Full-Length Sequencing of *CYP2D6* Variants with PacBio HiFi Reads

Abstract #: eP426

Lei Zhu, Aaron Wenger, Josiah Wilcots, John Harting, Jonas Korlach, Primo Baybayan
PacBio, 1305 O'Brien Drive, Menlo Park, CA 94025

Introduction

CYP2D6 is a highly polymorphic gene with more than 130 known variants, including deletions, duplications, single nucleotide polymorphisms, and other types of variations (Butler, 2018; Black et al., 2012).

(Image cited from Black et al., 2012)

These variants affect the rate of metabolism of approximately 25% of commonly prescribed drugs in human (Owen et al., 2019). PacBio SMRT Sequencing has been proven to be an effective tool for the interrogation of *CYP2D6* variants (Qiao et al., 2016; Buermans et al., 2017). With PacBio's highly accurate long reads (HiFi), we have developed a streamlined end-to-end workflow for more accurate detection of highly polymorphic *CYP2D6* locus. This study demonstrates the advantages of HiFi reads for full-length sequencing of *CYP2D6*, previously annotated by other technologies.

Methods

- Twenty-two Coriell pharmacogenomics samples with *CYP2D6* variants were amplified with long-range PCR.
- The primer sets for the amplification of upstream duplications, downstream *CYP2D6* genes, and for the *5 allele shown below were adapted from Qiao et al. (2019) and Fukuda et al. (2005).

- A two-step PCR with barcoded M13 Primers was used to enable pooling of 22 samples for a single SMRTbell library preparation, which was sequenced on the PacBio Sequel II and IIe Systems.
- HiFi reads (>QV20) were demultiplexed on SMRT Link v10.0 and clustered into haplotypes. The consensus reads of each haplotype were produced using the "pbaa" amplicon analysis from bioconda (https://github.com/PacificBiosciences/pbAA) and mapped to the human reference genome GRCh38 for the assignment of *CYP2D6* types.

Results

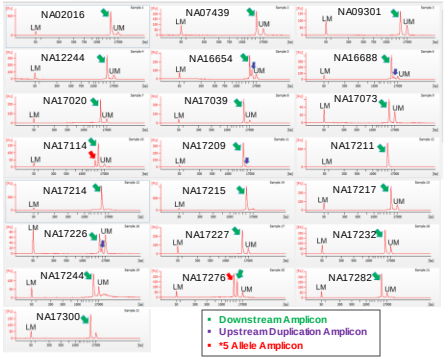

Figure 1. Barcoded *CYP2D6* Amplicons.
LM, lower marker; UM, upper marker. The sample name on each electropherogram.

More than 1,600,000 full-length HiFi reads were generated from one SMRT Cell 8M with an average read length of 8.2 kb and a median HiFi quality > 99.9% (QV31).

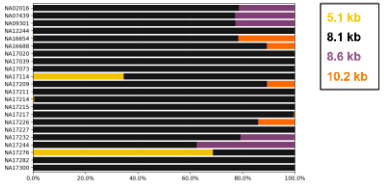

Figure 2. Nearly all (>99%) demultiplexed reads were on target to *CYP2D6* locus. 8.1 kb reads correspond to downstream gene; 8.6 kb or 10.2 kb (*36 allele) reads indicate upstream duplications; 5.1 kb reads indicate *5 allele (complete deletion).

Table 1. HiFi sequencing provides single-base resolution for diplotype calling.

Sample	Coriell *CYP2D6* diplotype[1]	pbaa calling	Sample	Coriell *CYP2D6* diplotype[1]	pbaa Calling
NA02016	*2×N/*17	*2×2/*17	NA17211	*2/*4	*2/*4
NA07439	*4×N/*41	*4×2/*41	NA17214	*2/*2	*2/*2
NA09301	Duplication	*1/*2×2	NA17215	*4/*41	*4/*41
NA12244	*35/*41	*35/*41	NA17217	*1/*41	*33/*41
NA16654	*10/*10	*10 + *36	NA17226	*4/*4	*4 + *36
NA16688	*2/*10	*2/*10 + *36	NA17227	*1/*9	*1/*9
NA17020	*1/*10	*1/*10	NA17232	*2/*2×N	*2×2/*35
NA17039	*2/*17	*2/*17	NA17244	DUP *4/*2A	*2/*4
NA17073	*1/*17	*1/*17	NA17276	*2/*5	*2
NA17114	*1/*5	*1	NA17282	*41/*41	*41/*41
NA17209	*1/*4	*1/*4 + *36	NA17300	*1/*6	*1/*6

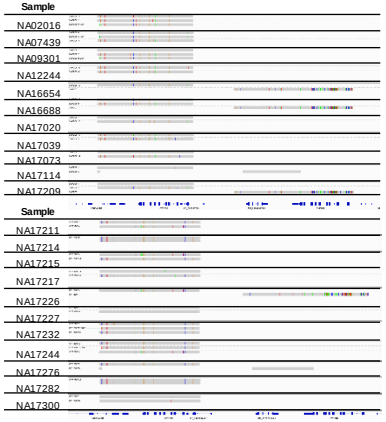

Figure 3. *CYP2D6* genotyping with PacBio HiFi reads.
HiFi reads provide full length sequences of each allele.

Discussion

- The PCR amplification of the *CYP2D6* region was robust and specific. And HiFi sequencing provides single-base resolution for diplotype calling.
- Coriell sample NA17217 was identified by microarray as *1/*41. HiFi sequencing produced a diplotype of *33/*41.
- For sample NA17232, an additional SNP was detected, making the diplotype to be *2x2/*35 instead of *2/*2xN.
- In addition, for 4 of 22 samples, including NA16654, NA16688, NA17209, and NA17226, HiFi sequencing identified duplications missed by microarray or real-time PCR.

Conclusions

- We have demonstrated an end-to-end workflow (amplification to analysis) for the targeted sequencing of *CYP2D6*.
- The workflow allows full-length sequencing of 5 kb, 8 kb, and 10 kb amplicons with consensus accuracy of >QV30.
- HiFi reads revealed diplotypes that were not well characterized by other technologies.

References

Black, J. L., et al. (2012). Frequency of undetected CYP2D6 hybrid genes in clinical samples: impact on phenotype prediction. *Drug metabolism and disposition: the biological fate of chemicals*, 40(1), 111–119.

Buermans, H. P., et al. (2017). Flexible and Scalable Full-Length CYP2D6 Long Amplicon PacBio Sequencing. *Human mutation*, 38(3), 310–316.

Butler M. G. (2018). Pharmacogenetics and Psychiatric Care: A Review and Commentary. *Journal of mental health & clinical psychology*, 2(2), 17–24.

Fukuda, T., et al. (2005). "Novel structure of the CYP2D6 gene that confuses genotyping for the CYP2D6*5 allele." Drug metabolism and pharmacokinetics vol. 20,5: 345-50.

Owen, R. Pet al. (2009). Cytochrome P450 2D6. *Pharmacogenetics and genomics*, 19(7), 559–562.

Qiao, W., et al. (2016). Long-Read Single Molecule Real-Time Full Gene Sequencing of Cytochrome P450-2D6. *Human mutation*, 37(3), 315–323.

Qiao, W., et al. (2019). Integrated CYP2D6 interrogation for multiethnic copy number and tandem allele detection. *Pharmacogenomics*, 20(1), 9–20.

Acknowledgements

We would like to thank Janet Ziegle for providing the Coriell pharmacogenomics samples.