P472

Highly scalable pharmacogenomic panel testing with hybrid capture and long-read sequencing

N. Gonzaludo[1], S. Kingan[1], J. Harting[1], P. Baybayan[1], T. Han[2], L. Arbiza[2], B. Li[3], Y. Yang[3,4], N. Hammond[4], S. A. Scott[3,4]

[1] PacBio, Menlo Park, CA, [2] Twist Bioscience, South San Francisco, CA, [3] Stanford University, Stanford, CA,
[4] Clinical Genomics Laboratory, Stanford Medicine, Palo Alto, CA

Introduction

In order to fully utilize pharmacogenomics (PGx) for precision medicine research, PGx assays must be:

- Comprehensive
- High-throughput and scalable
- Robustly cover difficult-to-sequence and difficult-to-map regions (present in many clinically actionable PGx loci)

Legacy low-cost technologies can be biased towards known variants and/or populations and may result in data that are incomplete or difficult to phase, potentially affecting phenotype prediction and subsequent clinical utility. With long-read PacBio HiFi sequencing and Twist Bioscience hybrid capture technology, we describe a pre-designed PGx panel that is comprehensive and cost-efficient, allowing for scalable application to precision medicine research programs.

Panel design

PGx research panel:

- 49 gene targets, including all 20 current genes with CPIC guidelines, as well as FDA PGx genes and genes of PGx research interest
- Probes optimized using proprietary algorithm to enable balanced capture of complex regions
- Available as a ready-made Twist Alliance Long-Read PGx Panel (www.twistbioscience.com/products/ngs/Long-Read-Sequencing-Panels)

CYP genes	HLA	Others	
CYP1A2*	**HLA-A**	**ABCB1**	HTR2C
CYP2B6⁺	**HLA-B**	**ABCG2**	**IFNL3**
CYP2C19	**HLA-DQA1**	ADD1	**MT-RNR1****
CYP2C8	**HLA-DRB1**	**ADRA2A**	**MTHFR**
CYP2C9		**ANKK1**	**NAGS**
CYP2D6		**APOL1**	**NAT2**
CYP3A4		**BCHE**	**NUDT15**
CYP3A5		CACNA1S	**OPRD1**
CYP4F2		CFTR	**OPRK1**
		COMT	**OPRM1**
		CTBP2P2	POLG
		DPYD	**RYR1**
		DRD2	**SLC6A4**
		F2	**SLCO1B1**
		F5	**TPMT**
		G6PD	**UGT1A1**
		GBA	**UGT2B15**
		GRIK4	**VKORC1**
			YEATS4

Table 1. 49 gene targets included in the Twist Alliance Long-Read PGx Panel. ***Bold** denotes full-gene coverage. ⁺Underline denotes inclusion in a CPIC Level A guideline. **full length mtDNA coverage is available for spike-in separately through Twist.

Sample preparation, capture, and sequencing

- HG002 and 23 Coriell PGx GeT-RM reference samples were sequenced on a PacBio Sequel IIe system
- 10 samples were selected for public data release and are described in this poster (www.pacb.com/connect/datasets/#targeted-datasets)
- Workflow:

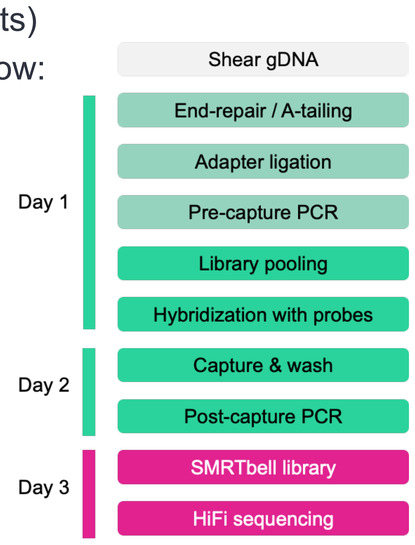

Figure 1. Workflow for sample preparation. PacBio steps in pink, Twist in green, and third party in grey. 24 samples were sequenced on a single SMRT Cell 8M.

Data analysis and results

SMRT Link v11.0 (HiFi reads, PCR duplicate marking and demultiplexing). Variants were called using a PacBio targeted sequencing pipeline, including DeepVariant, phasing with WhatsHap, and targeted metric calculation with Picard. Github: PacificBiosciences/HiFiTargetEnrichment.

Relevant PGx variants (CPIC database v1.22.2) and read coverage were assessed at each position, average read coverage for HLA variants (Fig. 2):

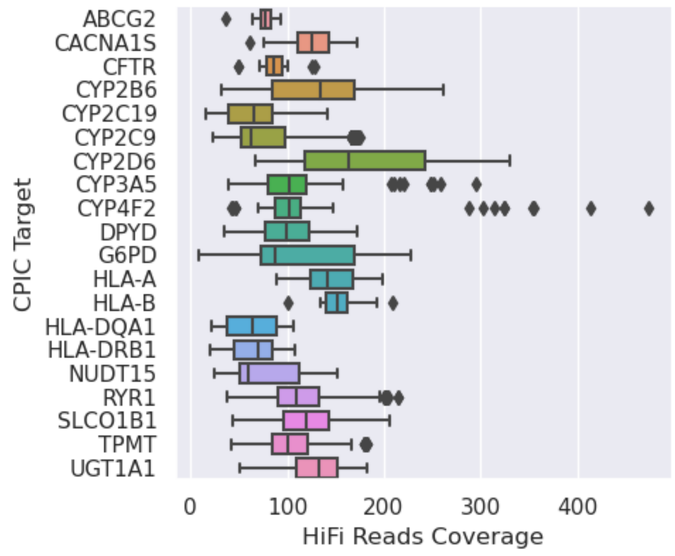

Figure 2. HiFi reads coverage of CPIC variants. Distribution of CPIC variant coverage across the 10 samples, summarized by gene.

Phased gVCFs produced as input into PharmCAT v2.1.2 for star (*) allele calling. Mapped bam files used to call CYP2D6 diplotypes using Pangu. Concordance for 9 samples with GeT-RM consensus calls[1] were compared (Fig. 3):

		HG00276	HG01190	NA07348	NA11832	NA18518	NA19109	NA19174	NA19207	NA19226
CYP2B6	GeT-RM	*2/(*4)	*1(*5)/*1(*27)	*1/*1	*1/*1	*1/*6	*1/*6	*6/*18	*1/*6	*18/(*20)
	PharmCat	*2/*4	*1/*5	*1/*1	*1/*1	*1/*6	*1/*6	*6/*18	*1/*6	*18/*20
CYP2C9	GeT-RM	*1/*2	*1/*2	*1/*1	*1/*3 (*18)	*1/*1	*1/*1	*1/*1	*1/*1	*1/*8
	PharmCat	*1/*2	*1/*61	*1/*1	*1/*3	*1/*1	*1/*1	*1/*1	*1/*1	*1/*8
CYP2C19	GeT-RM	*1/*1	*1/*2	*2/*17	*1/*2	*2/*17	*17/*17	*1/*2	*2/*17	*1/*2
	PharmCat	*1/*1	*1/*2	*2/*17	*2/*38	*2/*17	*17/*17	*1/*2	*2/*17	*1/*2
CYP2D6	GeT-RM	*4/*5	*4/*5	*1/*6	*1/*4	*17/*29	*2XN/*29	no consensus (*4/*40)	*2/*10/*XN	*2/*2XN
	PharmCat	*4/*5	*68+*4/*5	*1/*6	*1/*68+*4	*17/*29	*2x2/*29	*4/*40	*2x2/*10	*2x2/*2
CYP3A4	GeT-RM	*1/*2	*1/*1B	*1/*1	*1/*1	*1B/*1B	*1B/*1B (*15)	*1B/*1B	*1B/*1B	*1B/*1B (*15)
	PharmCat	*1/*2	*36/*36	*1/*1	*1/*1	*36/*36	*15/*36	*36/*36	*36/*36	*15/*36
CYP3A5	GeT-RM	*3/*3	*1/*1	*3/*3	*3/*3	*1/*6	*1/*3	*1/*6	*3/*7	*1/*6
	PharmCat	*3/*3	*1/*1	*3/*3	*3/*3	*1/*6	*1/*3	*1/*6	*3/*7	*1/*6
CYP4F2	GeT-RM	*1/*1	*1/*3	*1/*1	*1/*1	*1/*2	*1/*1	*1/*1	*1/*1	*1/(*2)
	PharmCat	*1/*1	*1/*3	*1/*1	*1/*1	*1/*2	*1/*1	*1/*1	*1/*1	*1/*2
SLCO1B1	GeT-RM	*1/*15	*1/*1	*1/*1	*1/*1	*1A/*1B	*1/*15	*1A/*1B	*1A/*1B	*1/*1
	PharmCat	*1/*15	*1/*1	*1/*1	*14/*20	*1/*37	*15/*37	*1/*27	*37/*41	*31/*37
TPMT	GeT-RM	*1/*1	*1/*1	*1/*1	*1/*1	*1/*1	*1/*1	*1/*1	*1/*1	*1/*1
	PharmCat	*1/*16	*1/*1	*1/*1	*1/*1	*1/*1	*1/*1	*1/*1	*1/*1	*1/*1
UGT1A1	GeT-RM	*60/*28	(*37)/*60	*60/*60	(*28 + *60)/*60	*60/*60	*1/*60	*60 / (*28 + *60)	(*36+*60)/(*28+ *60)	*1/*60
	PharmCat	*1/*80+*28	*1/*80+*37	*1/*1	*1/*80+*28	*1/*1	*1/*1	*1/*80+*28	*36/*80+*28	*1/*1
VKORC1 rs9923231	GeT-RM	GA	GG	GG	GG	GG	GG	GG	GG	GG
	PharmCat	GA	GG	GG	GG	GG	GG	GG	GG	GG

Table 3. Star (*) allele diplotype concordance or refinement for selected genes called by PharmCAT. GeT-RM rows indicate consensus calls from literature[1]. PharmCAT rows indicate output diplotypes from the HiFi targeted sequencing data. Discordant diplotypes in red. Green boxes: no change in phenotype, yellow boxes: PharmCAT call predicted a different phenotype compared to original GeT-RM call; Note: PharmVar has retired CYP3A4*36 due to lack of evidence for function; *36 and *1 can be considered functionally equivalent. PharmCAT does not detect the normal function UGT1A1*60 allele due to lack of guidance in the CPIC guideline; *60 and *1 are functionally equivalent.

Discussion

- Variants included in CPIC guidelines robustly covered by the panel (at least mean 50x across CPIC genes)
- Star allele refinement likely due to allele definition updates or larger number of variants for higher resolution star allele assignment
- Refined diplotypes did not result in predicted phenotype change in most cases, however, HiFi-based calls for *SLCO1B1* predict increased or decreased function in some samples, as well as an indeterminate rather than normal metabolizer call for *TPMT* in one sample.

Conclusions

- The described PGx panel for long-read HiFi sequencing provides unbiased coverage of medically-actionable pharmacogenes.
- Results show high accuracy for detecting key loci for pharmacogenomics research and implementation.
- Refinement using HiFi sequencing may impact drug response phenotype prediction. As HiFi data enables full-gene phasing across targets, future work will assess phasing impact on star allele calls, subsequent phenotype prediction, and overall clinical utility.

References and Resources

- PharmCAT: K Sangkuhl & M Whirl-Carrillo, et al. Pharmacogenomics Clinical Annotation Tool (PharmCAT). Clinical Pharmacology & Therapeutics (2020) 107(1):203-210.
- Pangu: Github: PacificBiosciences/pangu

1. Pratt VM, Everts RE, Aggarwal P, et al. Characterization of 137 Genomic DNA Reference Materials for 28 Pharmacogenetic Genes: A GeT-RM Collaborative Project. *J Mol Diagn.* 2016;18(1):109-123.