



From Linear to Pangenome: Improving Variant Detection in Clinical Genomics

Daniel Dörr
Molecular Diagnostics Symposium 2026

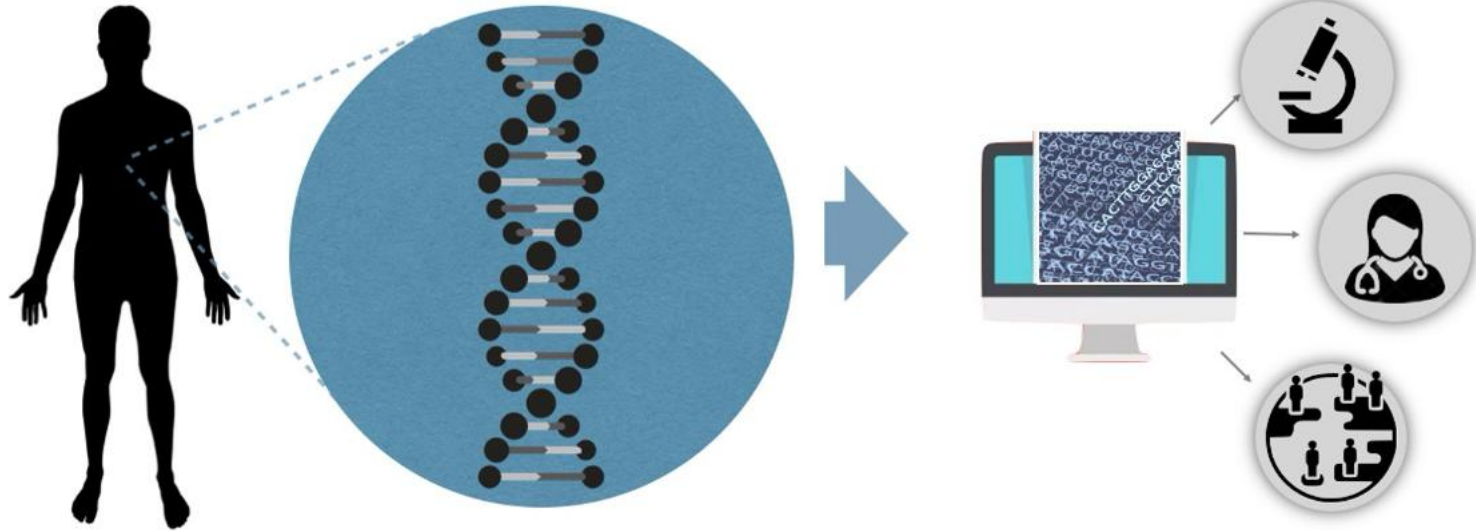
Milestones
Human
Genomics

The Human
Pangenome

Core
Applications

The Human Genome Project

Centralized coordinate system: critical for **sharing genomic data and data analysis standards** among international researchers.



2002

Standard Variant Calling Pipeline

Sample's DNA



Sequencing



Sequencing reads

```
ATACTGAC          GCACTTC GACTTCC
GTTACAC  GTACAGA          CAATACTG
          TTCACTG  CATGCCG
GTATCAT  GAATAC  ACGCTAC
          CGATCGT          GTACGCA
```



CGATCGACGCTAGCGCGTACGTACGTAGCTAGCATATCAACGATCA

Reference genome

Standard Variant Calling Pipeline

Sample's DNA



Sequencing



Sequencing reads



Read mapping



Reference genome



Standard Variant Calling Pipeline

Sample's DNA



Sequencing



Sequencing reads



Read mapping



Reference genome

Standard Variant Calling Pipeline

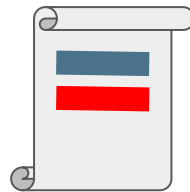
Sample's DNA



Sequencing



Sequencing reads



VCF file

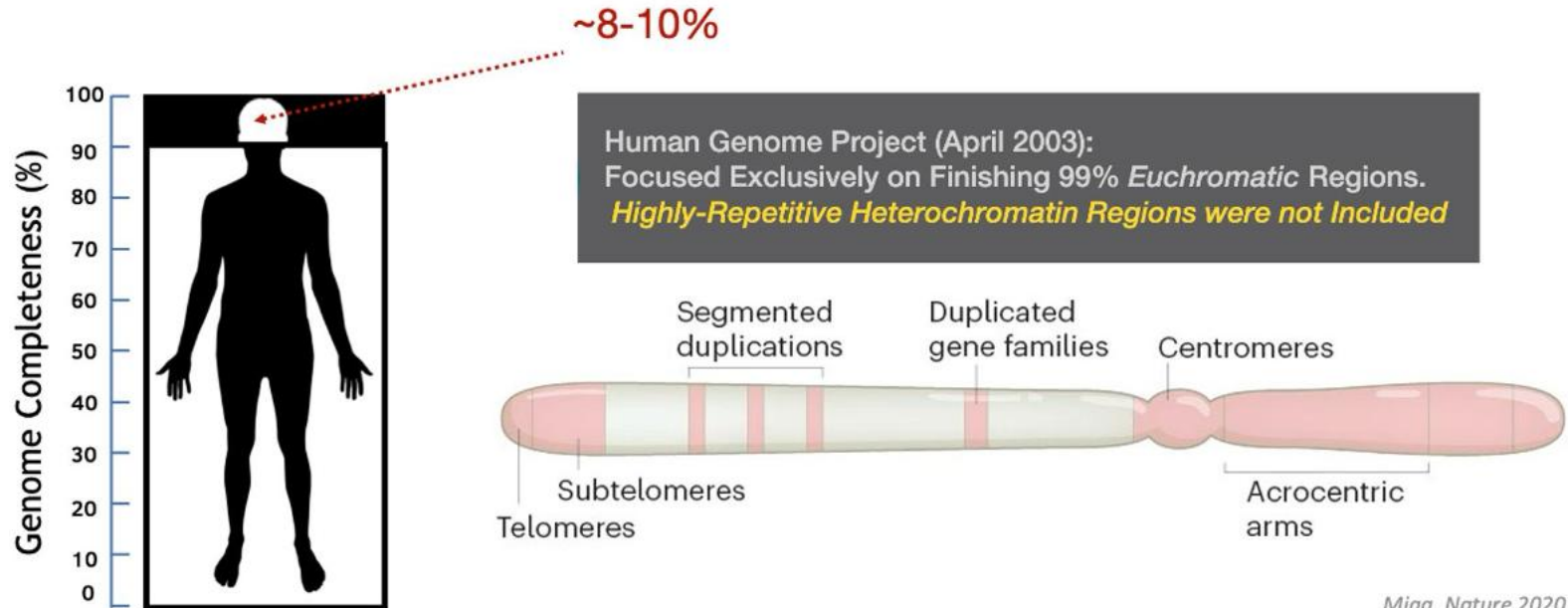
Variant calling



Reference genome

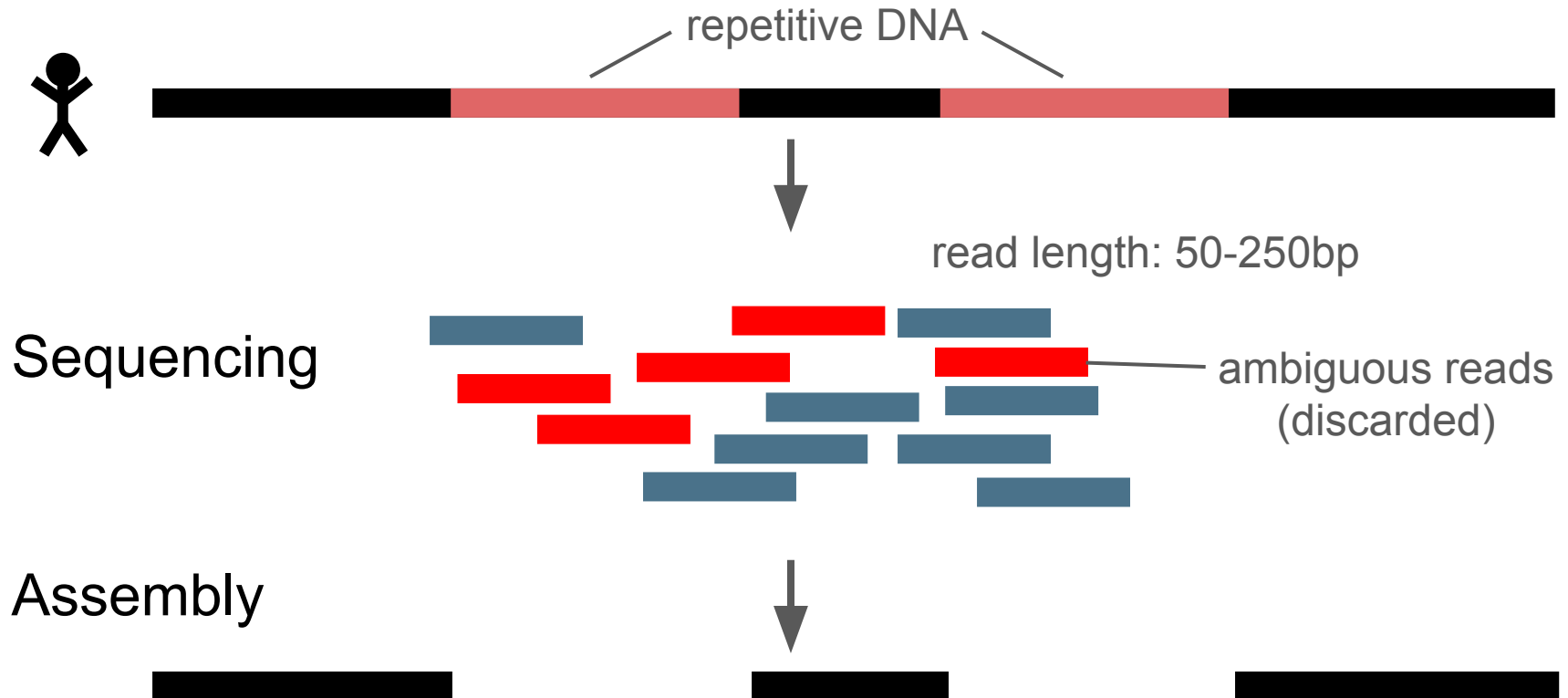
The Human Genome Project

Current version **GRCh38** still incomplete

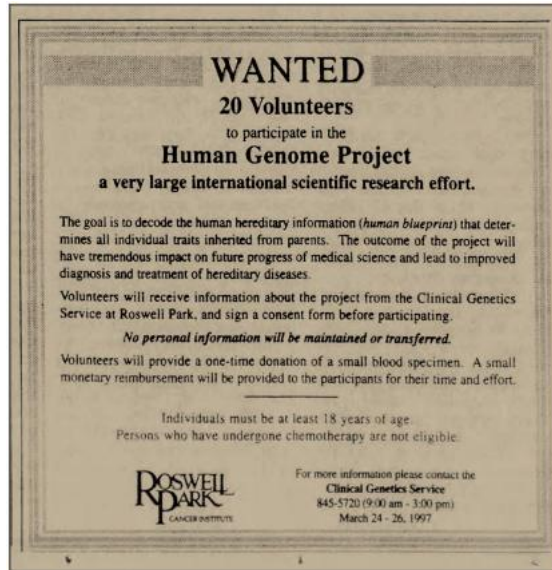


2002

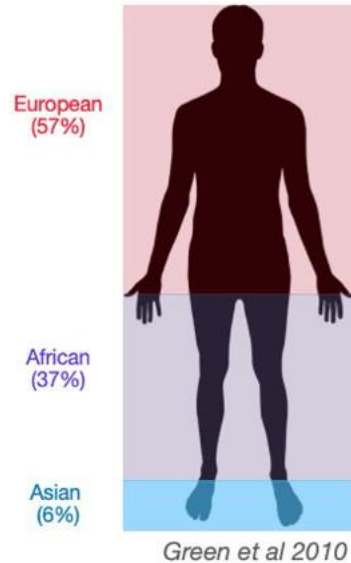
Shortcomings to Short-read Sequencing/Assembly



The Human Genome Project



News ad for recruiting volunteers for the Human Genome Project



- 70% of genome sequence from a single individual
- Biased towards **European descent**
- **Linear reference** (=single sequence) cannot represent diversity → reference bias

2002

Shortcomings to Standard Variant Calling

Sample's DNA



Sequencing



Sequencing reads

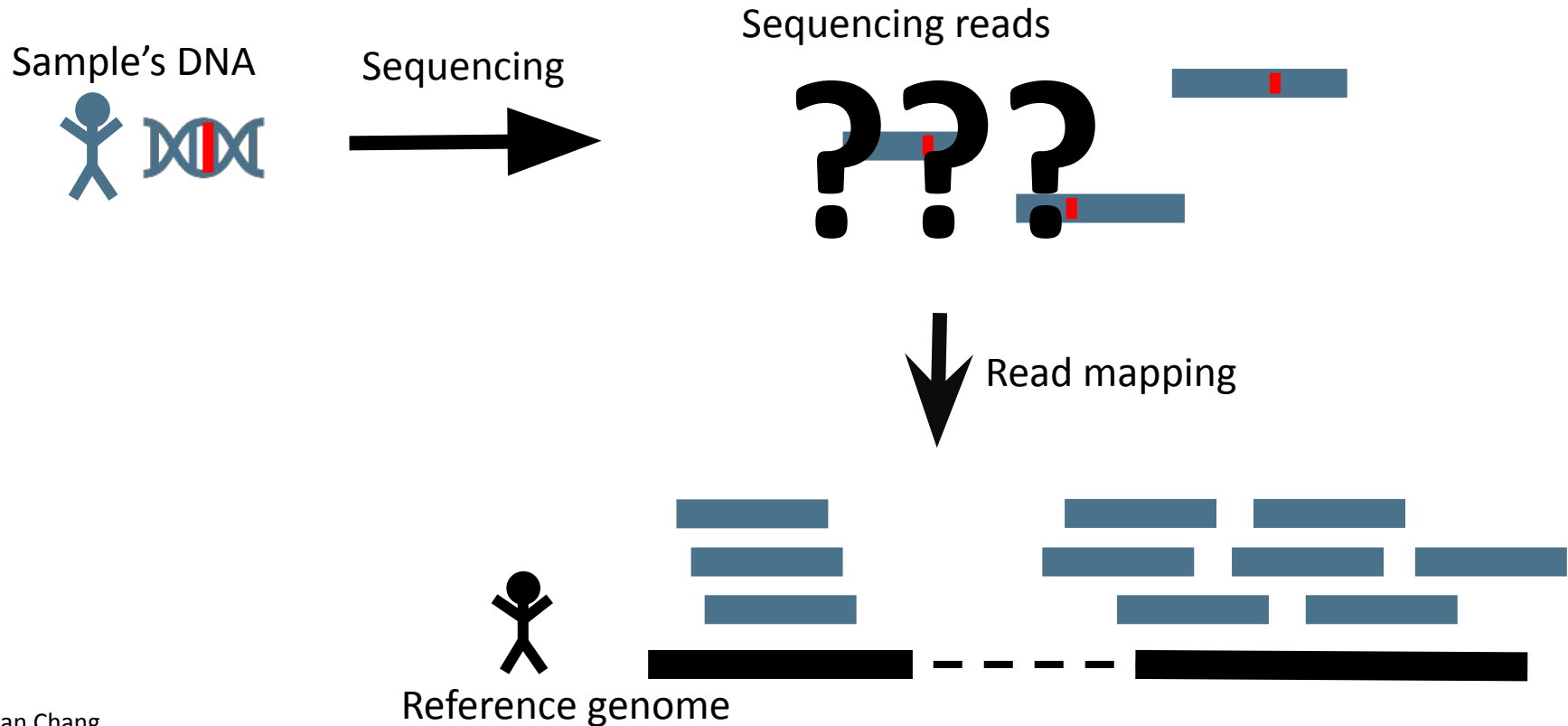


Read mapping



Reference genome

Shortcomings to Standard Variant Calling: Reference Bias



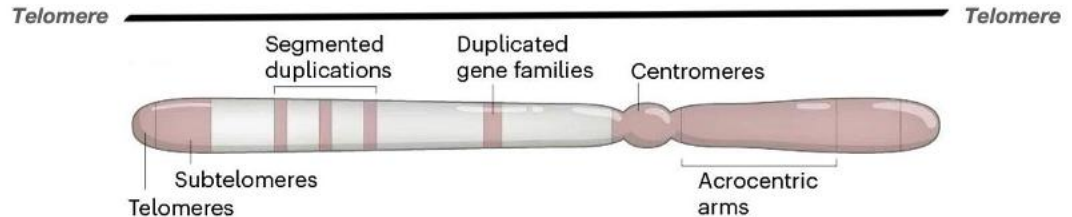
CHM13-T2T: A Complete Assembly of a Human Genome



Nurk, Koren, Rhie, Rautiainen, et al., [The complete sequence of a human genome](#), Science 2022

- Sequencing of CHM13, a completely **homozygous** human cell line
- Finally **complete human genome** with combination of high coverage long accurate reads (PacBio HiFi and ONT ultra-long reads)
- Introduced nearly **200 MB** of new sequence (vs GRCh38)

1 chromosome = 1 assembled contig



2002

2022

“Winning strategy” to T2T genome sequencing/assembly

PacBio SMRT

Circular Consensus Sequencing (CCS)
aka “HiFi”

- error rate: 0.1-0.2%
- read length: 20-25kb

Oxford Nanopore

Ultra-Long (UL) DNA Sequencing

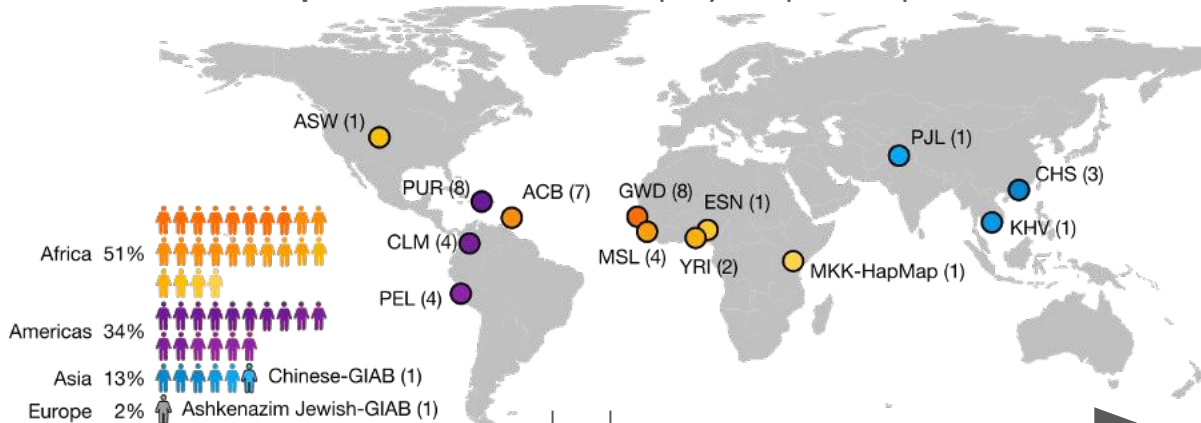
- error rate: 0.3-3%
- possible read length: >100kb

→ **Long sequence reads can span across repetitive regions**

The First Draft Release of a Human Pangenome



- **94 Haplotypes** from 47 fully phased diploid assemblies
 - 29 samples from the *1000 Genomes* (1KG) project
 - 18 from other sequencing projects
- **High-quality assemblies**, some near-T2T
- **119Mbp** of novel euchromatic polymorphic sequence

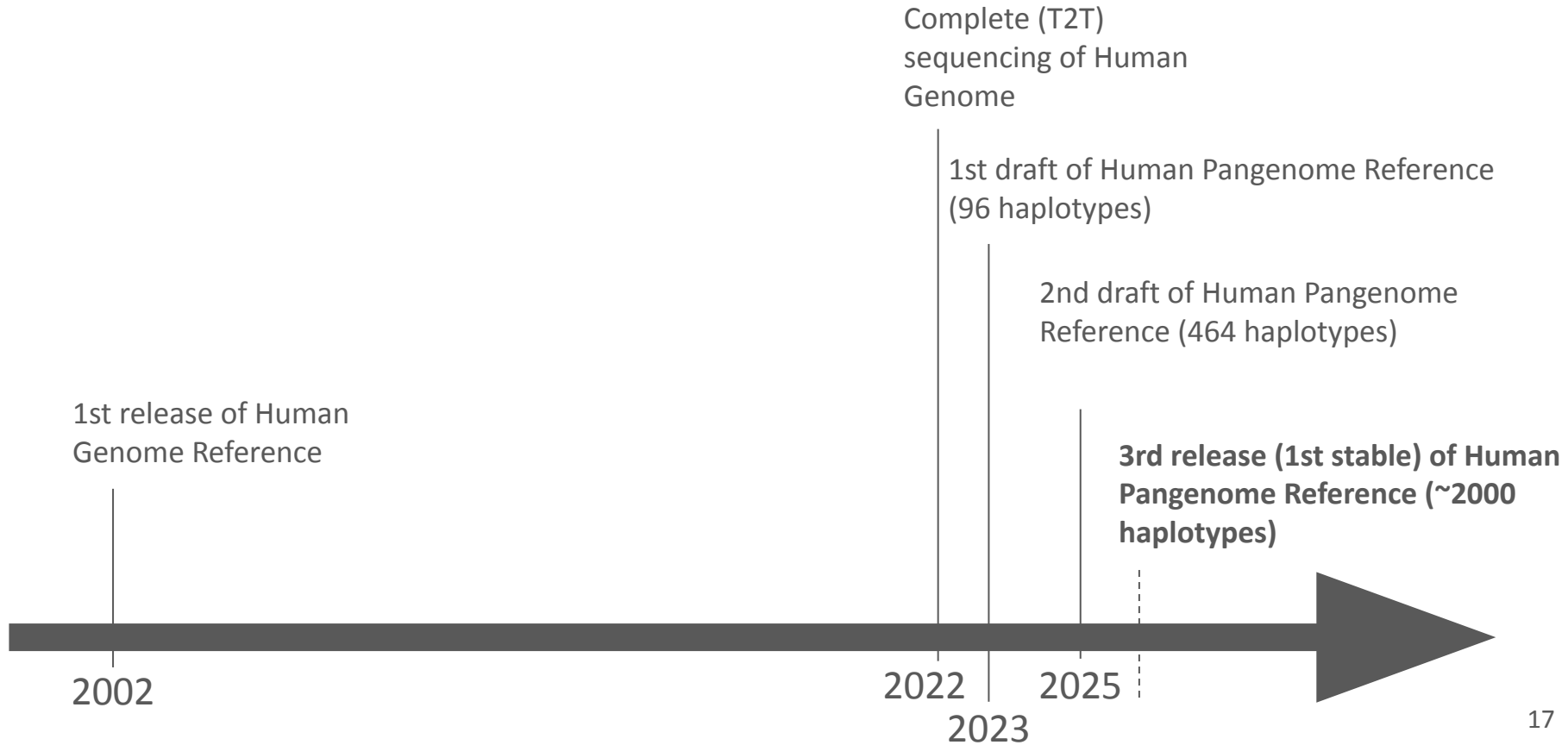


Liao, Asri, Ebler, Doerr, et al., [A Draft Human Pangenome Reference](#), Nature 2023

2002

2022
2023

The Human Genome Sequence in a Timeline

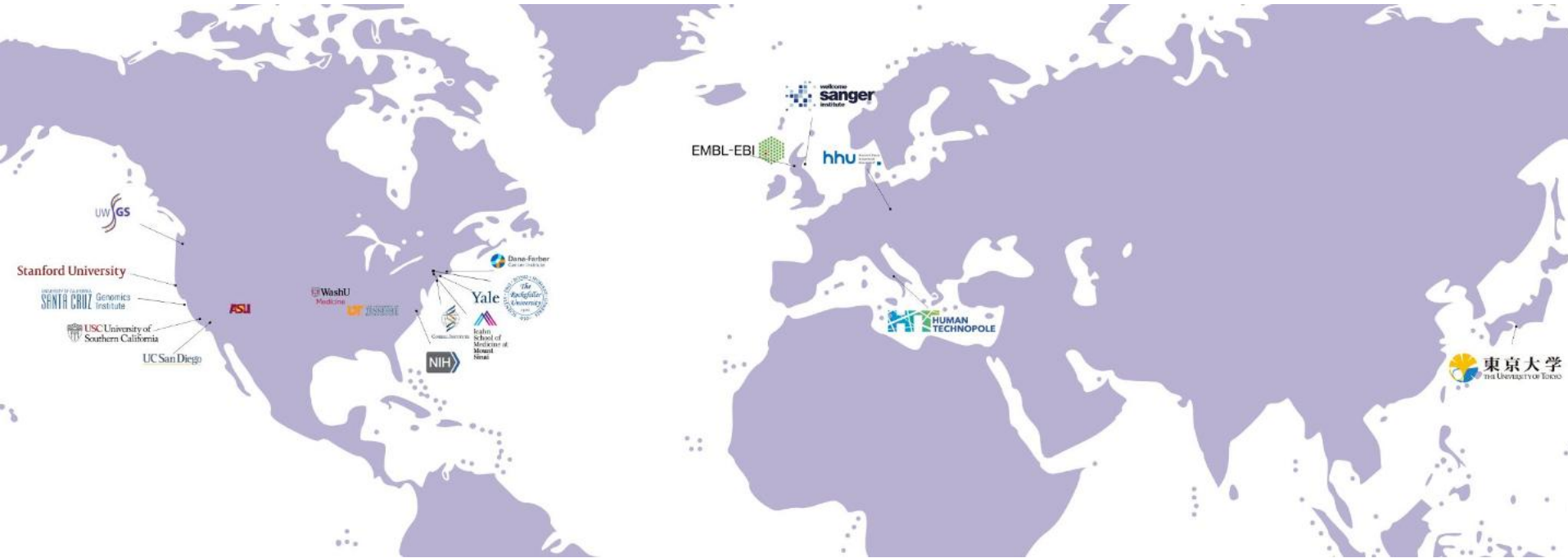


Milestones
Human
Genomics

The Human
Pangenome

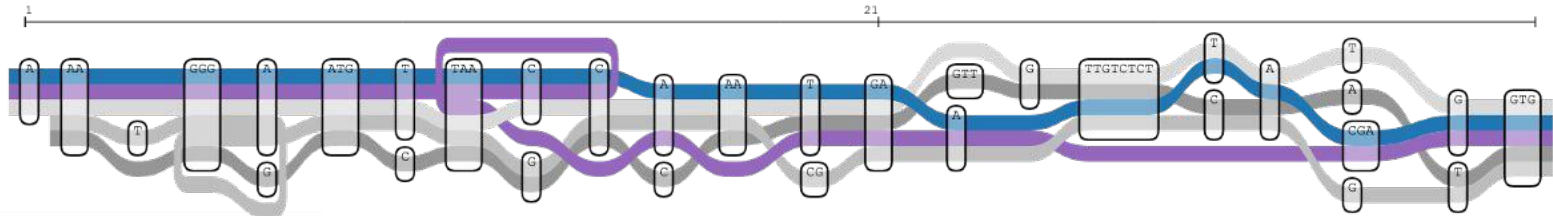
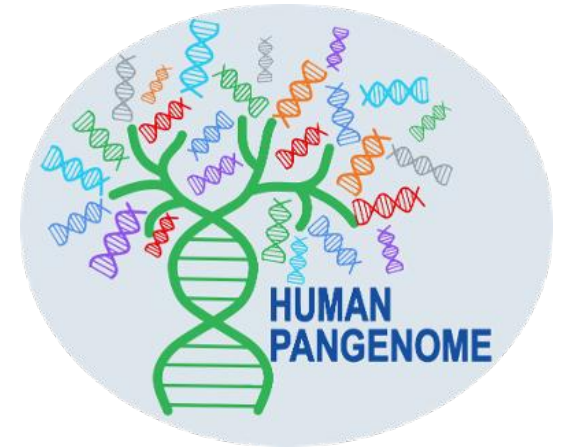
Core
Applications

The Human Pangenome Reference Consortium (HPRC)






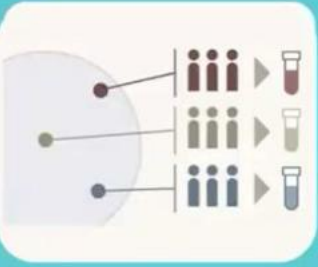

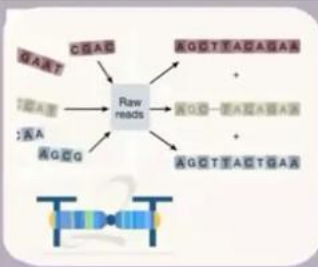
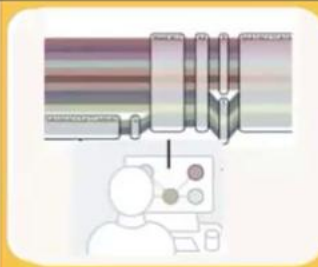



The Human Pangenome Reference Consortium (HPRC)

- Build the core resource
 - High-quality assemblies from >350 diverse humans
 - A map of common variants (>99%) and haplotypes
 - A reference data structure & coordinate system
- Nucleate and foster an ecosystem of tools
- Promote: adoption; guide deployment
- Establish a model for long-term growth via international collaborations



HPRC Working Groups

				
Population sampling and representation	Technology Production	Phased/Finished T2T Assemblies	Pangenome and new workflows/tooling	International Pangenome Project
				



**Embedded Ethics and Policy:
Inter-disciplinary ethics working group/oversight committee**

Pangenome Representations

~70%  1 Individual of blended ancestry

Original human genome reference sequence

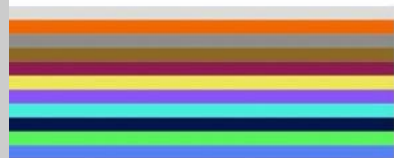
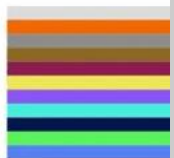
~30%  19 Individuals mostly of European ancestry

Collection

Pangenome

Collection of genome sequences from many individuals of the same species

Pangenome



Individual human genome reference sequences

Duplication

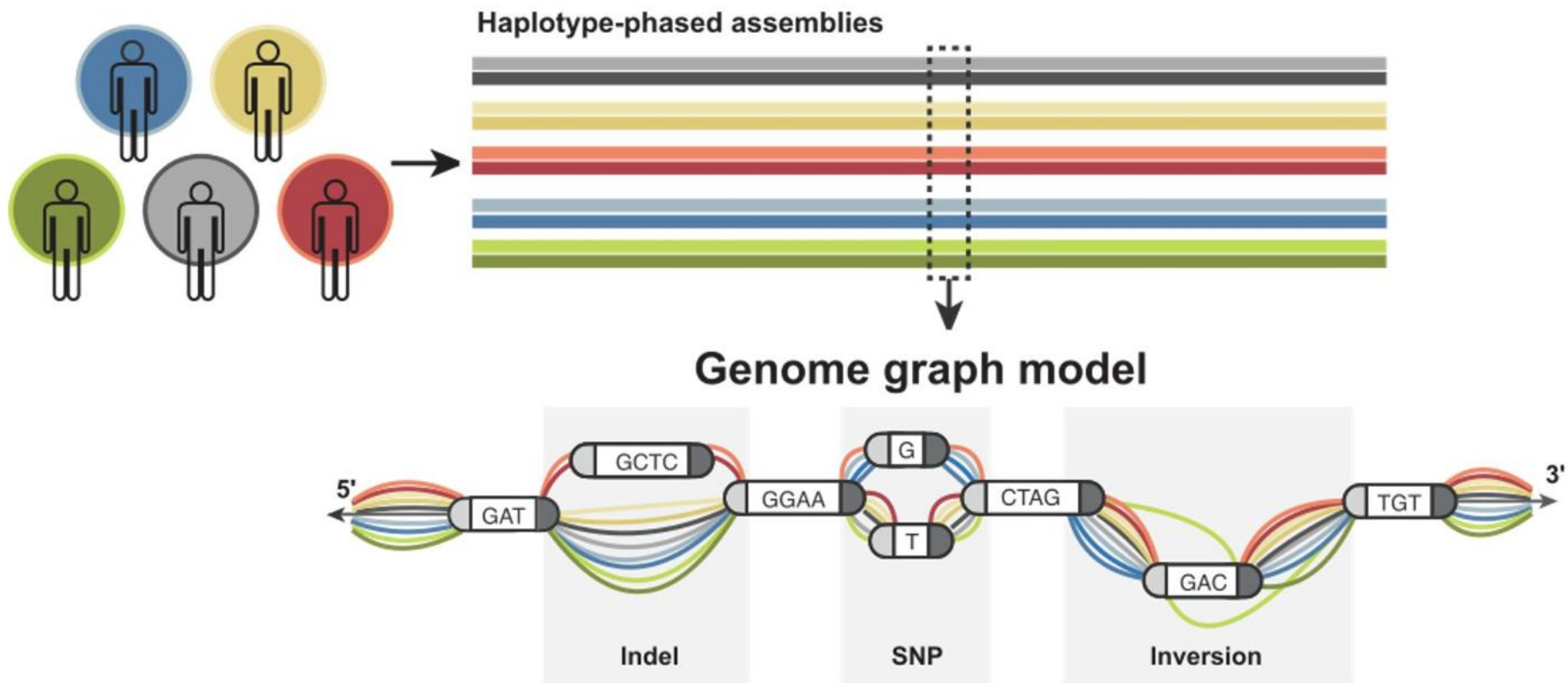
Deletion

Insertion

ATCGCG

Inversion

The Pangenome Graph

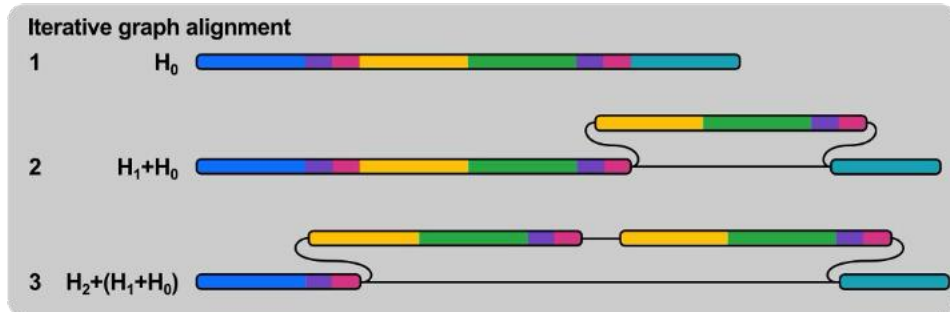


Pangenome Graph Construction

Input haplotypes assemblies



Progressive construction

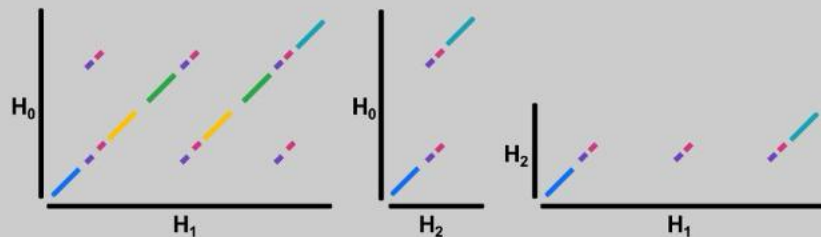


- 1st haplotype never folds back onto itself
- Segmental duplications remain uncollapsed

👉 Tools for graph construction

Order-independent construction

1 All-vs-all pairwise alignment



2 Segment-wise graph construction



Pangenome Graph Construction

Input haplotypes assemblies



Progressive construction



- 1st haplotype never folds back onto itself
- Segmental duplications remain uncollapsed

👉 Tools for graph construction

Order-independent construction

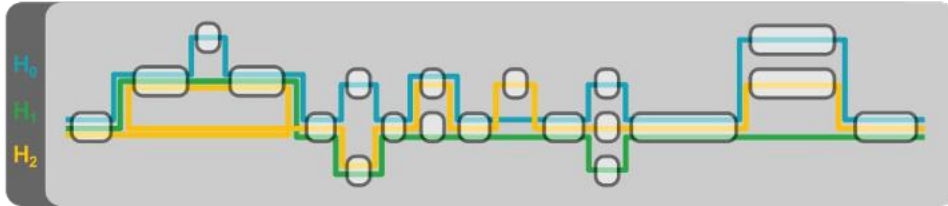


Graph types

Structural Variants (SVs): Length ≥ 50 bp

 **Tools for graph construction**

Base-level graphs



SV-level graphs

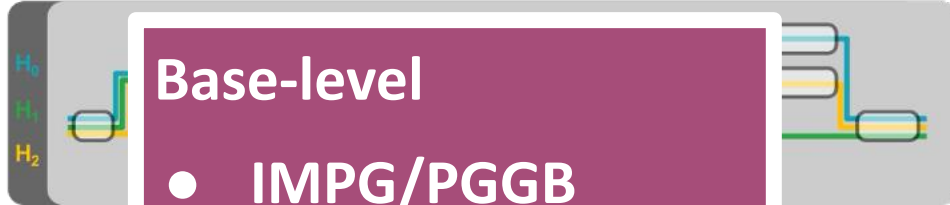


Graph types

Structural Variants (SVs): Length ≥ 50 bp

 **Tools for graph construction**

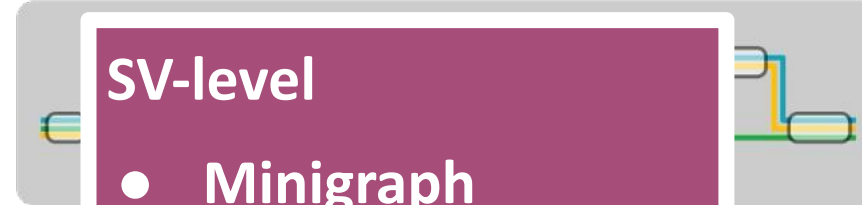
Base-level graphs



Base-level

- IMPG/PGGB
- Minigraph-Cactus

SV-level graphs



SV-level

- Minigraph

Current and Past Versions of HRPC Pangenome Graphs

1. Minigraph

- GRCh38-based
- CHM13-based

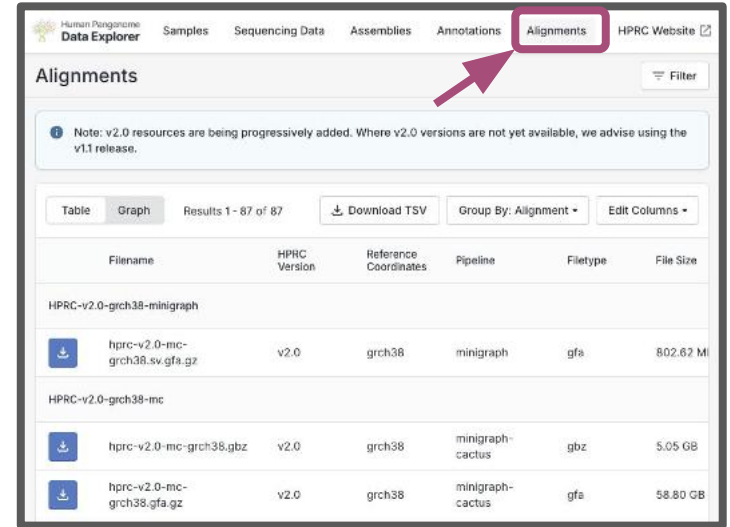
2. Minigraph-Cactus

- GRCh38-based
- CHM13-based

3. IMPG/PGGB

Which graph to choose depends on application:

- SV-analysis/fast visualization → Minigraph
- Segmental duplications/complex regions → PGGB, but graphs often too big for current methods
- Current “default” → Minigraph-Cactus: Most methods are tested on this graph type



Human Pangenome Data Explorer | Samples | Sequencing Data | Assemblies | Annotations | **Alignments** | HPRC Website

Alignments Filter

Note: v2.0 resources are being progressively added. Where v2.0 versions are not yet available, we advise using the v1.1 release.

Table | Graph | Results 1 - 87 of 87 | Download TSV | Group By: Alignment | Edit Columns

Filename	HPRC Version	Reference Coordinates	Pipeline	Filetype	File Size
HPRC-v2.0-grch38-minigraph					
hprc-v2.0-mc-grch38.sv.gfa.gz	v2.0	grch38	minigraph	gfa	802.62 Mi
HPRC-v2.0-grch38-mc					
hprc-v2.0-mc-grch38.gbz	v2.0	grch38	minigraph-cactus	gbz	5.05 GB
hprc-v2.0-mc-grch38.gfa.gz	v2.0	grch38	minigraph-cactus	gfa	58.80 GB

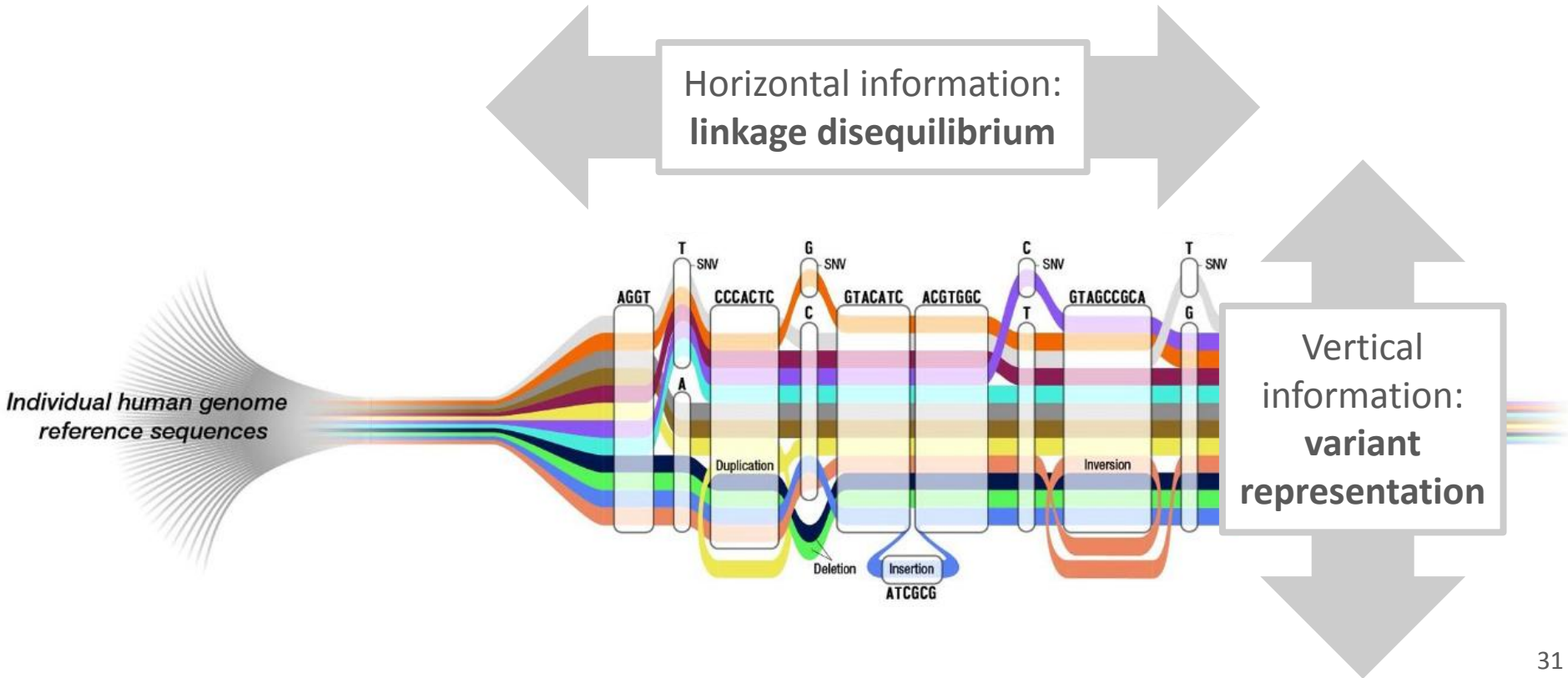
<https://data.humanpangenome.org/alignments>

Milestones
Human
Genomics

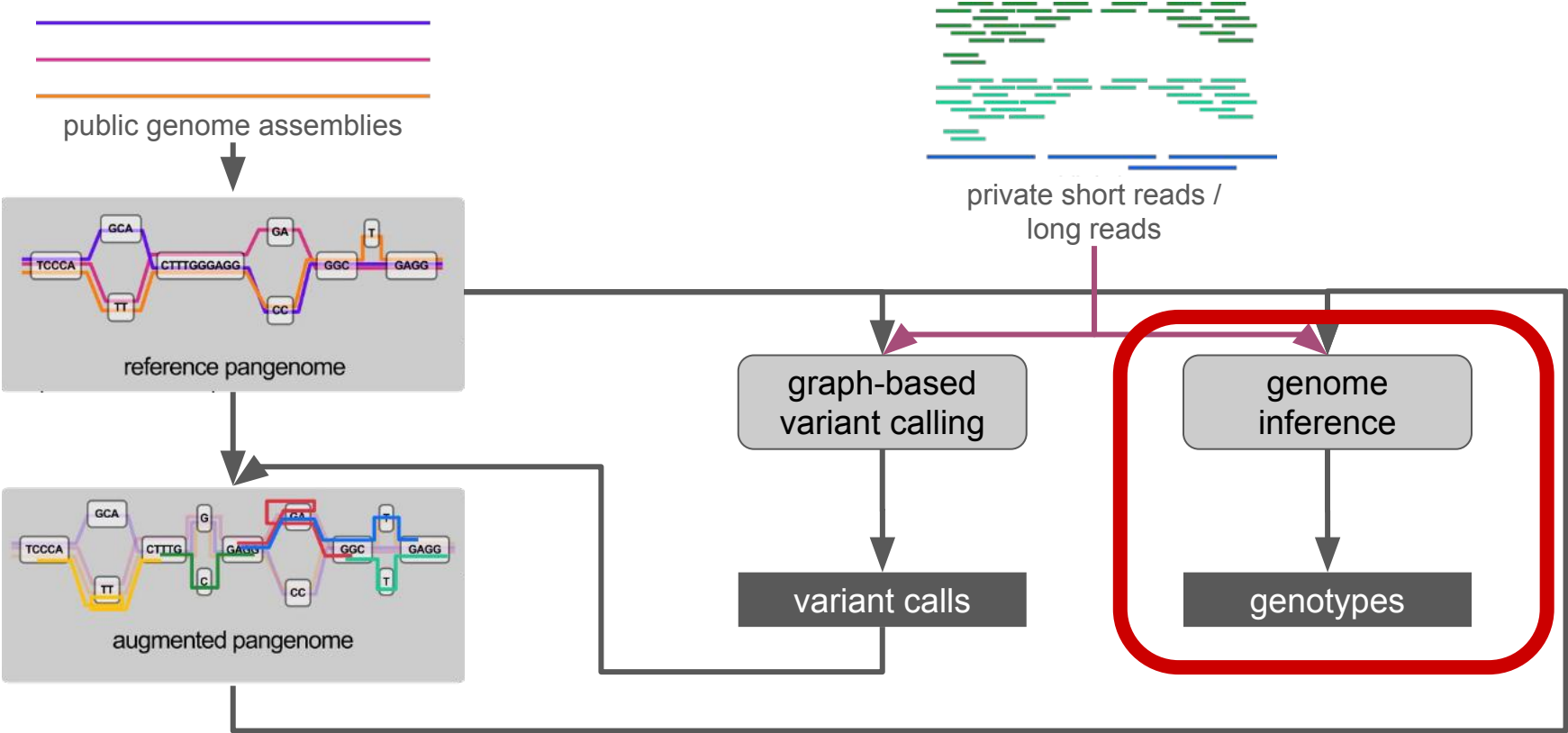
The Human
Pangenome

Core
Applications

From One Reference to Many Paths: The Pangenome Advantage



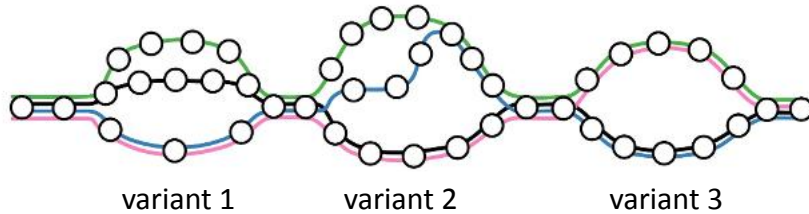
Pangenome-based workflows and core applications



Pangenome-based Genotyping

Input:

1. **Pangenome graph** constructed from known haplotypes



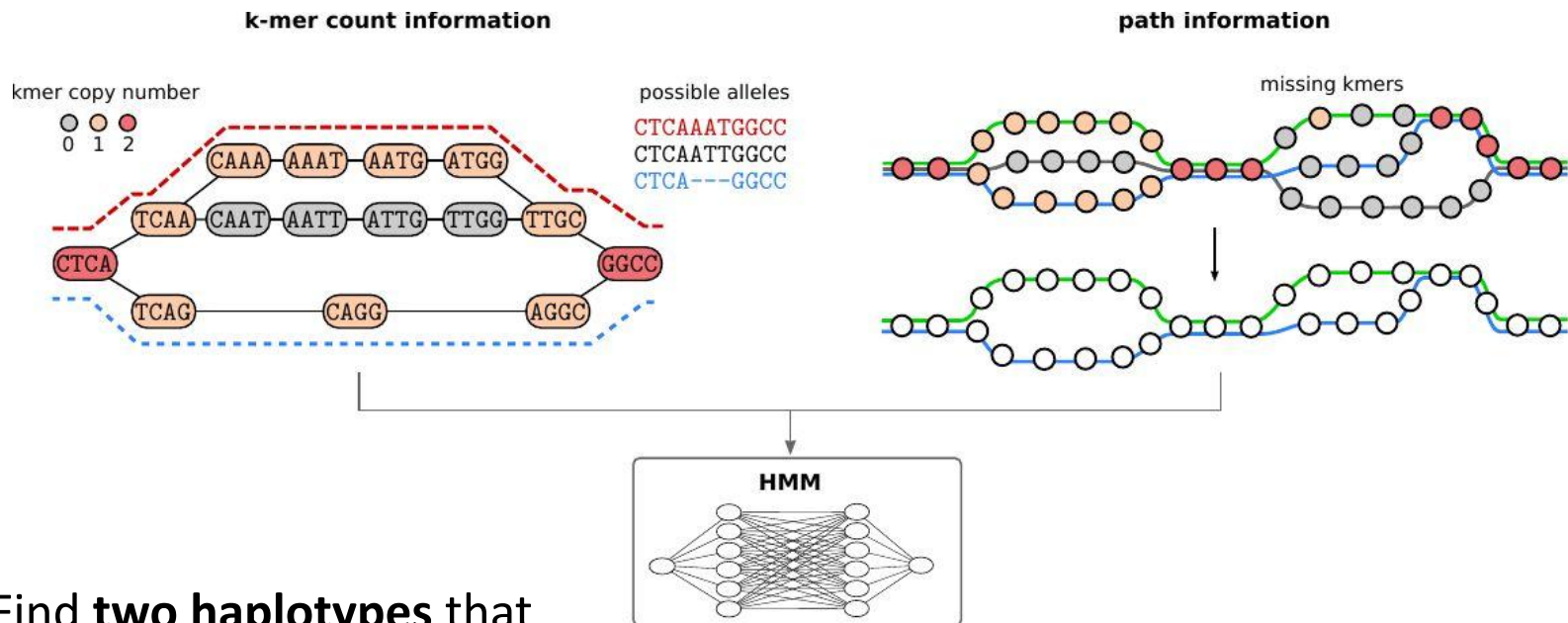
2. **Short reads** of the sample to be genotyped (not in graph!)



Goal: Determine genotypes of **unknown genome** (= not in graph) for all variants contained in graph



PanGenie: Idea



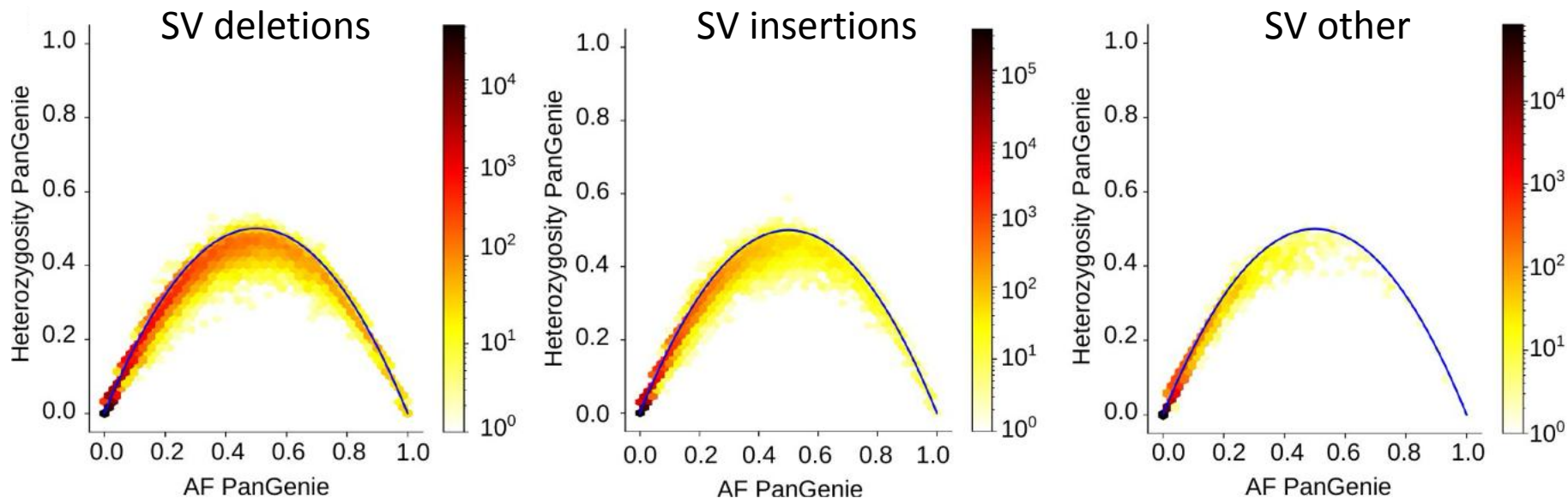
Find **two haplotypes** that

(a) best explain the observed k-mer counts and

(b) are mosaics of the known haplotypes present in the graph



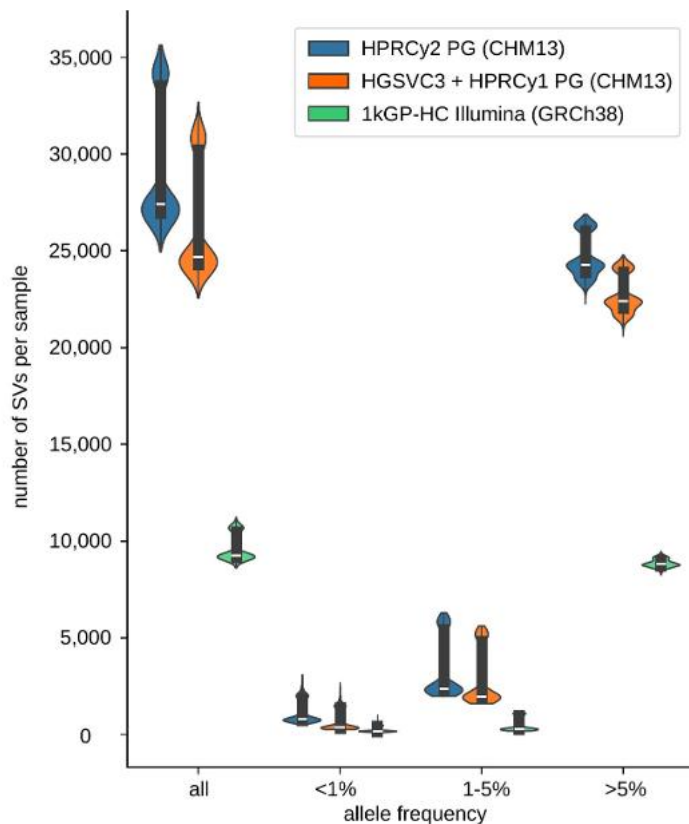
Pangenie: Genotyping SVs in 3,202 samples of the 1KG cohort



- Based on HPRC2 pangenome
- Genotyping in line with Hardy-Weinberg Equilibrium

Ebler *et al*, *Nat. Genet.* (2022)

Pangenie: Genotyping SVs in 3,202 samples of the 1kGP cohort



- Comparison of
 - **Pangenie** based on HPRC2 & HPRC1+HGSVC3 pangenomes
 - **traditional** alignment-based SV calling workflow (green)
- PanGenie can access **thousands of additional SVs** compared to traditional short-read based SV callers

Ebler *et al*, *Nat. Genet.* (2022)

Locityper: targeted locus genotyper

- Input: **local haplotypes** (e.g. from pangenome) + **WGS dataset** (fastq/cram)
- Output: **two haplotypes** most similar to the WGS dataset
+ **read alignments** to these haplotypes
- **Higher accuracy over complex medically relevant genes** compared to state-of-the-art variant calling pipelines

Compared to Pangenie:

- can handle more complex loci,
- but not scalable to the full genome,
- worse at predicting individual SNPs



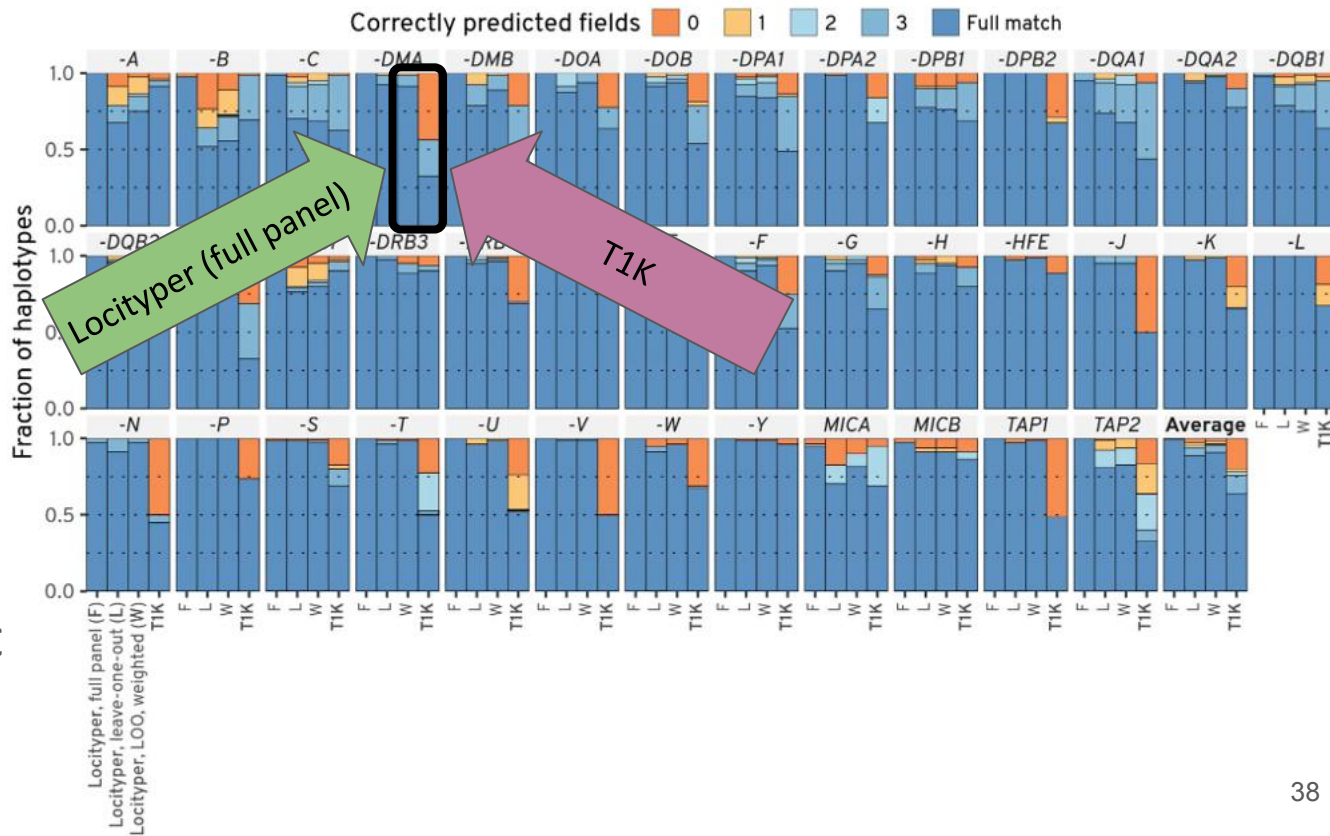
github.com/tprodanov/locityper

Prodanov *et al*, *Nat. Genet.* (2025)

Locityper: Genotyping the MHC locus

Locityper

- shows high accuracy in predicting *HLA* alleles
- can genotype
 - gene order,
 - gene deletions,
 - tandem duplications,
 - etc.

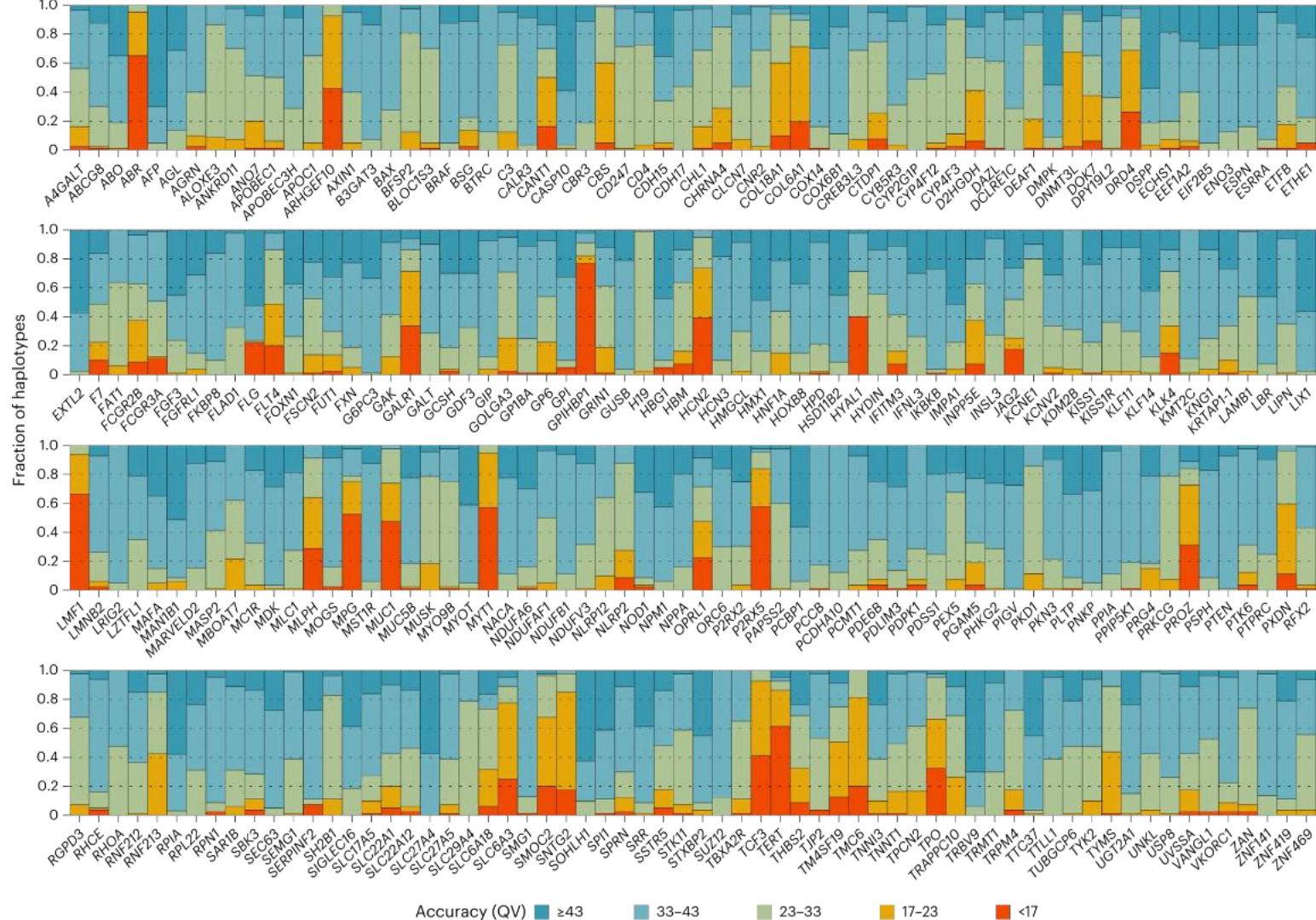


T1K: state of the art MHC genotyper

● 256 challenging medically relevant loci

● **Locityper** using leave-one-out HPRC1 reference panel

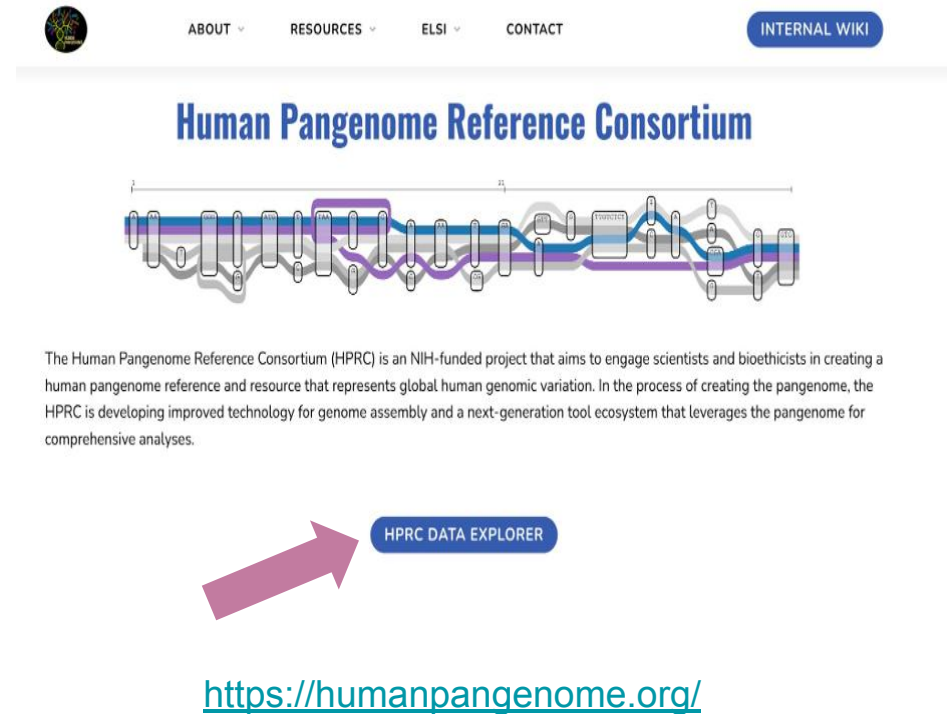
● 40 Illumina WGS datasets



Discussion & Outlook

Discussion

- Long-read sequencing technology transforms clinical genomics
- Human Pangenome Reference: a valuable resource of human genomic diversity in the workings
- Sequencing data, assemblies, annotations, and pangenome graphs produced by HPRC are open & available
- Pangenome-based genotypers already outperform “linear” state-of-the-art methods
- Illumina “DRAGON” makes already use of human pangenome



ABOUT ▾ RESOURCES ▾ ELSI ▾ CONTACT

INTERNAL WIKI

Human Pangenome Reference Consortium

The Human Pangenome Reference Consortium (HPRC) is an NIH-funded project that aims to engage scientists and bioethicists in creating a human pangenome reference and resource that represents global human genomic variation. In the process of creating the pangenome, the HPRC is developing improved technology for genome assembly and a next-generation tool ecosystem that leverages the pangenome for comprehensive analyses.

HPRC DATA EXPLORER

<https://humanpangenome.org/>

Outlook

- Human Pangenome Reference is not alone
 - Chinese pangenome reference (1000 haplotypes)
 - Arab pangenome reference
 - Japanese and Saudi Arabian pangenome
 - etc.
- Human Pangenome Project
 - Federated alliance of international genomic partners
 - Pangenomic software ecosystem and broad user base
 - Facilitate educational outreach and capacity building

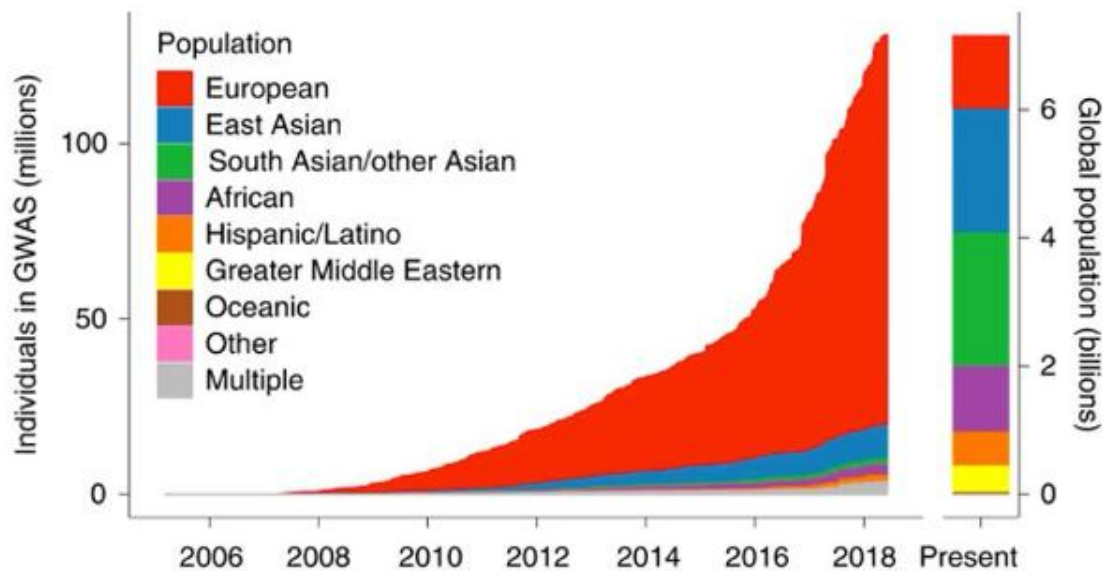


Global Alliance
for Genomics & Health

Collaborate. Innovate. Accelerate.

Outlook

- Lack of diversity & representation not specific to the Human Genome Project



Thank you!

HPRC, in particular:

Andrea Guarracino

Eimear Kenny

Erik Garrison

Glenn Hickey

Heather Lawson

Heng Li

Jana Ebler

Peter Heringer

Timofey Prodanov

Tobias Marschall

CoDia

Computational Diabetology



References

- Corazzini, L. et al. Recurrent DNA break clusters drive replication-stress-induced copy number variants and genome diversification. *Nat. Commun.* 17, 3627 (2026).
- Wang, Y. et al. The 1000 Chinese Pangenome empowers medical and population genetics. *Nature* (2026) doi:10.1038/s41586-026-10315-y.
- Marx, V. It's a colorful pangenome world. *Nat. Methods* 23, 487–492 (2026).
- Prodanov, T. et al. Locityper enables targeted genotyping of complex polymorphic genes. *Nat. Genet.* 57, 2901–2908 (2025).
- Ma, W. & Chaisson, M. J. P. Genotyping sequence-resolved copy number variation using pangenomes reveals paralog-specific global diversity and expression divergence of duplicated genes. *Nat. Genet.* 57, 2909–2919 (2025).
- Garimella, K. V. et al. Population-scale long-read sequencing in the All of Us Research Program. *medRxiv* (2025) doi:10.1101/2025.10.02.25336942.
- Marx, V. Choose your human genome reference wisely. *Nat. Methods* 22, 2010–2014 (2025).
- Chang, X. et al. Rapid, accurate long- and short-read mapping to large pangenome graphs with vg Giraffe. *bioRxiv* (2025) doi:10.1101/2025.09.29.678807.
- Nowbandegani, P. S., Zhang, S., Hu, H., Li, H. & O'Connor, L. J. Defining and cataloging variants in pangenome graphs. *bioRxiv* (2025) doi:10.1101/2025.08.04.668502.
- Collins, R. L. & Talkowski, M. E. Diversity and consequences of structural variation in the human genome. *Nat. Rev. Genet.* 26, 443–462 (2025).
- Asri, M. et al. Pangenome-aware DeepVariant. *bioRxiv* (2025) doi:10.1101/2025.06.05.657102.
- Steyaert, W. et al. Unraveling undiagnosed rare disease cases by HiFi long-read genome sequencing. *Genome Res.* 35, 755–768 (2025).
- Kulmanov, M. et al. Phased genome assemblies and pangenome graphs of human populations of Japan and Saudi Arabia. *Genomics* (2024).
- Olbrich, M. et al. An Emirati pangenome incorporating a diploid telomere-to-telomere reference. *Genomics* (2024).
- Mostovoy, Y. et al. Resolution of ring chromosomes, Robertsonian translocations, and complex structural variants from long-read sequencing and telomere-to-telomere assembly. *Am. J. Hum. Genet.* 111, 2693–2706 (2024).
- Matthews, C. A., Watson-Haigh, N. S., Burton, R. A. & Sheppard, A. E. A gentle introduction to pangenomics. *Brief. Bioinform.* 25, bbae588 (2024).
- Taylor, D. J. et al. Beyond the Human Genome Project: The Age of Complete Human Genome Sequences and Pangenome References. *Annu. Rev. Genomics Hum. Genet.* (2024) doi:10.1146/annurev-genom-021623-081639.
- Groza, C. et al. Pangenome graphs improve the analysis of structural variants in rare genetic diseases. *Nat. Commun.* 15, 657 (2024).
- Abondio, P., Bruno, F., Passarino, G., Montesanto, A. & Luiselli, D. Pangenomics: A new era in the field of neurodegenerative diseases. *Ageing Res. Rev.* 94, 102180 (2023).
- Chaisson, M. J. P., Sulovari, A., Valdmanis, P. N., Miller, D. E. & Eichler, E. E. Advances in the discovery and analyses of human tandem repeats. *Emerg Top Life Sci* 7, 361–381 (2023).
- Miga, K. H. & Eichler, E. E. Envisioning a new era: Complete genetic information from routine, telomere-to-telomere genomes. *Am. J. Hum. Genet.* 110, 1832–1840 (2023).
- Attwaters, M. A diverse and inclusive human pangenome. *Nat. Rev. Genet.* 24, 585 (2023).
- Lee, H. et al. Pan-conserved segment tags identify ultra-conserved sequences across assemblies in the human pangenome. *Cell Rep Methods* 3, 100543 (2023).
- Chin, C.-S. et al. Multiscale analysis of pangenomes enables improved representation of genomic diversity for repetitive and clinically relevant genes. *Nat. Methods* 20, 1213–1221 (2023).
- Gao, Y. et al. A pangenome reference of 36 Chinese populations. *Nature* 619, 112–121 (2023).
- Audano, P. A. & Beck, C. R. Small allelic variants are a source of ancestral bias in structural variant breakpoint placement. *bioRxiv* (2023) doi:10.1101/2023.06.25.546295.
- Song, L., Bai, G., Liu, X. S., Li, B. & Li, H. Efficient and accurate KIR and HLA genotyping with massively parallel sequencing data. *Genome Res.* 33, 923–931 (2023).
- Vollger, M. R. et al. Increased mutation and gene conversion within human segmental duplications. *Nature* 617, 325–334 (2023).
- Liao, W.-W. et al. A draft human pangenome reference. *Nature* 617, 312–324 (2023).
- Guarracino, A. et al. Recombination between heterologous human acrocentric chromosomes. *Nature* 617, 335–343 (2023).
- Porubsky, D. et al. Inversion polymorphism in a complete human genome assembly. *Genome Biol.* 24, 100 (2023).
- Sibbesen, J. A. et al. Haplotype-aware pantranscriptome analyses using spliced pangenome graphs. *Nat. Methods* 20, 239–247 (2023).
- Uddin, M. et al. A draft Arab pangenome reference. (2023).
- Porubsky, D. et al. Recurrent inversion polymorphisms in humans associate with genetic instability and genomic disorders. *Cell* 185, 1986–2005.e26 (2022).
- Nurk, S. et al. The complete sequence of a human genome. *Science* 376, 44–53 (2022).
- Wang, T. et al. The Human Pangenome Project: a global resource to map genomic diversity. *Nature* 604, 437–446 (2022).
- Ebler, J. et al. Pangenome-based genome inference allows efficient and accurate genotyping across a wide spectrum of variant classes. *Nat. Genet.* 54, 518–525 (2022).
- Sirén, J. et al. Pangenomics enables genotyping of known structural variants in 5202 diverse genomes. *Science* 374, abg8871 (2021).
- Lu, T.-Y., Human Genome Structural Variation Consortium & Chaisson, M. J. P. Profiling variable-number tandem repeat variation across populations using repeat-pangenome graphs. *Nat. Commun.* 12, 4250 (2021).
- Ebert, P. et al. Haplotype-resolved diverse human genomes and integrated analysis of structural variation. *Science* 372, (2021).