Development and Optimization of a 43 Gene Pharmacogenomic Panel Using Enrichment-Based Capture and PacBio HiFi Sequencing

Abstract #: 2021-A-1822-ASHG

Daniel Portik[1], Ting Hon[1], Josiah Wilcots[1], Nina Gonzaludo[1], Yao Yang[2,3], Nathan A. Hammond[3], Zev Kronenberg[1], Nathaniel Watson[3], John Harting[1], Euan Ashley[4], Janet Ziegle[1], Stuart A. Scott[2,3], Sarah Kingan[1]

1. PacBio, 1305 O'Brien Drive, Menlo Park, CA 94025; **2.** Department of Pathology, Stanford University, Stanford, CA 94305; **3.** Clinical Genomics Laboratory, Stanford Health Care, Palo Alto, CA 94304; **4.** Stanford Center for Inherited Cardiovascular Disease, Stanford, CA, 94305

Introduction

Pharmacogenomic testing has the potential to improve patient treatment outcomes and reduce healthcare costs associated with medication efficacy and adverse drug reactions. Several assays and technologies exist for germline pharmacogenomic testing, but many clinically-significant pharmacogenomic loci remain challenging to accurately interrogate due to low sequence complexity and/or the presence of highly homologous pseudogenes. Long-read amplicon sequencing using Pacific Biosciences (PacBio) technology has previously been reported to accurately and precisely interrogate problematic pharmacogenomic loci, including HLA, CYP2D6, and SLC6A4[1,2]. However, pharmacogenomic HiFi sequencing panels have not been described. We developed a novel method to comprehensively interrogate a panel of 43 pharmacogenomic genes using an enrichment-based capture strategy (IDT) coupled with PacBio highly accurate (HiFi) Sequencing.

Panel Design

IDT lockdown probes were designed for a set of 43 genes to capture a total of 935 kb (Table 1). The full-length coding sequence is captured for 19 genes. Probes were 120bp in length (N=399) and designed with 0.2X tiling.

Pharmacogenomics Panel		
ABCB1	CYP4F2	MTHFR
ABCG2	DPYD	**NAT2**
ADRA2A	DRD2	**NUDT15**
APOL1	F2	OPRD1
BCHE	F5	OPRK1
COMT	G6PD	OPRM1
CTBP2P2	GRIK4	SLC6A4
CYP1A2	**HLA-A**	SLCO1B1
CYP2B6	**HLA-B**	**TPMT**
CYP2C19	**HLA-DQA1**	UGT1A1
CYP2C8	**HLA-DRB1**	UGT2B15
CYP2C9	**HTR2A**	**VKORC1**
CYP2D6	HTR2C	YEATS4
CYP3A4	**IFNL3**	
CYP3A5	**IFNL4**	

Table 1. Targets included in the pharmacogenomics panel. Probes were designed to cover 54 regions from 43 genes, with the goal of obtaining relevant SNP sites (plain text) or full-length CDS (bold text).

Sample Preparation, Capture, and Sequencing

We used HG002 and 24 Coriell samples[3] to evaluate the gene panel. Laboratory methods are described below:

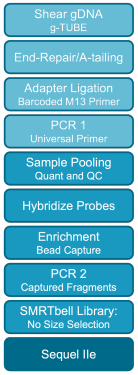

Figure 1. Workflow of sample preparation.

- 2µg of each gDNA sample was sheared to 10kb using the Covaris g-TUBE and ligated with barcoded adapters.
- Three equimolar pools of 8-plex barcoded gDNA library (2µg total) were input into the probe-based capture with a custom designed PGx gene panel (Coriell samples).
- A SMRTbell library was constructed using the captured and re-amplified gDNA (https://www.pacb.com/wp-content/uploads/Procedure-Checklist-%E2%80%93-Multiplex-Genomic-DNA-Target-Capture-Using-IDT-xGen-Lockdown-Probes.pdf).
- The resulting library was sequenced on one SMRT Cell 8M (30-hour movie) on the PacBio Sequel IIe System using v2.0 chemistry.

Data Analysis Workflow

SMRT Link was used to generate HiFi reads, remove PCR duplicates, and demultiplex, and a PacBio WGS pipeline was used to call variants for individual samples (Fig. 2).

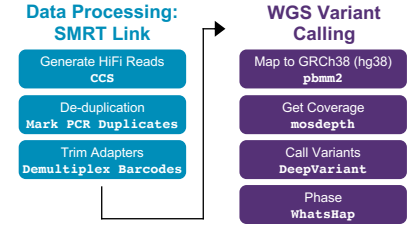

Figure 2. Example bioinformatics workflow for multiplexed sample captures. The PacBio human WGS workflow is publicly available on github: **PacificBiosciences/pb-human-wgs-workflow-snakemake**

Results

We benchmarked our variant calling in HG002 and found 100% concordance with SNP calls in the GIAB (ref) truth set (not shown). We also sequenced 24 Coriell samples on 1 SMRTCell 8M on the Sequel IIe system. On average, each sample had 300-fold coverage of ~7.2kb HiFi reads (Table 2). The average coverage across targets was 300-fold (Fig. 3), with 87% of targets exceeding 50x. Average coverage across samples (Fig. 4) was uniform. We benchmarked CYP2D6 star allele calls against Scott et al. (2020)[4] (not shown) and demonstrated the ability to call *5 allele (Fig. 5).

HiFi Reads	HiFi Yield	Median Read Length	Median Read Quality	Mean On-Target Coverage
2.2 M	16 Gb	7.2 kb	Q41	300-fold

Table 2. Yield statistics for 24-plex of Coriell Samples which were sequenced on one SMRT Cell 8M on the PacBio Sequel IIe System.

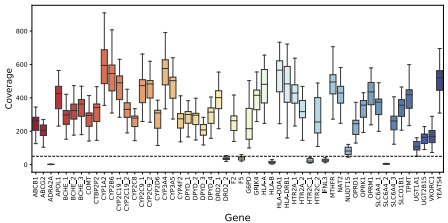

Figure 3. Distribution of coverages obtained for all target regions (54 regions, 43 genes) from the 24-plex dataset. The dotted line indicates 50-fold coverage; only 7 of 54 target regions fell below this mark.

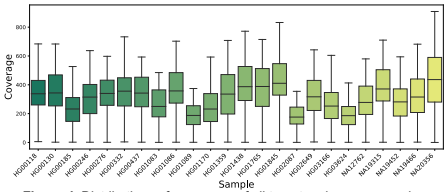

Figure 4. Distributions of coverages of all target regions per sample, for all samples included in the 24-plex capture.

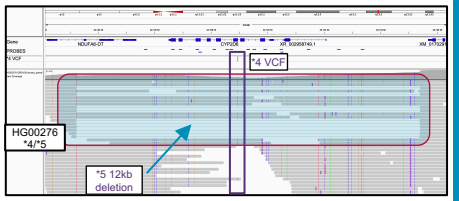

Figure 5. Capture of *5 deletion allele at CYP2D6.

Future Work

One round of probe re-design was performed for seven targets with low average coverage, resulting in increased reads on target (Fig. 5). Additional probe refinement may optimize coverage and increase multiplexing capacity.

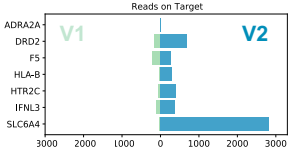

Figure 5. An improved probe design (V2) resulted in increased reads on target for 7 genes displaying low or uneven coverage with the initial probe design (V1).

Conclusion

We demonstrate a long-read capture method on a gene panel of 43 pharmacogenomic genes. This approach may be applied broadly to other gene panels and can be run in a cost-effective manner with high sample multiplexing.

References

1. Yang Y et al. (2017). Sequencing the *CYP2D6* gene: from variant allele discovery to clinical pharmacogenomic testing. *Pharmacogenomics*, 18, 673-685.

2. Botton MR et al. (2020). Phased haplotype resolution of the *SLC6A4* promoter using long-read single molecule real-time sequencing. *Genes*, 11, 1333.

3. Zook JM et al. (2016). Extensive sequencing of seven human genomes to characterize benchmark reference materials. *Scientific Data*, 3, 160025.

4. Scott SA et al. (2020). Development and analytical validation of a 29 gene clinical pharmacogenomic genotyping panel: multi-ethnic allele and copy number variant detection. *Clinical and Translational Science*, 14, 204-213.