StarPhase: Leveraging Long-Read Sequencing to Update Pharmacogenomic Benchmarks

P646

J. Matthew Holt, John Harting, Xiao Chen, Daniel Baker, Nina Gonzaludo, Zev Kronenberg, Christopher T. Saunders, Michael A. Eberle
PacBio, 1305 O'Brien Drive, Menlo Park, CA 94025

Long-read pharmacogenomics (PGx) with StarPhase

StarPhase is a PGx diplotyping tool that leverages long-read sequencing to resolve complex pharmacogenes. Key features include:

- **Diplotypes 21 CPIC Class A pharmacogenes** – Including complex genes *HLA-A*, *HLA-B*, and *CYP2D6*
- **Phase-aware diplotyping** – Generates haplotype-resolved PGx diplotypes using long reads
- **Full-length consensus sequences** – Generates complete genomic sequences for *HLA* genes and *CYP2D6*, enabling novel allele discovery at basepair resolution
- **Custom visualizations** – Creates sample-specific IGV visualizations to enable exploration of complex loci

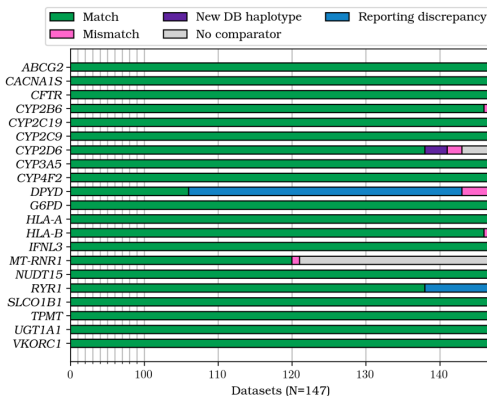### StarPhase generates high-accuracy PGx diplotype calls

Figure 1. **Comparison to HPRC benchmark.** The vast majority of diplotypes calls match the benchmark (98.1%). An additional 1.5% are caused by minor discrepancies or database updates. Manual inspection of the final nine diplotypes (0.3%) supports the StarPhase diplotypes calls.

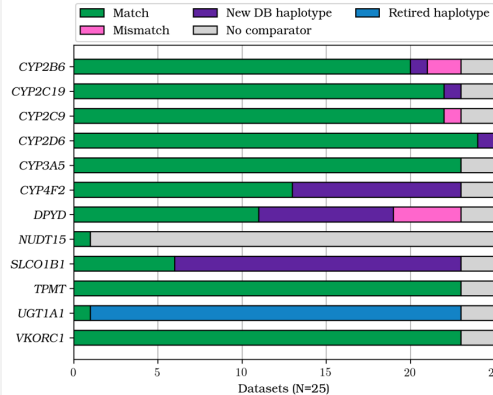## StarPhase updates PGx diplotype calls for GeT-RM samples

Figure 2. **Comparison to GeT-RM sample benchmark.** The majority of diplotypes calls match the GeT-RM call set (73.8%). Of the differences, 38 diplotypes (14.8%) match new haplotypes that were not defined when the benchmark was created. For *UGT1A1*, 22 diplotypes (8.6%) contain a retired haplotype, and are updated accordingly. Of the final seven mismatches, six can be explained by limitations of the benchmark assays. In total, StarPhase provided updated results for 26.2% of diplotypes in this GeT-RM sample benchmark.
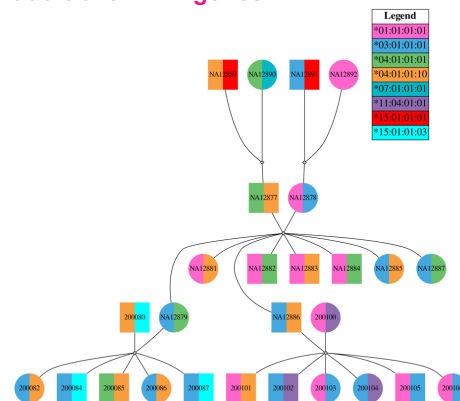
Extending the methods to nine additional HLA genes

Figure 3. *HLA-DRB1* **in CEPH pedigree 1463.** StarPhase has been extended to report nine more complex HLA genes. This figure shows the diplotypes calls for *HLA-DRB1* in CEPH pedigree 1463. StarPhase identifies each diplotype independently and generates no inheritance violations for the pedigree. Full list of additional HLA genes: *HLA-C, DPA1, DPB1, DQA1, DQB1, DRB1, DRB3, DRB4,* and *DRB5*.

StarPhase visualizations enhance interpretability of long-read PGx data
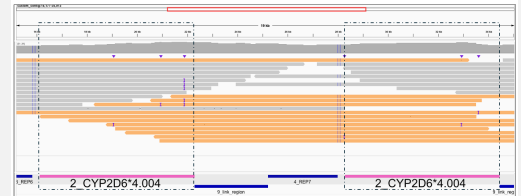

Figure 4. **Example *4.004x2 haplotype.** This IGV screenshot has been enhanced to highlight key evidence from a StarPhase *4.004x2 haplotype call. Assembled regions are labeled on the bottom track, with two *4.004 alleles highlighted in pink. Long reads that span both copies of *4.004 are highlighted in orange. These reads provide direct observational evidence of the duplication.

Conclusion

StarPhase is a long-read pharmacogenomic diplotyper that:

- Provides highly accurate diplotype results from long-read observations
- Provides refined PGx diplotypes for GeT-RM benchmark samples
- Generates full-length haplotype sequences and visualizations for complex pharmacogenes

StarPhase

Available on GitHub:
https://github.com/PacificBiosciences/pb-StarPhase
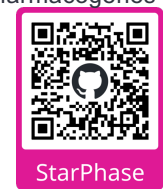
References

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.