PacBio

Application note

HiFi amplicon sequencing for pharmacogenetics: *CYP2D6*

Introduction

CYP2D6 is widely studied in the field of pharmacogenomics (PGx) since it directly impacts the metabolism of ~20% of the most prescribed medications, including antidepressants, opioids, and cancer drugs. However, *CYP2D6* is particularly challenging to study due to high levels of polymorphisms and structural variants (SVs), including complete gene deletions and duplications, tandem alleles, hybrids with the *CYP2D7* pseudogene, and gene conversion events (Nofziger et al., 2020). Additionally, this region is difficult to capture with short-read sequencing and other genotyping methods due to high sequence identity with neighboring pseudogenes *CYP2D7* and *CYP2D8* directly upstream, which have 97% and 92% exonic sequence similarity to *CYP2D6*, respectively; see Figure 1.

Assay designers typically need to use multiple workflows and platforms to fully capture variation, with specificity, in *CYP2D6*. While there are over 130 defined haplotypes, or star (*) alleles, of *CYP2D6* in PharmVar (Pharmacogene Variation Consortium, 2022),

© 2022 PacBio. All rights reserved. Research use only. Not for use in diagnostic procedures.
PN 102-326-548 REV01 OCT2022

PacBio

a)

b)

Figure 1. An IGV view of a) paired-end 100x short-read Illumina whole exome sequencing data mapping to *CYP2D6*, as well as off-target mapping to *CYP2D7*, a neighboring pseudogene, with coverage gaps in a segmental duplication region in between. b) Amplicon-based *CYP2D6* captured by PacBio HiFi reads in b) show specificity and full phasing across more than 40 heterozygous variants in *CYP2D6* and no misalignment to *CYP2D7*.

existing platforms that only capture a predefined set of single nucleotide variants make it difficult to identify novel or rare haplotypes and may rely on imputation. This makes it difficult to detect potentially actionable variants, particularly SVs, in underserved populations such as those with non-European ancestry (Del Tredici et al., 2018). Additionally, depending on which variants are assayed, some platforms may be unable to distinguish between certain haplotypes. Also, such platforms may only detect specific, known SVs and copy number variants, for example relying on only exon 9 and/or intron 2 to detect hybrid genes and copy number status. While short-read sequencing panels allow for higher resolution genotyping, it can be difficult or impossible to fully phase variants over kilobase-long stretches of sequence without imputation.

Variability in platform and assay design can have an impact on patient care (Pratt et al., 2021). Haplotypes of pharmacogenetic loci like *CYP2D6* are the basis for star (*) allele assignment. For *CYP2D6*, star (*) alleles are translated into an Activity Score, which predicts an individual's metabolizer status. Errors and ambiguities in haplotype assignment, which can stem from missing or incorrect variant calls and phasing, can result in an incorrect clinical phenotype prediction (Pratt et al.). This may prevent a provider from choosing the optimal drug and/or dose for a potential patient based on prescribing guidelines or clinical decision support.

With PacBio® HiFi sequencing, *CYP2D6* variation can be captured using highly accurate long reads on a single platform. Amplicon-based HiFi sequencing fully and directly resolves and phases complex loci like *CYP2D6* without assembly or inference, allowing for ancestry-agnostic haplotype assignment.

In this Application Note, we propose a streamlined PCR amplicon workflow that characterizes full-length *CYP2D6* alleles with PacBio HiFi sequencing. The proposed workflow includes analysis tools for haplotype resolution and star (*) allele calling. *Note that the PCR assay design, including primer sequences and PCR conditions, is intended as an example workflow, and is neither optimized nor supported by PacBio.*

Workflow overview

Here, we present an amplicon sequencing workflow for characterization of full-length *CYP2D6* alleles from samples of 100 ng of human genomic DNA (gDNA). In the first-round PCR, LA Taq DNA polymerase (Takara: RR02AG) and two primer pairs listed in Table 1 are used to generate 9 kb and 12 kb PCR products. In the second-round PCR, Barcoded M13 primers are used to barcode each sample. The resulting barcoded PCR products are pooled for SMRTbell® library construction and sequenced on a PacBio Sequel® II or IIe system.

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

PacBio

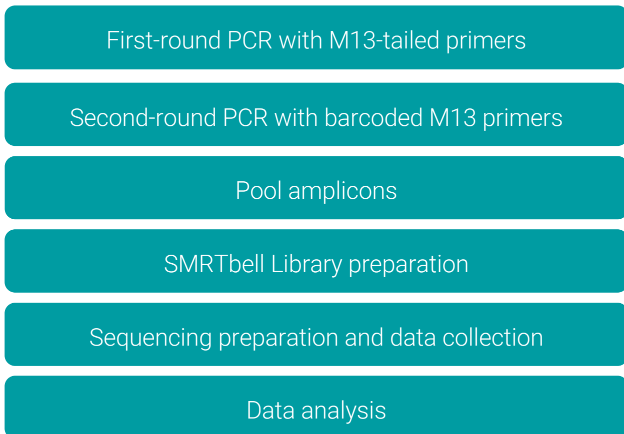

Figure 2. Overview of amplicon-based workflow.

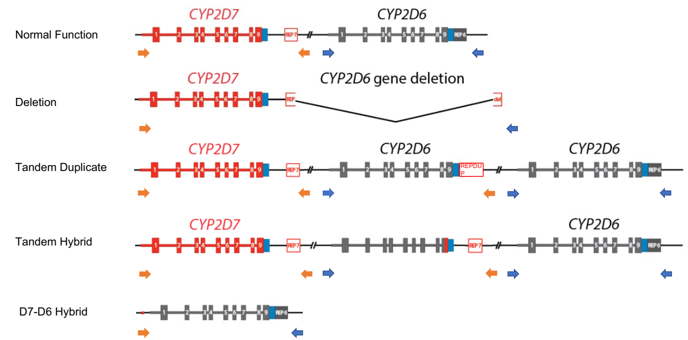

Figure 3. Primers are designed to capture structural variation affecting the *CYP2D6* locus. Primers upstream and downstream of CYP2D7 (orange) and CYP2D6 (blue) amplify both genes in wild-type samples, while deletion, duplication and hybrid alleles are amplified by combinations of upstream and downstream primers.

Multiplexing approach

This assay uses a two-step PCR workflow with barcodes added in the second PCR using PacBio's Barcoded M13 primer plate (102-135-500). This product supports up to 384 unique sample barcodes with dual indexing. Samples are pooled prior to SMRTbell library construction following PacBio guidance and the library is run on a single SMRT® Cell 8M. In this document and companion data release, we demonstrate multiplexing of 22 reference samples. The number of samples that can be pooled and run on a single SMRT Cell may be increased for a high-throughput assay.

Amplicon design

Table 1 lists the sequences of the forward and reverse M13-tailed primers recommended for amplification of *CYP2D6*. The primer set includes two primer pairs: 1. A_Fwd - Downstream_Rev and 2. H_Fwd - Upstream_Rev. The two pairs allow for the interrogation of 1) full gene plus downstream region of *CYP2D6*, 2) *CYP2D6* upstream duplication, 3) *CYP2D7* full gene, 4) *CYP2D6-CYP2D7* hybrids, and 5) *CYP2D7-CYP2D6* hybrids. Expected amplicon sizes range from 9 kb to 12 kb.

PCR conditions

The PCR cycling conditions and recommended reaction volumes for first-round PCR are described in Table 2 and 3, respectively. PCR reaction volumes for second-round PCR are in Table 4 and PCR conditions are in Procedure & Checklist - Preparing SMRTbell libraries using PacBio barcoded M13 primers for multiplex SMRT® sequencing (see Resources) under Barcoded M13 program. The final products (barcoded amplicons) are purified with 1x AMPure® PB beads, pooled using equal mass pooling, and finally concentrated with 1x AMPure PB beads for SMRTbell library prep following Procedure & Checklist - Preparing multiplexed amplicon libraries using SMRTbell prep kit 3.0

SMRT sequencing

Sample setup and sequencing on the Sequel IIe following guidance in SMRT® Link v11, using Binding Kit 3.2 (102-333-300), Sequel II sequencing kit 2.0 (101-820-200), and SMRT Cell 8M (101-389-001).

M13 Tailed Primer for *CYP2D6*	
Primer A_Fwd	**/5AmMC6/GTAAAACGACGGCCAGT**TCACCCCCAGCGGACTTATCAACC
Primer Downstream_Rev	**/5AmMC6/CAGGAAACAGCTATGAC**CAGGCATGAGCTAAGGCACCCAGA
Primer H_Fwd	**/5AmMC6/GTAAAACGACGGCCAGT**TCCGACCAGGCCTTTCTACCAC
Primer Upstream_Rev	**/5AmMC6/CAGGAAACAGCTATGAC**CGGCAGTGGTCAGCTAATGAC

Table 1. M13-tailed primers for genotyping *CYP2D6*. Each oligo contains a 5' M13 sequence (in bold), and forward or reverse primer sequences.

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

Cycles	Temperature	Time
1x	94°C	1 min
20x	94°C	30 sec
	57°C	30 sec
	72°C	10 min
1x	72°C	5 min
	4°C	hold

Table 2. PCR components and thermocycler conditions for first-round PCR.

Component	Volume per sample (µL)
2X GC buffer I	10.0 µL
dNTP Mixture (2.5 mM each)	3.2 µL
Takara LA Taq (5 U/µL)	0.2 µL
10 uM each primer	0.4 µL (each primer)
Human DNA (100 ng/µL)	1.0 µL
Water	4.0 µL
Total	**20.0 µL**

Table 3. Reaction volumes for first-round PCR

Component	Volume (µL)
2X GC buffer I	25.0
dNTP Mixture (2.5 mM each)	8.0
Takara LA Taq (5 U/µL)	0.5
Barcoded M13 primer pair	1.0
AMPure-purified round 1 PCR product	14.5
Water	1.0
Total	**50.0 µL**

Table 4. Reaction volumes for second-round PCR.

Data analysis

Due to the complexity of the *CYP2D6* locus, standard variant-calling workflows that include mapping reads to a reference prior to variant calling do not produce accurate results for all alleles. Instead, we clustered the reads for each sample to produce high quality consensus sequences, which was used for star (*) allele calls.

Resolving haplotypes

To capitalize on the accuracy of HiFi data, we used an amplicon analysis tool called PacBio Amplicon Analysis (pbAA v 1.0.3), which uses a sequence clustering algorithm to rapidly deconvolve a mixture of haplotypes. When used after standard demultiplexing in SMRT Link v10.2 or later or at the command-line with lima v2.3+, pbAA can effectively separate alleles with one or many variants, including SNVs and large indels contained within the target region for haplotype resolution. This enables precise diplotyping and star (*) allele classification in complicated loci such as *CYP2D6*.

Accurate star (*) allele assignment

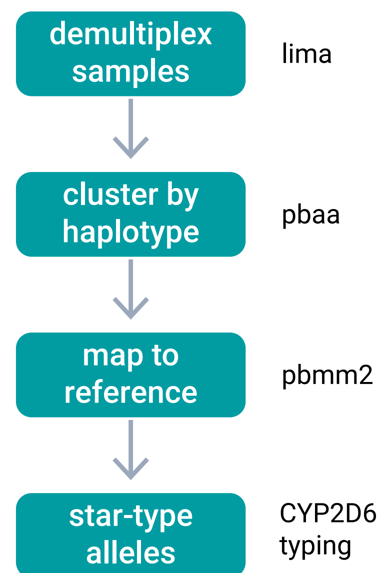

Figure 4. Analysis workflow for star (*) allele typing of *CYP2D6*.

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

Access PacBio Hifi analysis tools for *CYP2D6*, amplicon sequencing, barcoding, and more at https://github.com/PacificBiosciences.

Once *CYP2D6* haplotypes are resolved with pbAA, the phased consensus amplicons can be used to unambiguously determine star (*) allele diplotypes. Briefly, variants are matched from each unique *CYP2D6* amplicon to star (*) allele definitions based on those cataloged by PharmVar. Without using imputation, phased HiFi data are used to identify structural breakpoints for characterization of tandem arrangements (such as *CYP2D6*36+*10)*, deletions, and hybrid alleles.

Alleles are then sorted into haplotypes and assigned star (*) alleles based on PharmVar definitions, which enables assignment of diplotypes and translation into phenotype.

Analysis results

Using 22 Coriell GeT-RM samples (Pratt et al., 2016; Gaedigk et al., 2019) we performed the amplicon workflow and analysis described here to obtain star (*) allele diplotypes. Diplotypes were fully concordant with GeT-RM consensus calls for all samples.

This example end-to-end workflow uses full-length HiFi sequencing of PCR amplicons to produce complete and unambiguous characterization of *CYP2D6*.

See for yourself

Demo data for the 22 GeT-RM samples produced using the primers described in this document can be found at https://www.pacb.com/connect/datasets/.

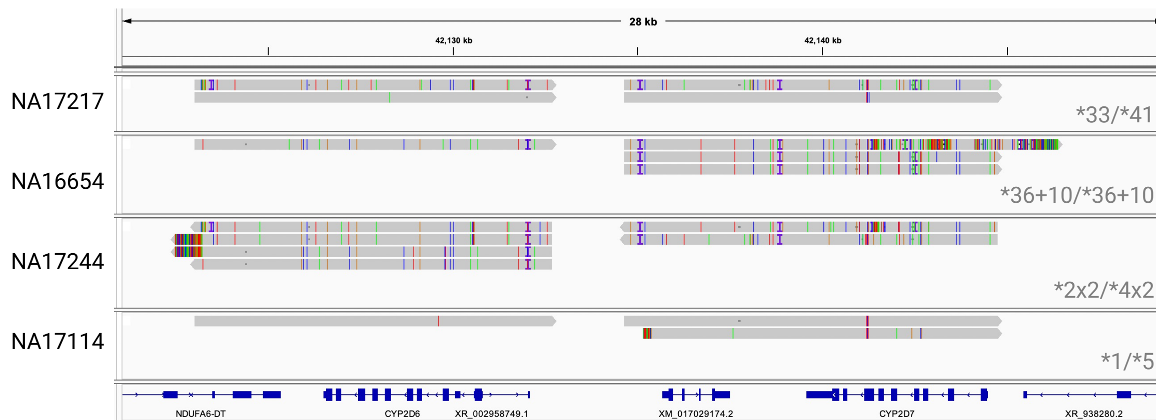

Figure 5. Example amplicon sequences covering samples with complex diplotypes. NA17217 covers less commonly-assayed variants, such as *33. Samples also included *CYP2D6-CYP2D7* hybrids in tandem arrangement (NA16654), double duplications (NA17244), and *CYP2D6* full gene deletion (NA17114).

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

Resources and References

Reagent and consumable for *CYP2D6* amplification (also see Procedure & checklist in Resources)

Products	Vendor	Product #
M13 tailed PCR primers	Any MLS	-
Barcoded M13 primer plate	PacBio	102-135-500
SMRTbell prep kit 3.0	PacBio	102-182-700
LA Taq DNA Polymerase kit (including buffer and dNTP)	Takara	RR02AG

Resources

Preparing multiplexed amplicon libraries using SMRTbell prep kit 3.0

Preparing SMRTbell libraries using PacBio barcoded M13 primers for multiplex SMRT sequencing

Amplicon Best Practices guide

PacBio GitHub Repositories

Acknowledgements

We acknowledge Andrea Gaedigk, PhD (Children's Mercy Research Institute Kansas City and Director, PharmVar Consortium) for primer design and feedback.

References

Del Tredici A.L., et al. (2018) Frequency of *CYP2D6* Alleles Including Structural Variants in the United States. *Frontiers in Pharmacology*, 5(9),305.

Gaedigk A., et al. (2019) Characterization of Reference Materials for Genetic Testing of *CYP2D6* Alleles: A GeT-RM Collaborative Project. *Journal of Molecular Diagnostics*, 21(6),1034-1052.

Nofziger C., et al. (2020) PharmVar GeneFocus: *CYP2D6*. *Clinical Pharmacology & Theraputics*, 107(1):154-170.

"Pharmacogene Variation Consortium (PharmVar)," last accessed 4 June 2022, www.PharmVar.org

Pratt V.M., et al. (2021) Recommendations for Clinical *CYP2D6* Genotyping Allele Selection: A Joint Consensus Recommendation of the Association for Molecular Pathology, College of American Pathologists, Dutch Pharmacogenetics Working Group of the Royal Dutch Pharmacists Association, and the European Society for Pharmacogenomics and Personalized Therapy. *Journal of Molecular Diagnostics*, 23(9),1047-1064.

Pratt V.M., (2016) Characterization of 137 Genomic DNA Reference Materials for 28 Pharmacogenetic Genes: A GeT-RM Collaborative Project. *Journal of Molecular Diagnostics*, 18(1), 109-23.

Research use only. Not for use in diagnostic procedures. © 2022 Pacific Biosciences of California, Inc. ("PacBio"). All rights reserved. Information in this document is subject to change without notice. PacBio assumes no responsibility for any errors or omissions in this document. Certain notices, terms, conditions and/or use restrictions may pertain to your use of PacBio products and/or third-party products. Refer to the applicable PacBio terms and conditions of sale and to the applicable license terms at pacb.com/license. Pacific Biosciences, the PacBio logo, PacBio, Circulomics, Omniome, SMRT, SMRTbell, Iso-Seq, Sequel, Nanobind, and SBB are trademarks of PacBio. All other trademarks are the sole property of their respective owner.