

Long read pharmacogenomics (PGx)

Pharmacogenomics (PGx) is critically important to precision medicine, informing the use of medications at an individual level. PGx diplotyping relies on the ability to accurately detect distinct haplotypes as "star (*) alleles". Detection of these haplotypes with PacBio HiFi sequencing has several key benefits:

- **Phased variants** – Accurate phasing of variants is critical for PGx diplotyping. Long reads provide phase blocks that can fully span a PGx gene, allowing for complete phasing of variants into haplotypes for accurate star (*) allele assignment.
- **Direct observation of copy number changes** – Long reads will often fully span two adjacent *CYP2D6* alleles, allowing for direct observation of duplications or additional hybrid alleles on a haplotype.
- **Identification of novel haplotype sequences** – Long reads can clearly distinguish haplotypes, which when compared compared to database sequences, enable novel allele discovery with basepair precision.

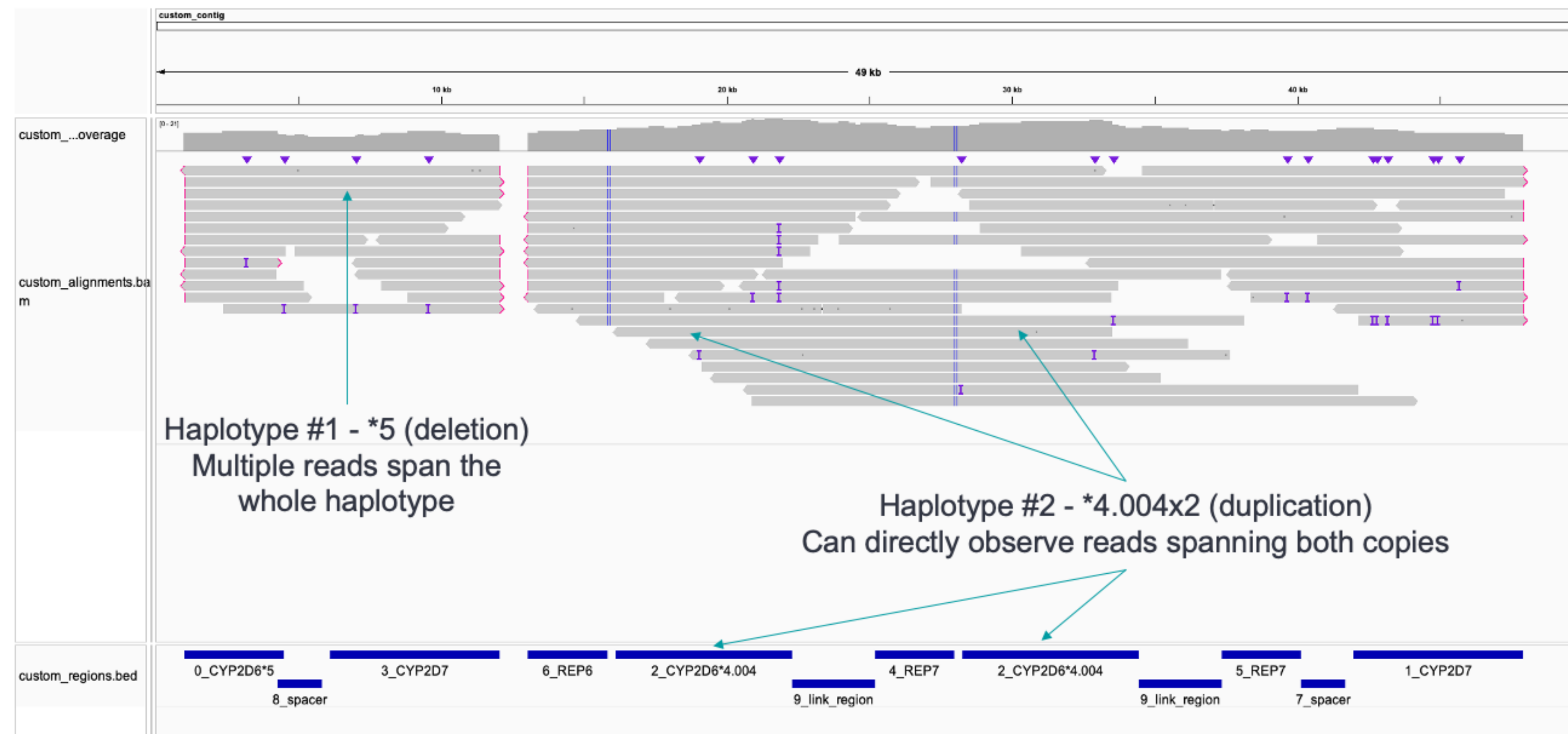


Figure 1. StarPhase custom *CYP2D6* diplotype visualization. StarPhase generated this custom IGV view to show a *CYP2D6**4.004x2 (duplication) / *5 (deletion) in NA20129 (HPRC dataset). The view includes a custom reference genome with both haplotype consensus and annotations indicating the location of alleles and nearby regions. Reads spanning the duplication event can be directly observed with ease.

StarPhase for long read pharmacogenomic diplotyping

StarPhase is a PGx diplotyper that leverages HiFi sequencing to accurately resolve PGx haplotypes. StarPhase provides several key benefits over existing PGx diplotyping tools:

- **Diploypes 21 CPIC Class A genes** – StarPhase matches haplotypes to those described by the CPIC, IMGT/HLA, and PharmVar databases.
- **Phase-aware diplotyping** – StarPhase correctly handles fully- and partially-phased genes.
- **Generates full-length consensus sequences** – For *HLA-A*, *HLA-B*, and *CYP2D6*, StarPhase generates a full-length DNA sequence. The tool reports 4-field haplotypes for HLA genes and sub-alleles (e.g., *4.004 instead of *4) for *CYP2D6*.
- **Provides visualization of complex haplotypes** – For *HLA-A*, *HLA-B*, and *CYP2D6*, StarPhase provides custom visualizations of the consensus haplotypes.

StarPhase provides high accuracy PGx diplotyping for 21 CPIC Class A genes

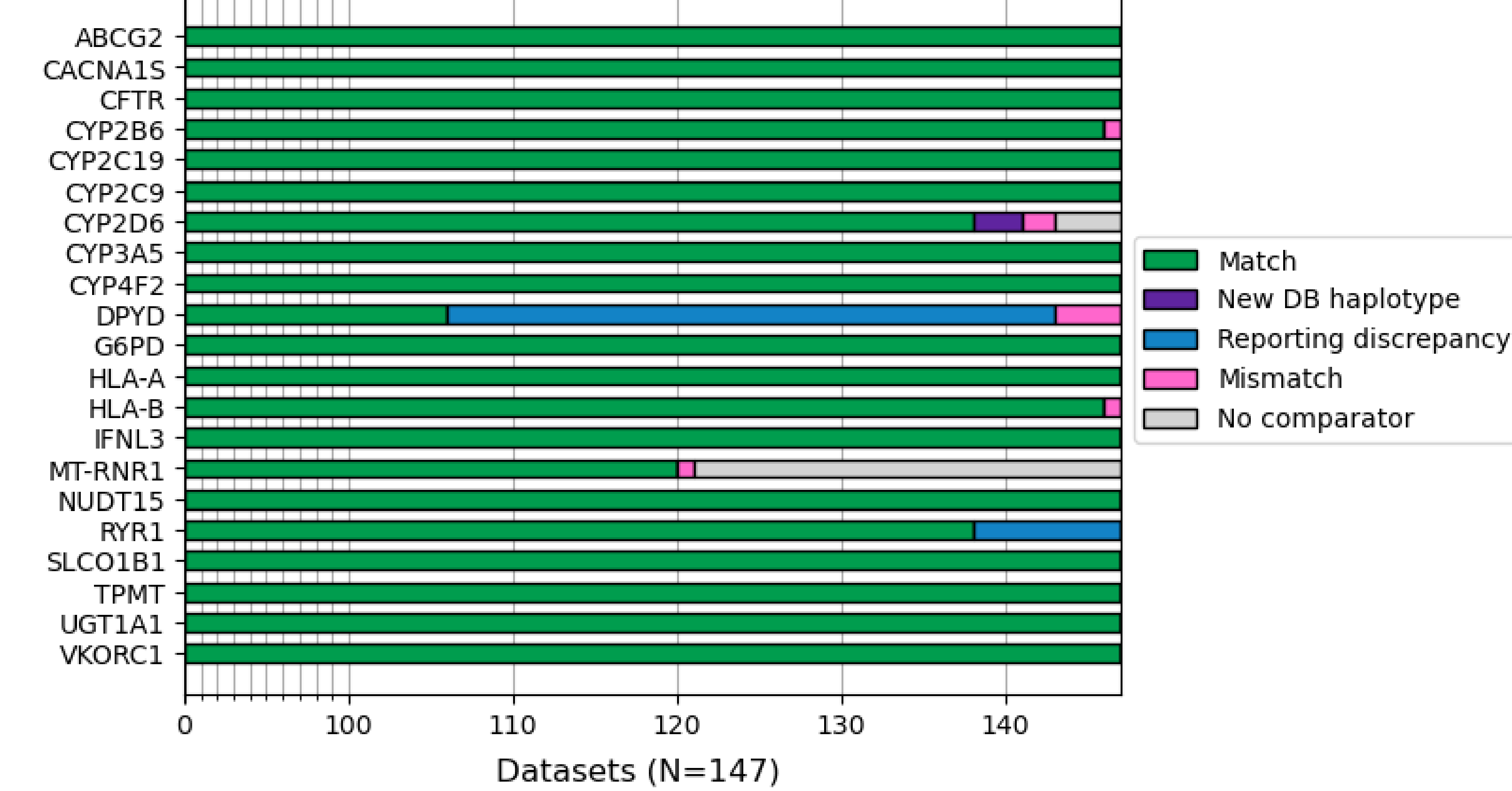


Figure 2. HPRC Comparisons. Evaluation of StarPhase diplotyping calls against 147 WGS datasets from the Human Pangenome Reference Consortium (HPRC). Comparator sets were generated from multiple sources¹⁻⁵. For each of the nine total mismatches, manual inspection of the data supported the StarPhase diplotype.

StarPhase provides updated PGx diploypes for reference datasets (GeT-RM)

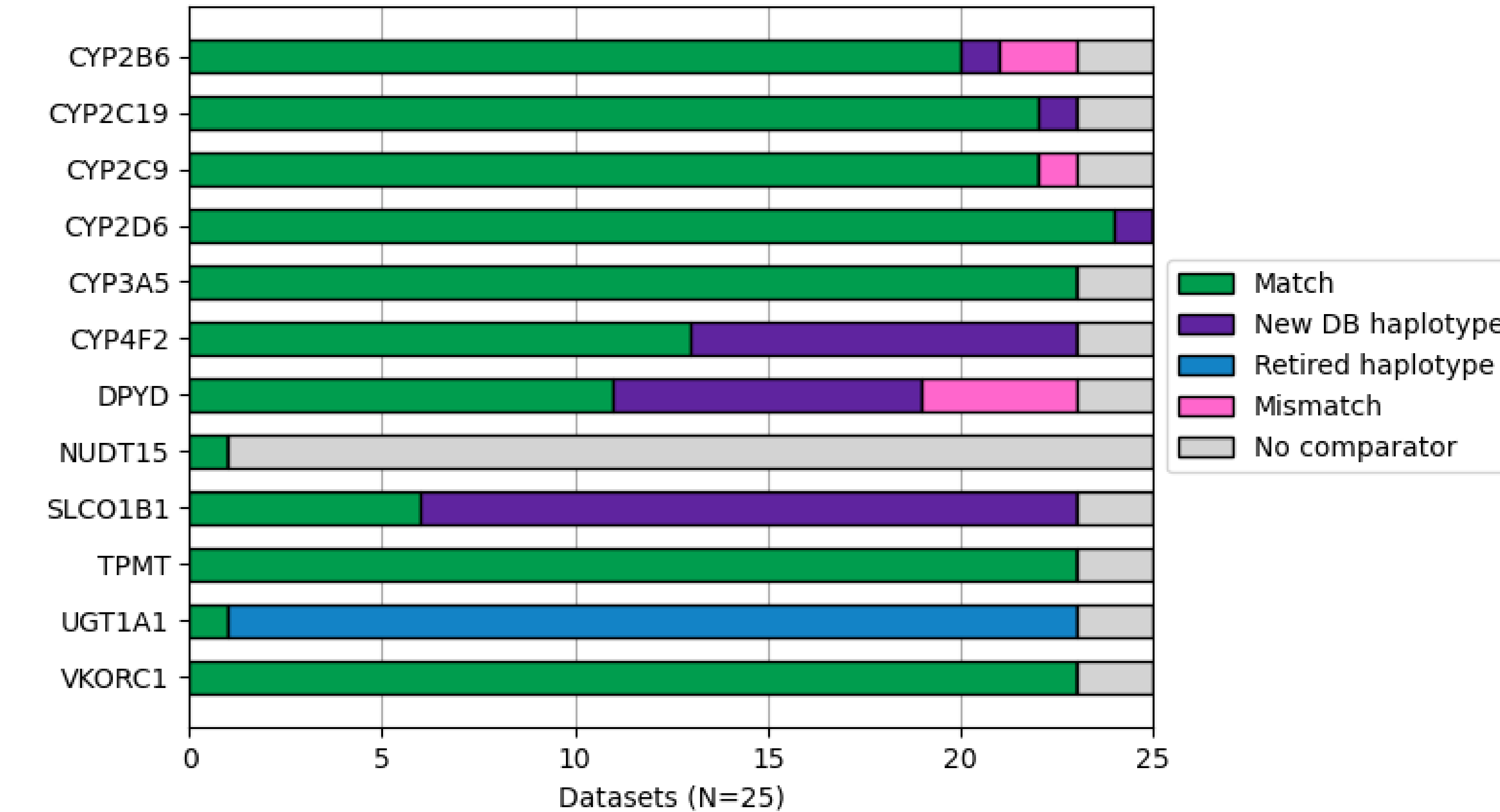


Figure 3. GeT-RM Comparisons. Evaluation of StarPhase diplotyping on 25 GeT-RM dataset. For all seven mismatches, manual inspection supported the StarPhase diplotype, and most had explanations for the mismatch (e.g., phasing limitations in the GeT-RM assay). StarPhase provides updated diploypes for many genes (purple and blue).

StarPhase diploypes match population expectations

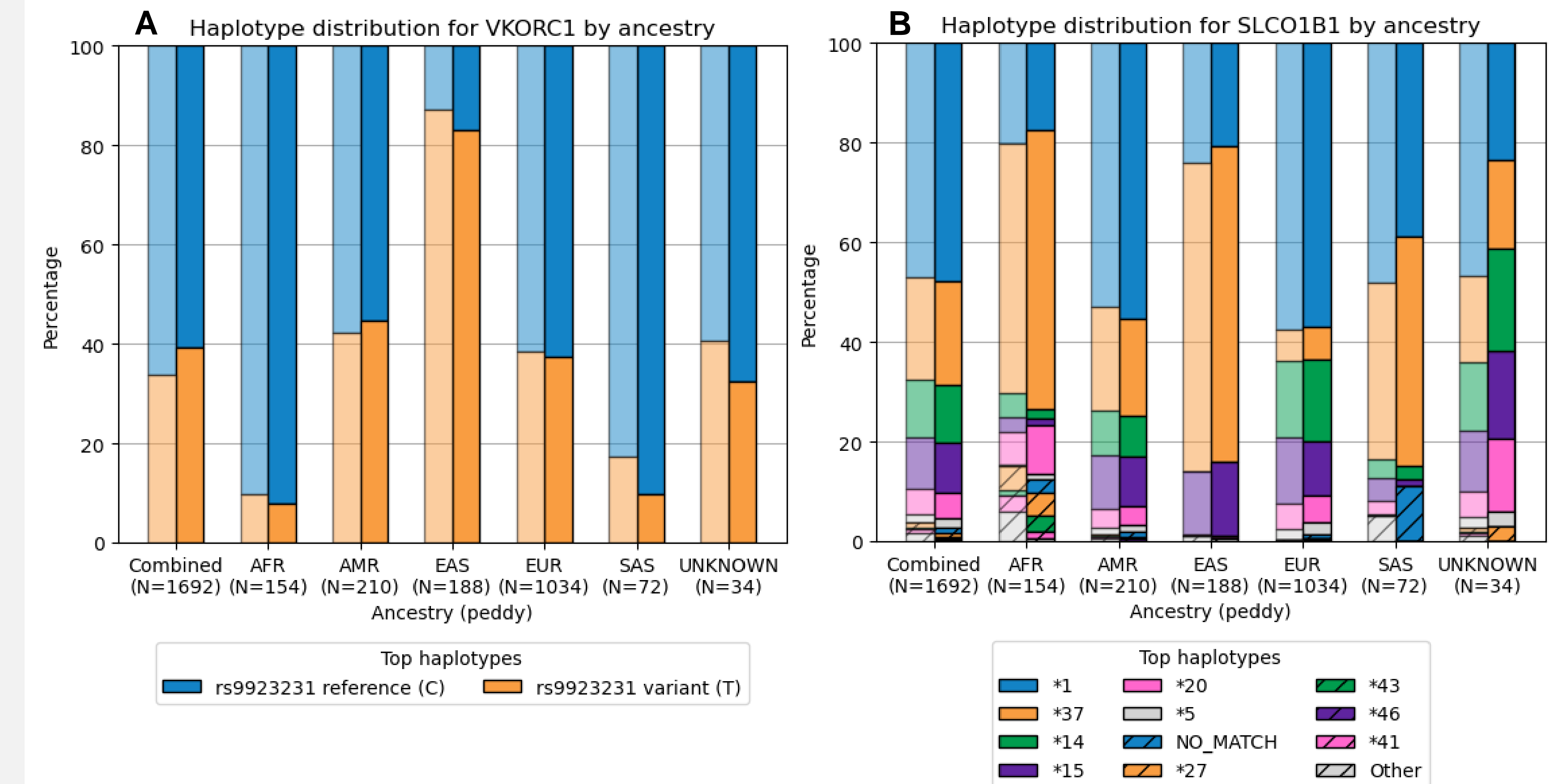
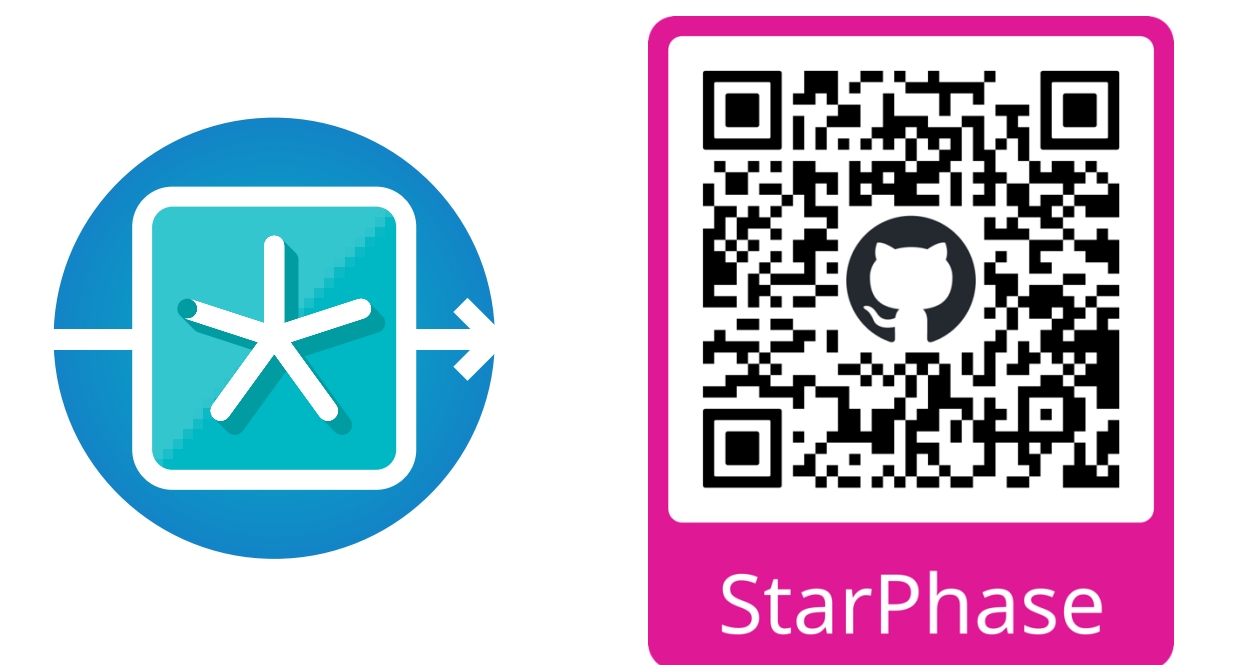


Figure 4. Population Comparisons. The population haplotype distributions for a multi-site cohort (N=846) for two genes: *VKORC1* (A) and *SLCO1B1* (B). The left faded bars show population expectations based on PharmCAT diplotyping from the AIOfUs project, while the right solid bars show the haplotype distributions from our StarPhase cohort.

Conclusion

StarPhase is a pharmacogenomic diplotyper for PacBio HiFi reads, providing:

- Accurate diplotyping of 21 Class A pharmacogenomic genes, including detailed 4-field HLA haplotypes and sub-alleles for *CYP2D6*
- Full-length consensus sequences for complex PGx genes
- Visualizations for complex PGx genes, enabling easier validation of novel haplotypes



Available on GitHub:

<https://github.com/PacificBiosciences/pb-StarPhase>

References

1. Behera, Sairam, et al. "Comprehensive and accurate genome analysis at scale using DRAGEN accelerated algorithms." *bioRxiv* (2024).
2. Chen, Xiao, et al. "Cyrius: accurate *CYP2D6* genotyping using whole-genome sequencing data." *The pharmacogenomics journal* 21.2 (2021): 251-261.
3. Lai, Sheng-Kai, et al. "A novel framework for human leukocyte antigen (HLA) genotyping using probe capture-based targeted next-generation sequencing and computational analysis." *Computational and Structural Biotechnology Journal* 23 (2024): 1562-1571.
4. Sangkuhl, Katrin, et al. "Pharmacogenomics clinical annotation tool (PharmCAT)." *Clinical Pharmacology & Therapeutics* 107.1 (2020): 203-210.
5. Wang, Ting, et al. "The Human Pangenome Project: a global resource to map genomic diversity." *Nature* 604.7906 (2022): 437-446.

Acknowledgements

The authors would like to thank our collaborators who provided data for the population comparisons: Genomic Answers for Kids (GA4K) program at Children's Mercy Kansas City and the Singapore Undiagnosed Disease Program.