Long-Read Amplicon Sequencing of the Polymorphic *CYP2D6* Locus

Abstract #: 2021-A-1673-ASHG
Lei Zhu, John Harting, Aaron Wenger, Zev Kronenberg, Nina Gonzaludo, 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 SNPs, duplications, deletions, and other types of variations (Butler, 2018; Black et al., 2012). These variants affect the rate of metabolism of approximately 25% of most prescribed drugs in human individuals (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).

Using a long-amplicon sequencing approach that is simple and cost effective, we have developed a streamlined end-to-end workflow for the detection of highly polymorphic *CYP2D6* locus with PacBio both long and accurate HiFi reads. This study demonstrates that HiFi sequencing enables generation of full-length sequences of *CYP2D6* alleles for accurate diplotype calling.

Materials and Methods

Figure 1. End-to-end workflow of *CYP2D6* sequencing.

- Two-step PCR approach was developed for this assay. In PCR1, M13-tailed gene-specific primer sets were used for the amplification of upstream duplications, downstream *CYP2D6* genes, and for the *5 allele (complete deletion), adapted from Qiao et al. (2019) and Fukuda et al. (2005). PCR2 amplifies PCR1 products with barcoded M13 primers.
- Twenty-two Coriell pharmacogenomic reference samples with *CYP2D6* variants were tested.
- Tested forty-one human saliva samples using the same assay.
- Barcoded amplicons were pooled for a single SMRTbell library preparation and sequenced on the PacBio Sequel II and IIe Systems.
- HiFi reads (>QV20) were demultiplexed on SMRT Link v10.1. The consensus reads of each haplotype were produced using the "pbaa" amplicon analysis from bioconda (https://github.com/PacificBiosciences/pbAA) and assigned *CYP2D6* diplotypes.

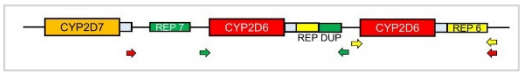

Figure 2. Location of M13-tailed *CYP2D6*-specific primers. Red arrows, pair of *5 allele primers; green arrows, pair of Upstream Dup primers; yellow arrows, pair of downstream primers.

Results

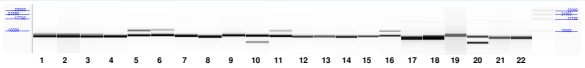

Figure 3. Barcoded amplicons from 22 Coriell reference samples analyzed on Femto Pulse.

Sample Number	Sample Name	Coriell *CYP2D6* diplotype	pbaa calling	Sampler Number	Sample Name	Coriell *CYP2D6* diplotype	pbaa calling
1	NA02016	*2×N/*17	*2×2/*17	12	NA17211	*2/*4	*2/*4
2	NA07439	*4×N/*41	*4×2/*41	13	NA17214	*2/*2	*2/*2
3	NA09301	Duplication	*1/*2×2	14	NA17215	*4/*41	*4/*41
4	NA12244	*35/*41	*35/*41	15	NA17217	*1/*41	*33/*41
5	NA16654	*10/*10	*10 + *36	16	NA17226	*4/*4	*4 + *36
6	NA16688	*2/*10	*2/*10 + *36	17	NA17227	*1/*9	*1/*9
7	NA17020	*1/*10	*1/*10	18	NA17232	*2/*2×N	*2×2/*35
8	NA17039	*2/*17	*2/*17	19	NA17244	DUP *4/*2A	*2/*4
9	NA17073	*1/*17	*1/*17	20	NA17276	*2/*5	*2
10	NA17114	*1/*5	*1	21	NA17282	*41/*41	*41/*41
11	NA17209	*1/*4	*1/*4 + *36	22	NA17300	*1/*6	*1/*6

Table 1. Diplotype calling of Coriell reference samples.

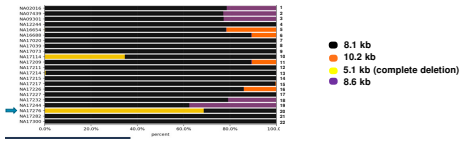

Figure 4. Proportion of total HiFi reads for each amplicon size per sample (e.g., NA17276 with 72% reads are *5 allele and 28% reads are normal 8.1 kb size allele). >99% demuxed HiFi reads from 22 Coriell samples were on target to *CYP2D6*.

Figure 5. The 2-step PCR assay was tested on 41 human saliva samples. The resulting amplicons from 41 samples are specific.

Discussion

- The 2-step barcoded PCR assay for *CYP2D6* interrogation is robust and specific for both reference and saliva samples.
- HiFi sequencing detected a SNP in Coriell sample NA17217, which is resolved to be *33/*41, instead of *1/*41 previously annotated by microarray.
- Similarly, a SNP was detected in sample NA17232, resulting in a diplotype call of *2x2/*35, instead of *2/*2xN.
- For 4 samples (NA16654, NA16688, NA17209, and NA17226), HiFi sequencing identified duplications (*36 allele), not detected by microarray or real-time PCR.

Conclusion

- PacBio provides an end-to-end workflow, from PCR to diplotype calling, for the characterization of *CYP2D6* variants.
- HiFi amplicon sequencing generates full-length 5-10 kb reads ≥Q30.
- HiFi reads provide base-level resolution, revealing diplotypes that were not detected 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. P., et 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

The authors would like to thank everyone who helped on analysis and review the poster.