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

John Harting[1], Zev Kronenberg[1], Nina Gonzaludo*[1], Jenny Ekholm[1], Geoff Henno[1], Edd Lee[1]

[1]PacBio, 1305 O'Brien Drive, Menlo Park, CA 94025, USA

Introduction

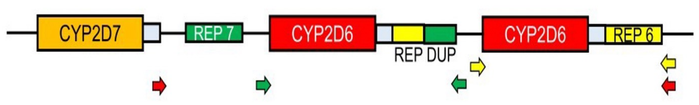
The *CYP2D6* locus is known for its importance to pharmacogenetics as well as for its high diversity and complex genomic setting.

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 *pbCYP2D6typer2*
- Typing results validated against GeT RM pharmacogenetics panel



➡ *5 allele primers: 5.1 kb amplicon

➡ Upstream dup primers: 8.6 or 10.2 kb amplicon

➡ Downstream primers: 8.2 kb amplicon

Qiao et al., 2019; Fukuda et al., 2005

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

Star-Typing Workflow

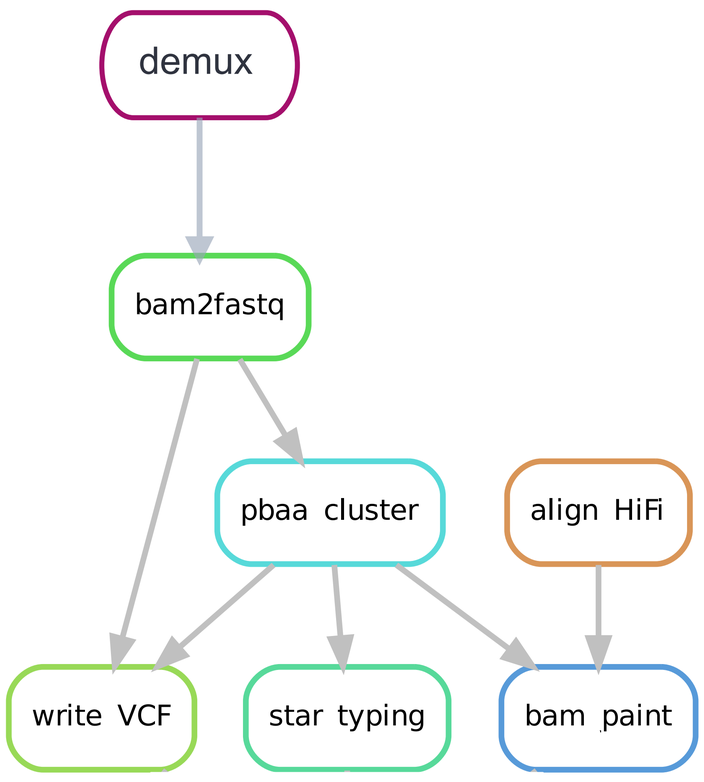


Figure 2. CYP2D6 Workflow. HiFi reads are demultiplexed by sample barcode and converted to fastq. Pbaa deconvolves alleles and generates consensus. Star types and VCF called from consensus. Optionally color HiFi reads by cluster for visual inspection.

HiFi Read Clustering

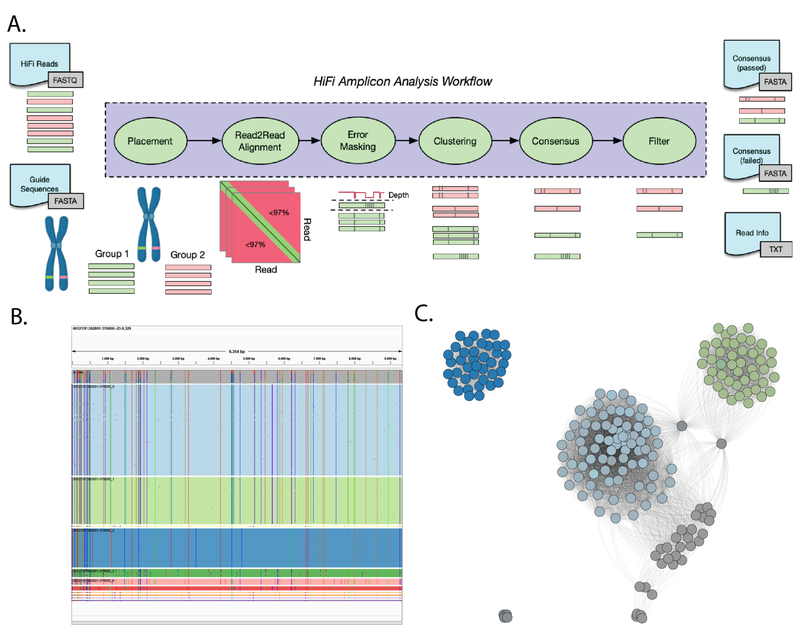


Figure 3. Pbaa Workflow and Visualization. (A) Clustering workflow. HiFi reads are assigned to guides and errors are masked within groups. Corrected reads are clustered and consensuses are generated. Post process filters separate pass/fail clusters. **(B)** Clustered and painted aligned HiFi reads in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

Star Allele Selection

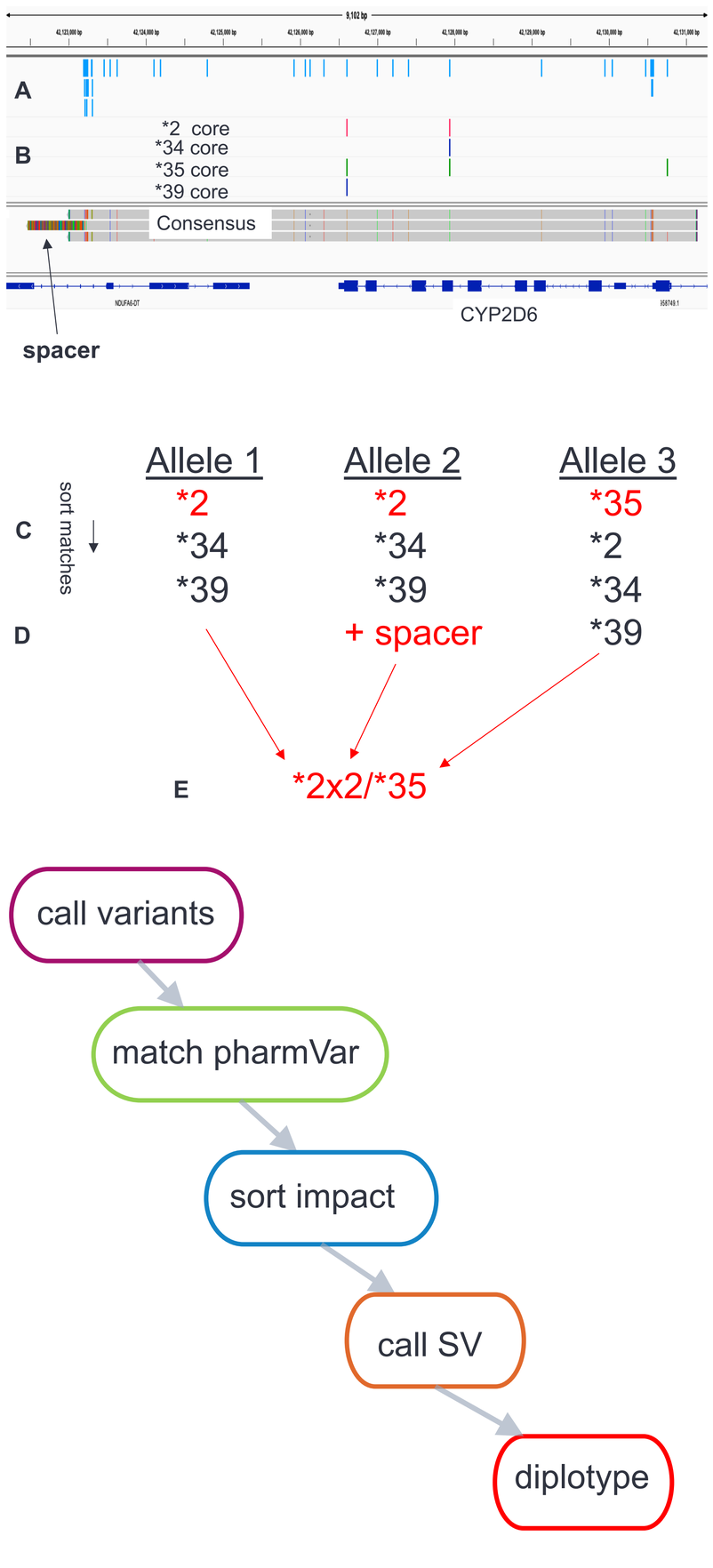


Figure 4. Star Typing Workflow. Example call for NA17232, ***2x2/*35 (A)** Call all variants with reference GRCh38. **(B)** Match core variants from pharmVar definitions. **(C)** For each allele, the star call is the first candidate when sorting matches by phenotypic impact, number of matched variants, and core number**. (D)** Assign SV status where appropriate (hybrid and duplicate alleles). **(E)** Assign alleles to haplotypes.

Results

Sample	CYP2D6 Reference	HiFi + pbaa Calling	Sample	CYP2D6 Reference	HiFi + pbaa Calling
NA02016	*2xN/*17	*2x2 /*17	NA17211	*2/*4	*2/*4
NA07439	*4xN/*41	*4x2/*41	NA17214	*2/*2	*2/*2
NA09301	Duplication	*1/*2x2	NA17215	*4/*41	*4/*41
NA12244	*35/*41	*35/*41	NA17217	*1/*41	*33/*41
NA16654	*10/*10	*10 + *36	NA17226	*4/*4	*4/*4 + *4.013
NA16688	*2/*10	*2/*10 + *36	NA17227	*1/*9	*1/*9
NA17020	*1/*10	*1/*10	NA17232	*2/*2xN	*2x2 /*35
NA17039	*2/*17	*2/*17	NA17244	DUP *4/*2A	*2x2/*4x2
NA17073	*1/*17	*1/*17	NA17276	*2/*5	*2/*5
NA17114	*1/*5	*1/*5	NA17282	*41/*41	*41/*41
NA17209	*1/*4	*1/*4 + *4.013	NA17300	*1/*6	*1/*6

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

	100x	200x	300x	400x	500x	1000x
TP	53	53	53	53	53	53
FN (filtered)	0	0	0	0	0	0
FN (missing)	0	0	0	0	0	0
FP	1	0	0	0	0	0
Accuracy	0.98	1.00	1.00	1.00	1.00	1.00
Precision	0.98	1.00	1.00	1.00	1.00	1.00
Recall	1.00	1.00	1.00	1.00	1.00	1.00
Avg. edit distance	0.02	0	0	0	0	0
Avg. PHRED QV	56	>56	>56	>56	>56	>56

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
- GeT- RM: Multiply-Confirmed-Mutations-GeT-RM
- PharmVar: https://www.pharmvar.org/gene/CYP2D6

P19.024.C