

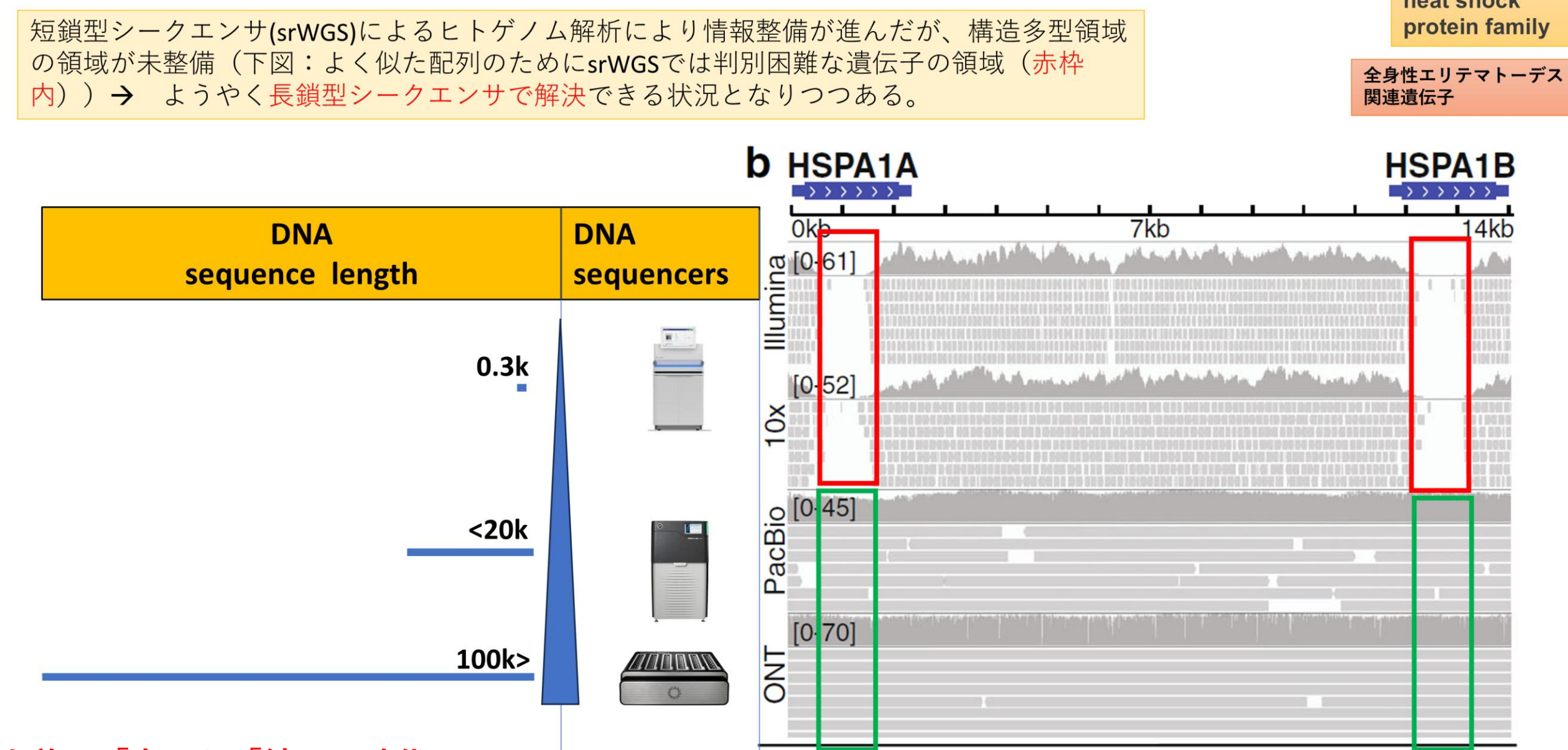
また、本解析で扱う公開可能なヒトゲノム情報を対象に、CPU・GPUおよび大規模ストレージを備える各大学の計算資源、mdx、パブリッククラウドを連携させる。令和6・7年度に進めた試験実装を踏まえ、国立情報学研究所の学認クラウドオンデマンド構築サービスで提供されるソフトウェア群を活用し、九州大学のオンプレミス環境(OASIS)を管理基盤として、複数クラウド上の計算資源とストレージを横断的かつシームレスに利用できる複数拠点のバッチジョブ実行基盤の構築を進める。

これらの辞書が拡充されることで、遺伝子本体に加えて発現制御に関わる領域の集団差を明らかにし、日本人集団の特性や疾患研究に資する制御領域のハプロタイプパターンを理解を深化させる。あわせて、大規模ゲノム解析を複数拠点・複数クラウドで効率的かつ再現可能に実行するための、ゲノムサイエンス解析環境のリファレンス実装をめざす。

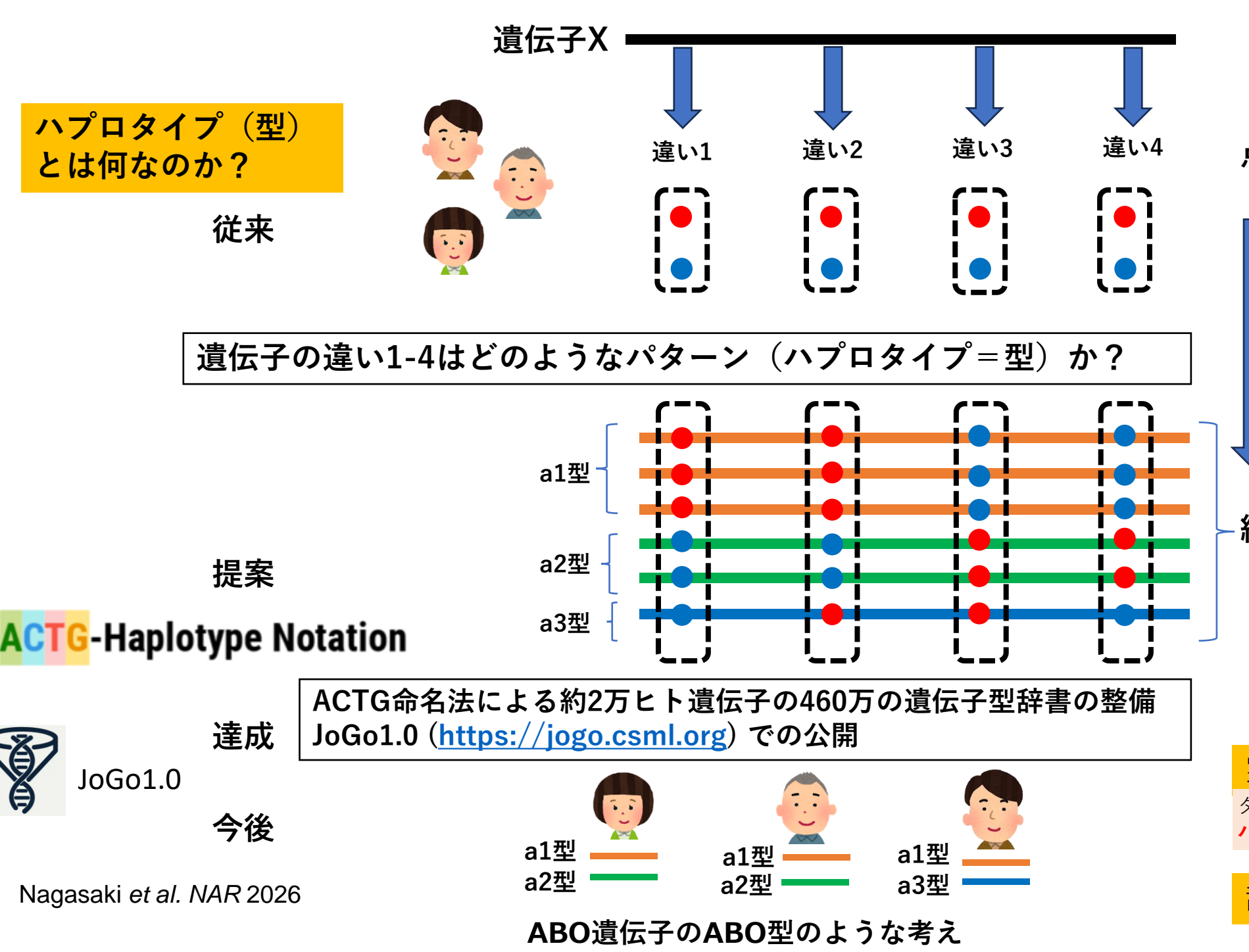
本研究課題の拠点とメンバ構成

- 【九州大学】
 - 生体防御医学研究所
 - 研究課題代表者 長崎 正明
 - 副代表者 大川 恭行
 - 他メンバ 土井 淳、関谷 弥生、浅倉 章宏、男澤 良子、寺岡 凌、橋本 洋希、町田 宗聡、松原 太一、南 めぐみ、Yifan Huangfu, Jiandong Chen, Samuel Owusu, Yueyi Tang
- 情報基盤研究開発センター 南里 豪志
- 【東京大学】
 - 情報基盤センター 塙 敏博
 - 大学院情報理工学系研究科 関谷 勇司
- 【京都大学】
 - ゲノム医学センター 松田 文彦
- 【国立情報学研究所】
 - アーキテクチャ科学系 竹房 あつ子 合田 憲人
 - クラウド基盤研究開発センター 大江 和一
 - 生命情報・DBJ研究センター 丹生 智也
- 【情報通信研究機構】
 - 総合テストベッド研究開発推進センター 村田 健史
- 【総合地球環境学研究所】
 - 基盤研究部 深沢 圭一郎

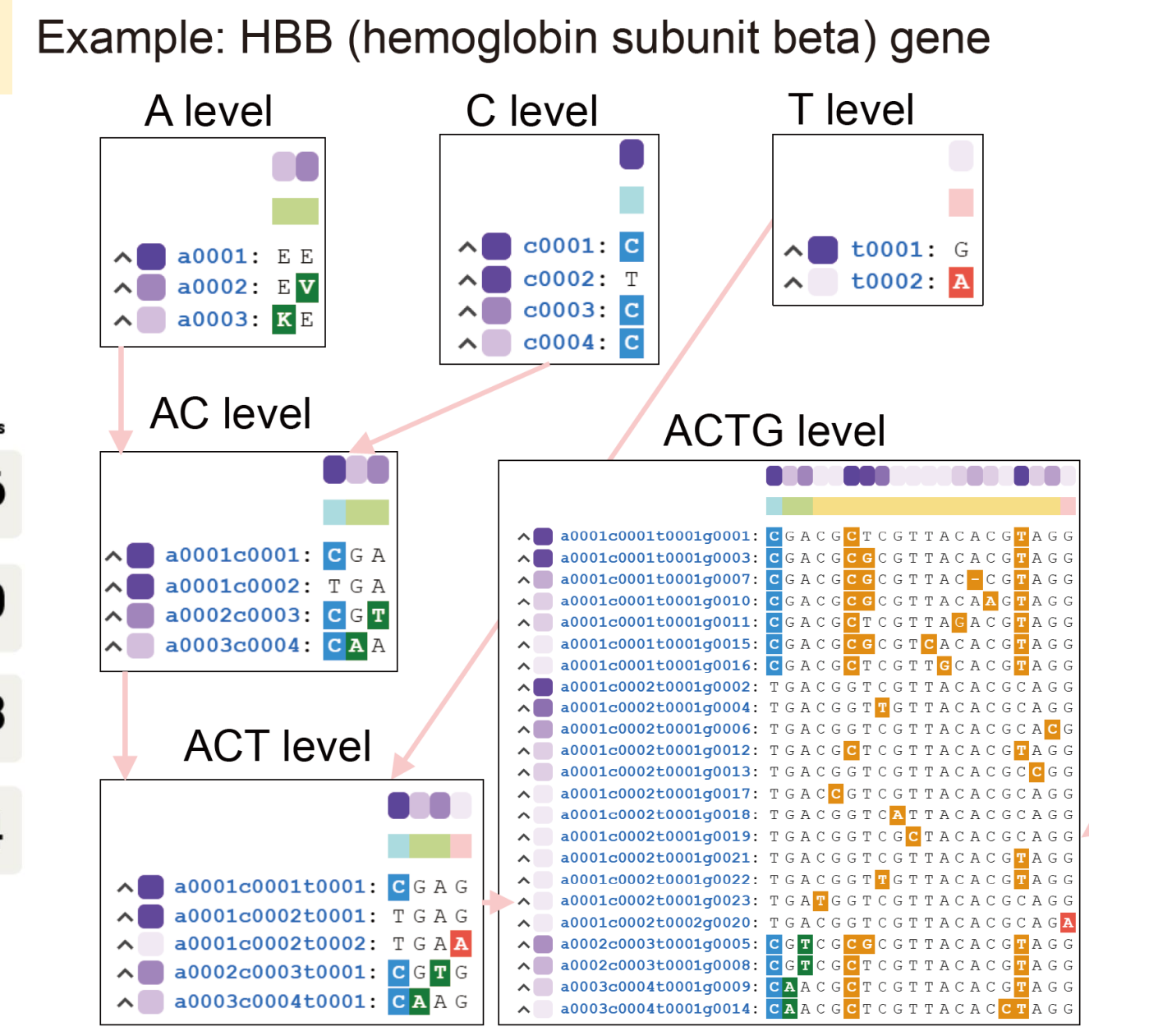
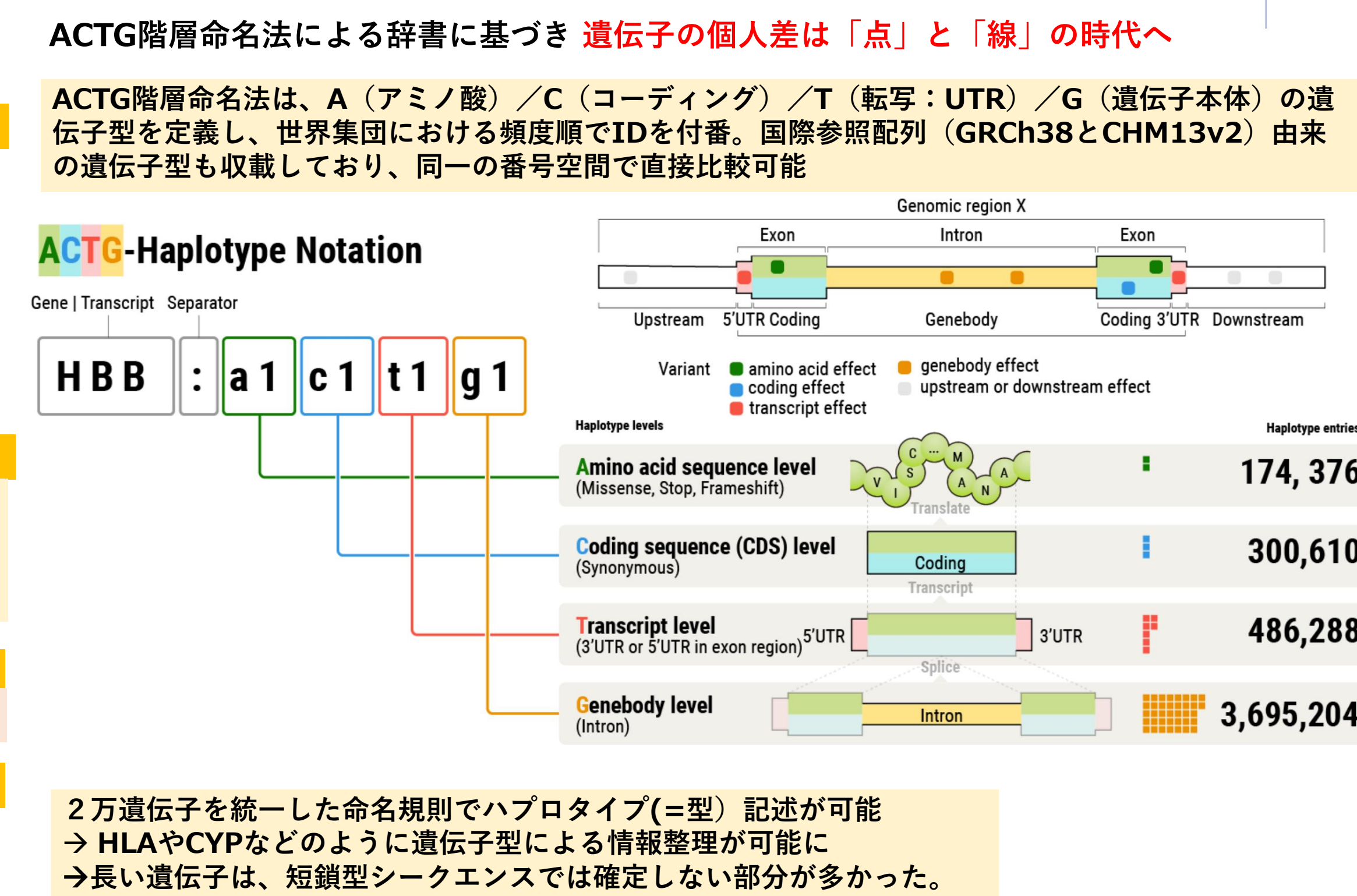
背景(1) ACTG塩基配列を読み取ることができる機器の種類



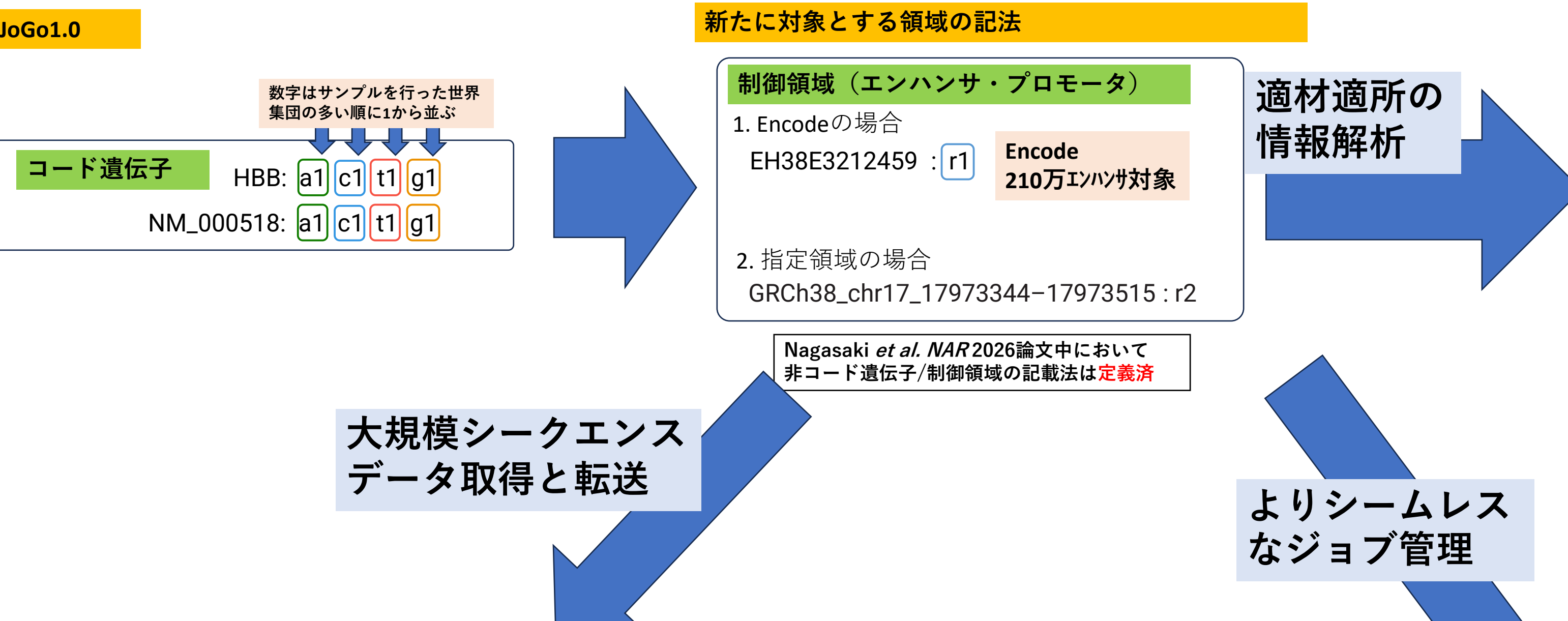
背景(2) ハプロタイプとは？



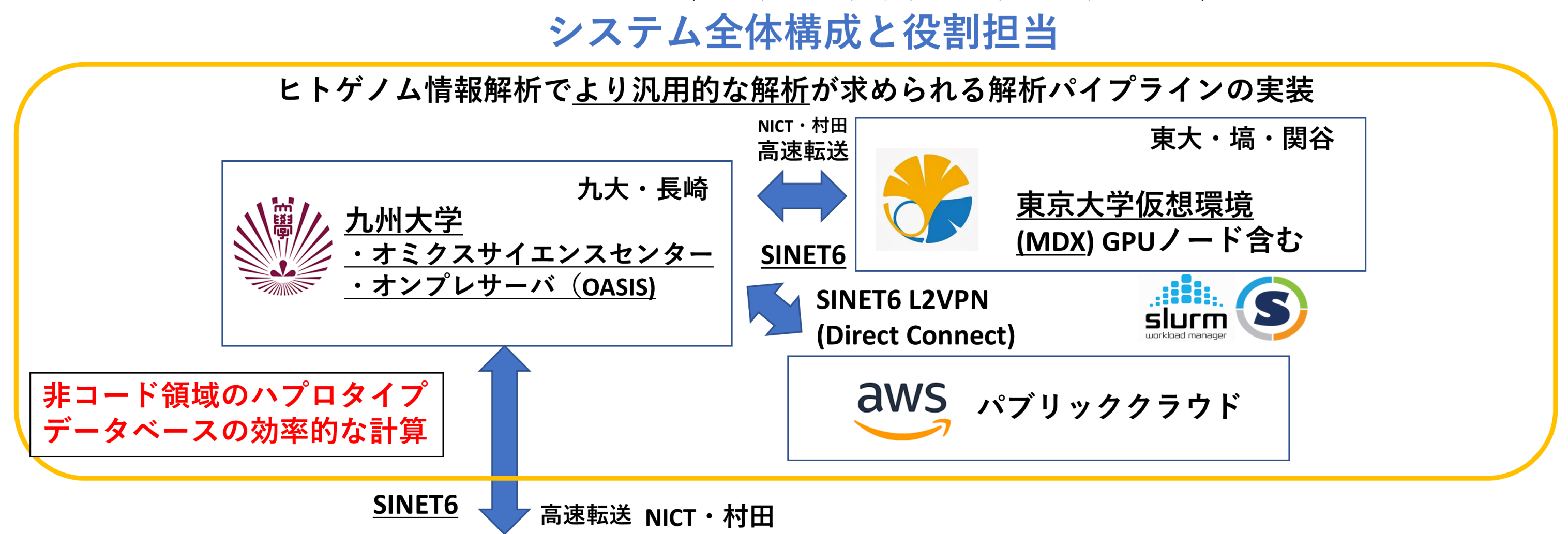
背景(3) ACTGハプロタイプ階層命名法



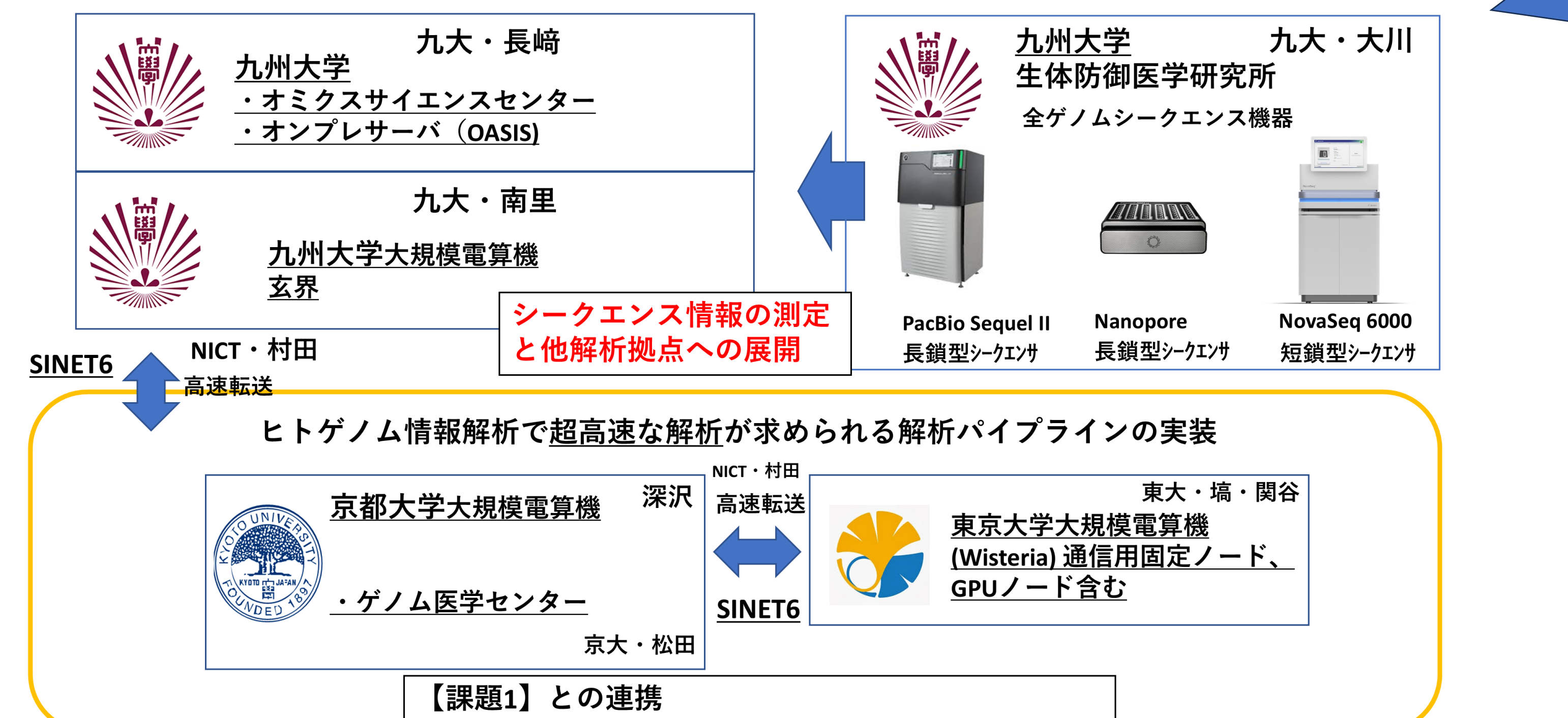
情報解析基盤で対象とする新たな領域



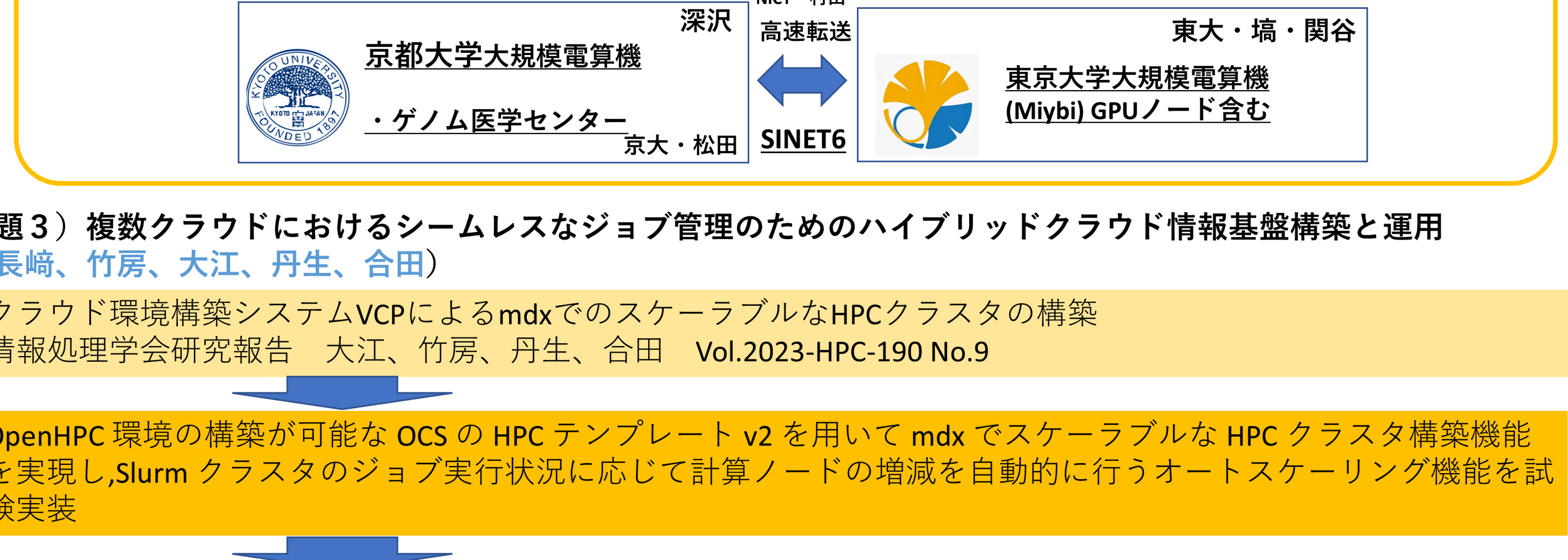
課題1) 中高深度長鎖シーケンス情報に基づく制御領域のハプロタイプリファレンスパネルの構築とそのための複数拠点間のハイブリッドクラウド情報基盤の運用(長崎、関谷、塙、深沢、大川、松田他)



課題2) 長鎖及び短鎖シーケンスから取得する情報を他拠点に効率良く展開するための設計検討と実装(大川、南里、長崎、深沢、村田)

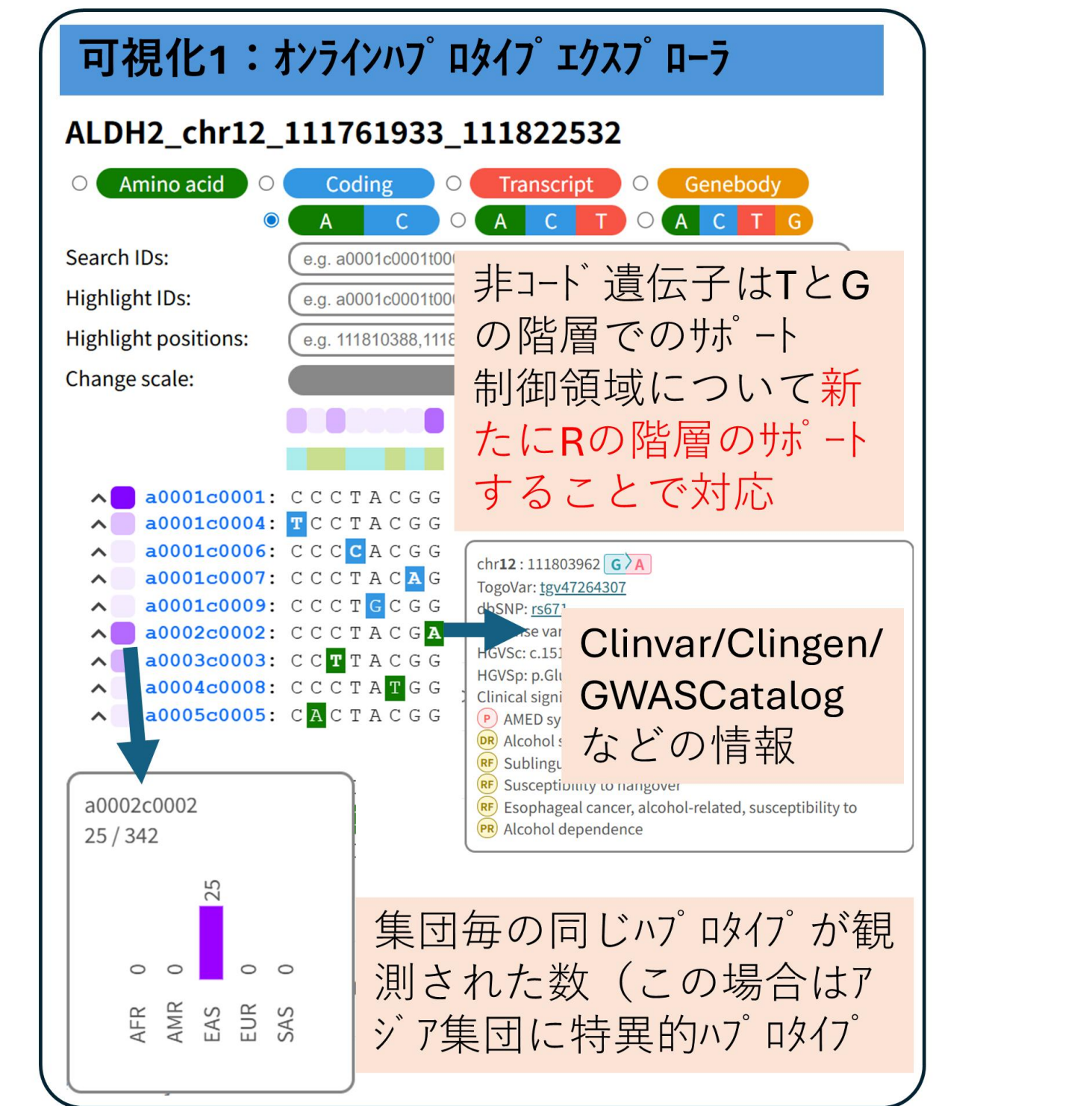
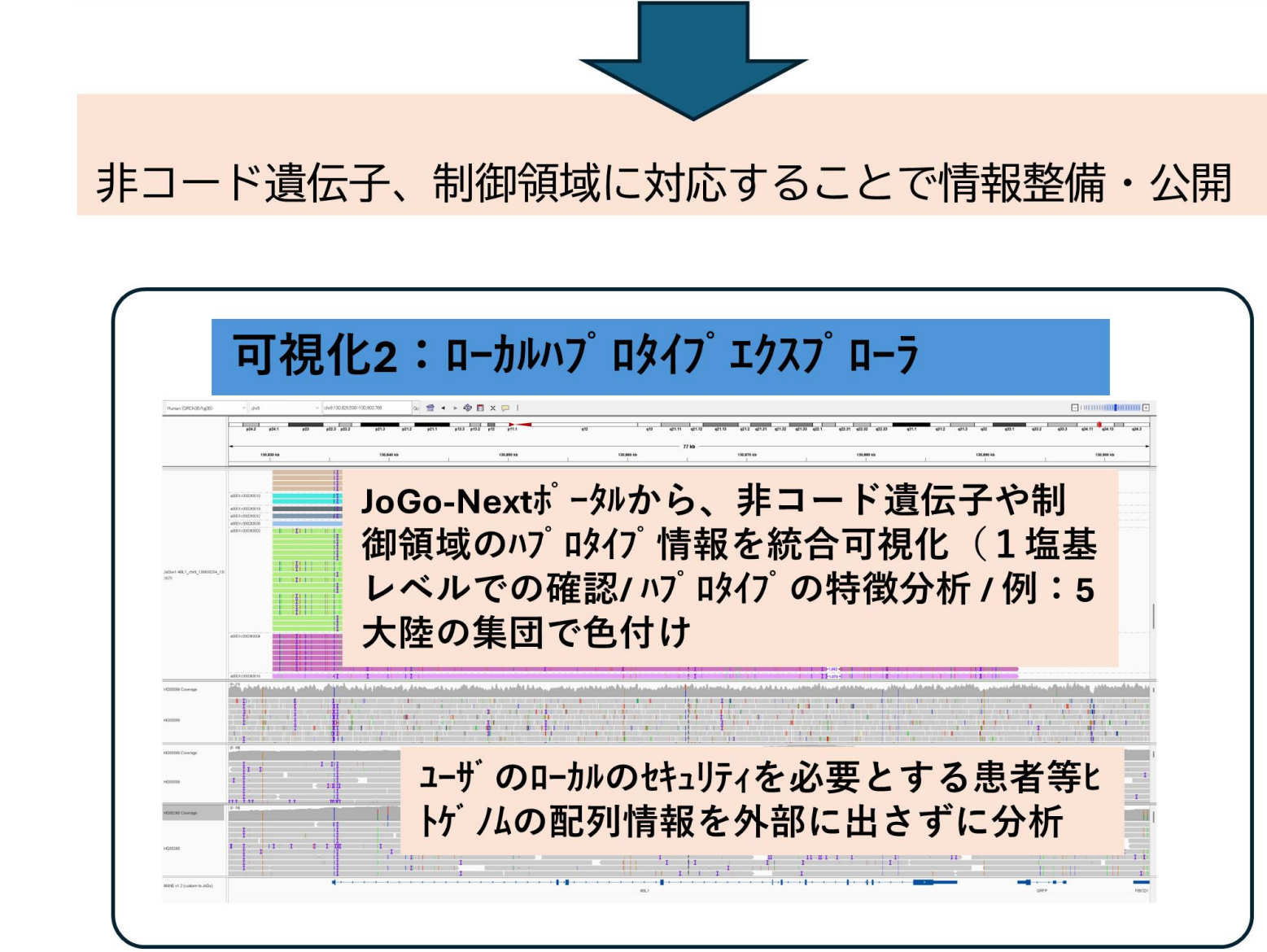


課題3) 複数クラウドにおけるシームレスなジョブ管理のためのハイブリッドクラウド情報基盤構築と運用(長崎、竹房、大江、丹生、合田)

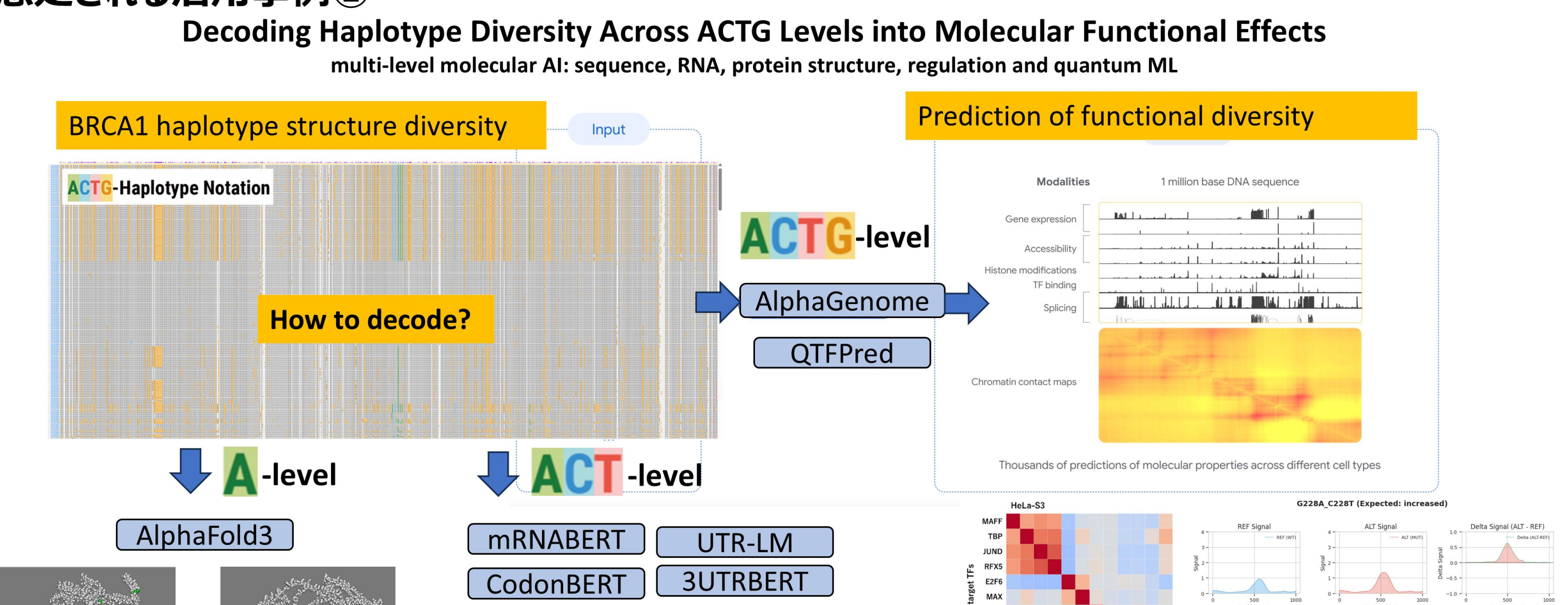


想定される活用事例①

「JoGo1.0」(可視化1)ウェブポータル上での可視化ツール(右図→)
②「T-Tango...」(3名制) M.Nagasaki. Practical guide for managing large-scale human genome data in research. *J Hum Genet*. 66(1): 39-52, 2021.
③ M.Nagasaki, K.Hirayasu, S.S.Khor, et al. JoGo-LLR Caller: Unveiling and Navigating the Complex Diversity of LILRB3-LILRA6 Copy Number Haplotype Structures with Whole-Genome Sequencing. *Human Immunology*. 96(9):11172, 2025.
④ K.Hirayasu, ... (14名制) M.Nagasaki. Identification of the hybrid gene LILRB3-3 by long-read sequencing and implication of its novel signaling function. *Front Immunol*. 15:1398935, 2024.
⑤ M.Nagasaki, T.Katayama, Y.Moriya, et al. Variant Information Standardization Collegium. JoGo 1.0: the ACTG hierarchical nomenclature and database covering 4.7 million haplotypes across 19,194 human genes. *Nucleic Acids Res*. 2025.
⑥ T.Matsubara, ... (5名制) M.Nagasaki. QTFPred: robust high-performance quantum machine learning modeling that predicts main and cooperative transcription factor bindings with base resolution. *Brief Bioinform*. 2025.
⑦ M.Kulmanov, S.Ashouri, Y.Liu, M.Abdelhakim, E.Alsolme, M.Nagasaki, et al. Phased genome assemblies and pangenome graphs of human populations of Japan and Saudi Arabia. *Sci Data*. 12(1):1316, 2025.



想定される活用事例②



1) M. Nagasaki, Y. Sekiya, A. Asakura, et al. Design and implementation of a hybrid cloud system for large-scale human genomic research, *Hum Genome Var*, 10: 6, 2023.
2) T. Tango, ... (3名制) M. Nagasaki. Practical guide for managing large-scale human genome data in research. *J Hum Genet*, 66(1): 39-52, 2021.
3) M. Nagasaki, K. Hirayasu, S.S. Khor, et al. JoGo-LLR Caller: Unveiling and Navigating the Complex Diversity of LILRB3-LILRA6 Copy Number Haplotype Structures with Whole-Genome Sequencing. *Human Immunology*, 96(9):11172, 2025.
4) K. Hirayasu, ... (14名制) M. Nagasaki. Identification of the hybrid gene LILRB3-3 by long-read sequencing and implication of its novel signaling function. *Front Immunol*, 15:1398935, 2024.
5) M. Nagasaki, T. Katayama, Y. Moriya, et al. Variant Information Standardization Collegium. JoGo 1.0: the ACTG hierarchical nomenclature and database covering 4.7 million haplotypes across 19,194 human genes. *Nucleic Acids Res*, 2025.
6) T. Matsubara, ... (5名制) M. Nagasaki. QTFPred: robust high-performance quantum machine learning modeling that predicts main and cooperative transcription factor bindings with base resolution. *Brief Bioinform*, 2025.
7) M. Kulmanov, S. Ashouri, Y. Liu, M. Abdelhakim, E. Alsolme, M. Nagasaki, et al. Phased genome assemblies and pangenome graphs of human populations of Japan and Saudi Arabia. *Sci Data*, 12(1):1316, 2025.