Typing CYP2D6 star alleles from fully phased variants using PacBio HiFi reads

John Harting1, Zev Kronenberg1, Nina Gonzaludo1*, Jenny Ekholm1, Geoff Henno1, Edd Lee1

1PacBio, 1305 O’Brann Drive, Menlo Park, CA 94025, USA

Introduction

The CYP2D6 locus is known for its importance to pharmacogenomics as well as for its high diversity and complex genomic setting. Gene duplications, gene fusions, gene conversion, and large deletions are common at this locus. Resolving and phasing individual alleles without imputation requires long and highly accurate reads. We demonstrate and benchmark the accuracy of PacBio HiFi reads and the pbaa clustering algorithm for resolving these important loci.

- 22 Coriell samples
- 3 Amplicon primer design
- 1 SMRT Cell 8M
- Barcoded and pooled
- HiFi reads analyzed by pbaa and pCYP2D6Typer2
- Typing results validated against GetRM pharmacogenetics panel

HiFi Read Clustering

Figure 1. CYP2D6 Primer Design. Three amplicon design captures duplicates, hybrids, and deletion alleles in one assay.

Figure 2. CYP2D6 Workflow. HiFi reads are demultiplexed by sample barcode and converted to fastq. Pbaa deconvolves alleles and generates consensus. Star types and VCF called from consensus. Optionally color HiFi reads by cluster for visual inspection.

Figure 3. Pbaa Workflow and Visualization. (A) Clustering workflow. HiFi reads are assigned to guides and errors are masked within groups. Corrected reads are clustered and consensuses are generated. Post process filters separate pass/fail clusters. (B) Clustered and painted aligned HiFi reads in IGV. (C) Corrected HiFi read graph, colors match alignments with passing clusters in panel B.

Figure 4. Star Typing Workflow. Example call for NA17732. *2x2*35 (A) Call all variants with reference GRCh38. (B) Match core variants from pharmVar definitions. (C) For each allele, the star call is the first candidate when sorting matches by phenotypic impact, number of matched variants, and core number. (D) Assign SV status where appropriate (hybrid and duplicate alleles). (E) Assign alleles to haplotypes.

Results

Table 1. HI F CYP2D6 *Allele Calls. Published calls compared to calls generated from long read HiFi amplicons. Calls in red are improved with respect to published results.

Table 2. CYP2D6 Accuracy Titration. Pbaa consensus results are highly accurate over a wide range of coverage when compared to truth set.

Conclusion and Availability

Direct star-typing of CYP2D6 using clustered PacBio HiFi reads generates detailed and accurate results over a wide range of coverage.

Code Availability:
- CCS: https://ccs.how/
- Demux: https://lima.how/
- pbaa: https://github.com/PacificBiosciences/pbAA
- Star Typer: https://github.com/PacificBiosciences/apps-scripts/tree/master/CYP2D6tools

Resources:
- Sequencing Data: https://github.com/PacificBiosciences/apps-scripts/tree/master/CYP2D6tools
- GoT-RM: Multiply-Confirmed-Mutations-GoT-RM
- PharmVar: https://www.pharmvar.org/gene/CYP2D6

For research use only. Not for use in diagnostic procedures. © 2022 PacBio. All rights reserved.