From sample to star alleles: a long-read pharmacogenomics pipeline powered by Twist target enrichment and PacBio HiFi sequencing

N. Gonzaludo[1], T. Han[2], L. Arbiza[2], A. Souppe[1], C. Lambert[1], S. Zhang[1], P. Baybayan[1], H. Ferraro[1], S. Kingan[1], B. Li[3], K. Sangkuhl[3], M. Woon[3], R. Whaley[3], M. Whirl-Carrillo[3], Y. Yang[3], T. E. Klein[3], S. A. Scott[3], J. Harting[1]

[1] PacBio, Menlo Park, CA, [2] Twist Bioscience, South San Francisco, CA, [3] Stanford University, Stanford, CA

Introduction

Routine implementation of pharmacogenomics (PGx) can enable an increase medication safety and efficacy. Long-read PacBio HiFi sequencing technology allows for:

- High accuracy calling of SNVs, indels, and SVs
- Ancestry-agnostic content
- Direct phasing (critical for unambiguous star (*) allele assignment)
- Cost-effective, high-throughput sequencing

However, data analysis remains challenging, with manual bioinformatics expertise often needed to stitch together tools and workflows. We describe a demonstrative workflow for analysis, interpretation, and reporting of long-read sequencing data for PGx to support scalable research and implementation.

Panel design

A pharmacogenomics research panel was designed to capture 50 gene targets, including all 20 current genes with CPIC guidelines, as well as FDA PGx genes and genes of PGx research interest. Probes were optimized using a proprietary algorithm to enable balanced capture of complex regions. The design is available as a ready-made panel called the Twist Alliance Long-Read PGx Panel at **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. Targets included in the Twist Alliance Long-Read PGx Panel. Fifty gene targets; ***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.

Pharmacogenomics sequencing pipeline

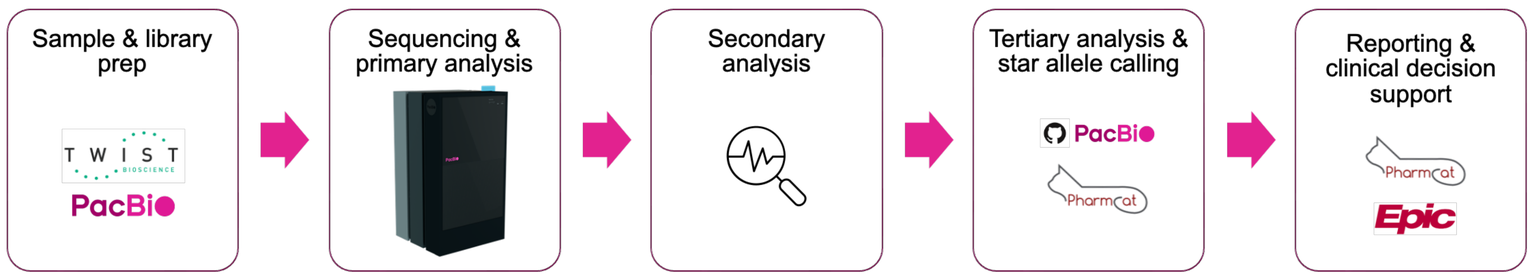

Figure 1. A demonstrative workflow for pharmacogenomics implementation. We demonstrate the use of the Twist Alliance Long-Read PGx panel for use with PacBio HiFi sequencing systems, with an analysis pipeline that includes primary analysis (on-instrument), bioinformatics tools for secondary and tertiary analysis, star allele calling with Pangu and PharmCAT, and interpretation and reporting with PharmCAT and custom EHR clinical decision support (CDS) (e.g. Epic Best Practice Alerts).

Sample and library preparation, HiFi sequencing

HG002 and 23 Coriell PGx GeT-RM reference samples were used to evaluate the panel on a PacBio Sequel IIe system. Ten samples were selected for public data release as described here. Laboratory methods are described below (**Fig. 2**).

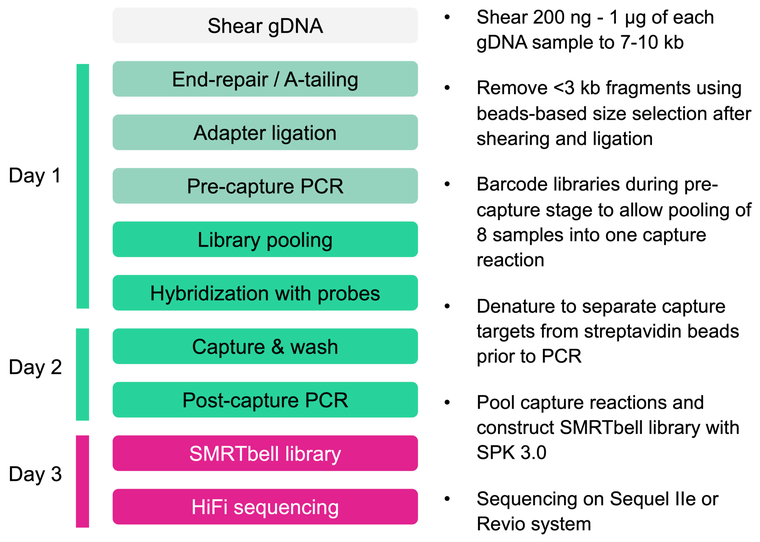

Figure 2. Workflow for sample preparation. Protocol and reagents from PacBio used for steps in pink, Twist in green, and third party in grey. 24 samples were sequenced on a single SMRT Cell 8M.

HiFi reads	HiFi yield	Med. on-target read length	Med. read quality	Mean on-target coverage
1.38 M	7.3 Gb	5,216 bp	QV 32	174x

Table 2. HiFi sequencing metrics for the 10 publicly available sample. CPIC variants were covered robustly in the panel (at least mean 50X).

PharmCAT

The Pharmacogenomics Clinical Annotation Tool (PharmCAT, **https://pharmcat.org**) is a bioinformatics tool that provides individual PGx annotations for genetic data. PharmCAT extracts PGx-relevant variants from a VCF file derived from sequencing or genotyping technologies, predicts PGx gene haplotypes and diplotypes, infers corresponding phenotypes, and presents phenotype-based clinical PGx recommendations from CPIC and PharmGKB-curated DPWG guidelines.

Secondary, tertiary analysis, and star allele calling

SMRT Link v11.0 was used to generate HiFi reads, mark PCR duplicates, and demultiplex. Variants were called using a PacBio targeted sequencing pipeline, including DeepVariant, phasing with WhatsHap, and targeted metric calculation with Picard. More information is available on **Github: PacificBiosciences/HiFiTargetEnrichment**.

Phased gVCFs were produced as input into PharmCAT v2.1.2 for star (*) allele calling, and mapped bam files were used to call *CYP2D6* diplotypes using Pangu (**Fig. 3**). Star allele diplotypes for the 10 samples were concordant with GeT-RM consensus calls for most genes. For *SLCO1B1, TPMT* and *UGT1A1,* allele calls were refined due to increased variant-level resolution, which led to changes in predicted allele function for some samples.

The Phenotyper and Reporter components of PharmCAT can be used to generate an output report based on CPIC guidelines, with content in development that is intended to support Epic genomic indicators and CDS.

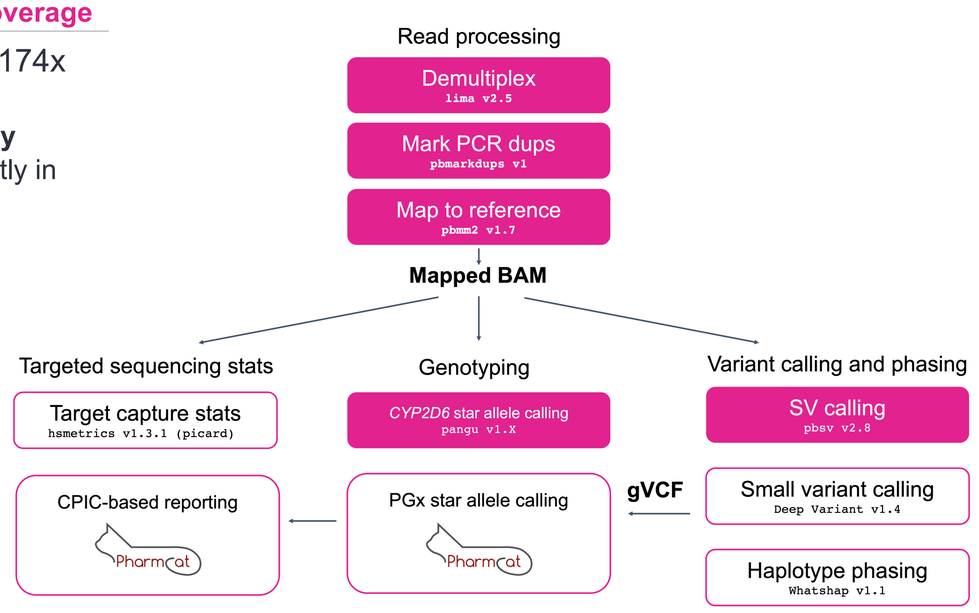

Figure 3. Bioinformatics workflow. PacBio-developed tools in pink boxes, 3rd party tools in white boxes.

Conclusion

This example workflow demonstrates how the use of long-read HiFi sequencing can enable the development of scalable pharmacogenomics and its implementation, streamlining the translation of sequencing data into phenotype information for precision medicine programs.

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.

- 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
- Data from this poster are publicly available at **www.pacb.com/connect/datasets/#targeted-datasets**