Enablement of long-read targeted sequencing panels using Twist hybrid capture and PacBio HiFi sequencing

S. Kingan[1], J. Harting[1], T. Hon[1], Y-C. Tsai[1], I. McLaughlin[1], J. Ziegle[1], T. Han[2], L. Arbiza[2], S. Kloet[3], L. Busscher[3], G. Henno*[1], E. Lee[1], N. Gonzaludo[1]

1. Pacific Biosciences, Menlo Park, California, USA; 2. Twist Bioscience, South San Francisco, California, USA;
3. Leiden University Medical Center, Leiden, Netherlands

P16.026.B

Introduction

While many assays and technologies exist for germline genomic testing, several clinically-relevant loci remain challenging to characterize due to low sequence complexity and/or the presence of highly homologous pseudogenes. Long and accurate PacBio HiFi sequencing enables:

- Consolidation of multiple assays on one platform with comprehensive detection of genetic variation, including SNPs, indels and structural variants
- Unambiguous haplotype resolution through direct phasing, without the need for imputation
- Ancestry-agnostic capture of novel and rare variants

Targeted sequencing allows for high-resolution characterization of gene panels at a scale and cost that is more accessible than whole genome sequencing. We describe a method to leverage Twist Bioscience's double-stranded DNA probes that can be individually tuned to enrich target regions with exceptional uniformity.

We describe results from a custom panel designed to fully capture 20+ pharmacogenes related to drug response.

PacBio HiFi reads

HiFi reads are both long (up to 25 kb) and accurate (99.95% accuracy). HiFi reads are generated with SMRT sequencing (Fig. 1). Briefly, a linear template sequence is ligated to SMRTbell adapters. DNA polymerase synthesizes complementary sequences to both strands of the original linear template, leading to rolling circle sequencing and multiple passes of the original template. CCS uses the individual subreads to generate a highly accurate consensus sequence (HiFi read).

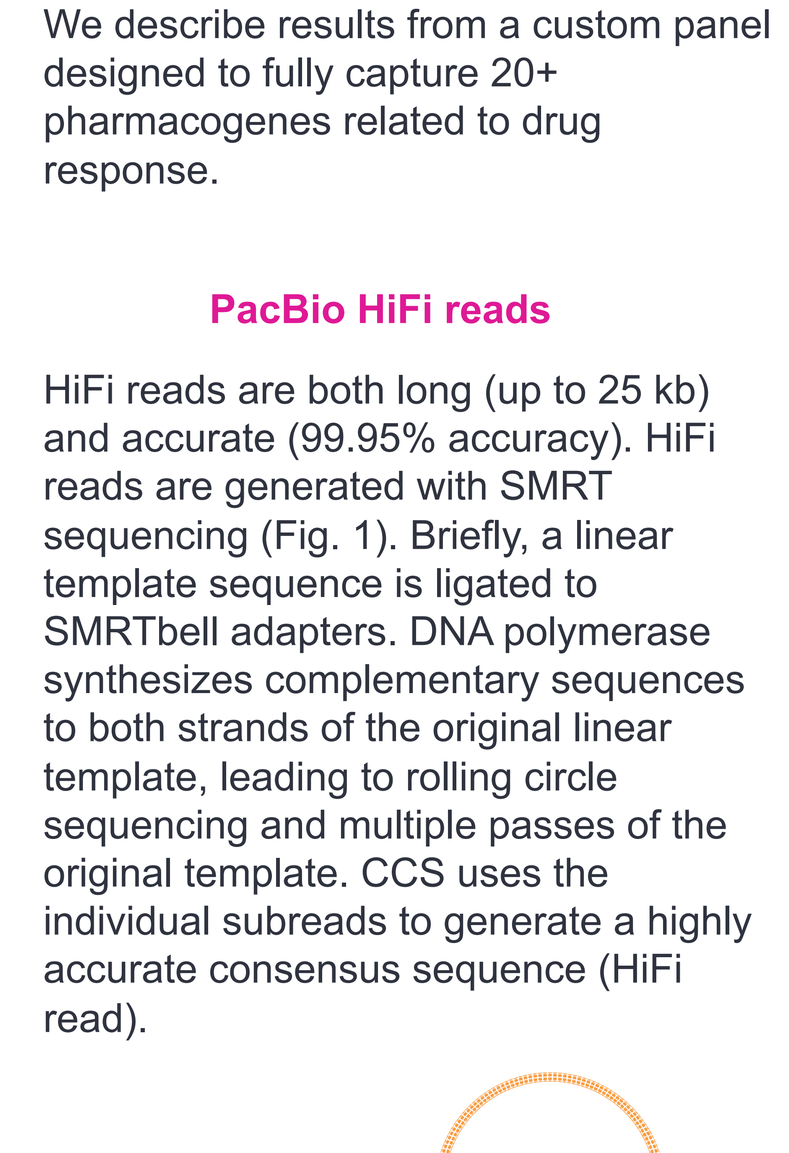

Figure 1. Schematic of PacBio SMRT sequencing and generation of HiFi reads. More information is available at **http://ccs.how**

Panel design

A pharmacogenomics research panel was developed through the Twist Bioscience custom panel design process. Probes were optimized using a proprietary algorithm to enable balanced capture of complex regions. Probes were designed to cover a 2 Mb target region of interest with sparse tiling density.

CYP genes	HLA class I	HLA class II	Others
CYP2B6			ASL
CYP2C19		HLA-DPB1	CFTR
CYP2C9	HLA-A	HLA-DQA1	DPYD
CYP2D6	HLA-B	HLA-DQB1	F5
CYP3A4	HLA-C	HLA-DRA	SLCO1B1
CYP3A5		HLA-DRB1	TPMT
CYP3A7			UGT1A1
			VKORC1

Table 1. Targets included in the pharmacogenomics panel.

Sample preparation, capture, and sequencing

We used HG002 and 23 Coriell GeT-RM samples[1] to evaluate the gene panel on a single SMRT Cell 8M. Laboratory methods are described below (Fig. 2):

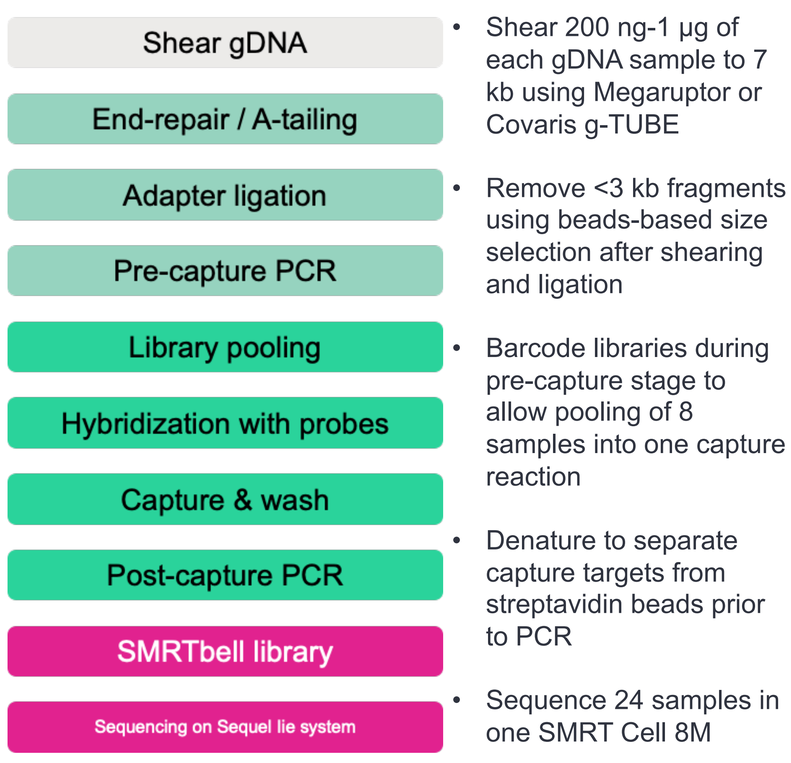

Figure 2. Workflow for sample preparation. Protocol and reagents from PacBio[2] used for steps in pink, Twist in green, and third party in grey.

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. 3).

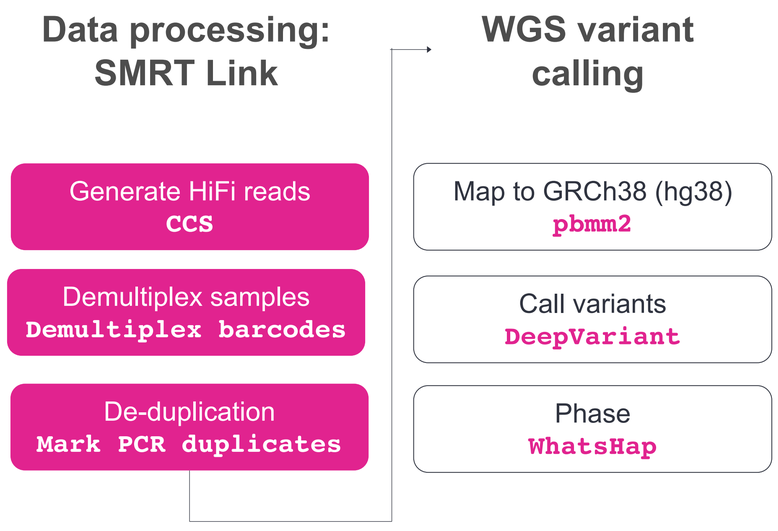

Figure 3. Example bioinformatics workflow for multiplexed sample captures. The PacBio human WGS workflow is publicly available on github: **PacificBiosciences/pb-human-wgs-workflow-snakemake**

Results

24 GeT-RM Coriell samples were sequenced on 1 SMRT Cell 8M on the Sequel IIe system (Table 2). Samples had on average 117 k HiFi reads, with a mean on-target read length of ~6.5 kb (Fig. 3). The percent of targeted regions covered was fairly uniform except for two samples, NA18518 and NA18868 (Fig. 4). Across all samples, 94% of target regions exceeded 30x coverage.

HiFi reads	HiFi yield	Median on-target read length	Median read quality	Mean on-target coverage
2.88 M	18.6 Gb	6,513 bp	QV 41.7	190x

Table 2. HiFi sequencing metrics.

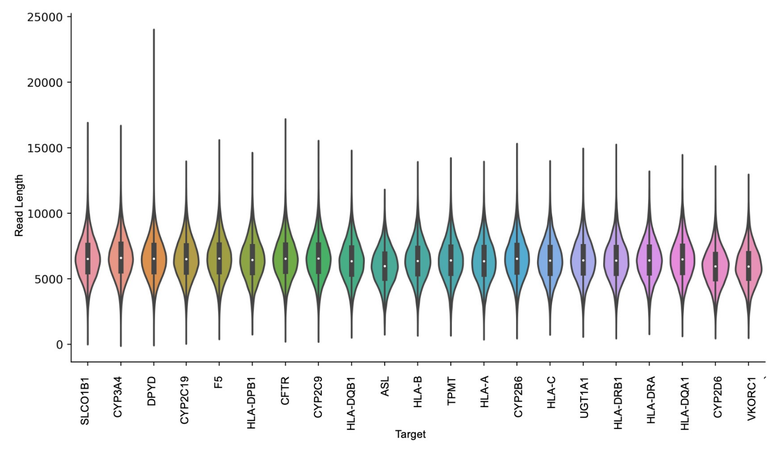

Figure 3. Read length distribution across all pharmacogene targets.

Star alleles were called using Aldy v4[3], which has been adapted to run on long reads. Star allele diplotypes were 100% concordant with GeT-RM and WGS consensus calls (data not shown) for all samples in the following genes: CFTR, CYP2B6, CYP2C19, CYP2C9, CYP3A4, DPYD, SLCO1B1, TPMT, UGT1A1, and VKORC1.

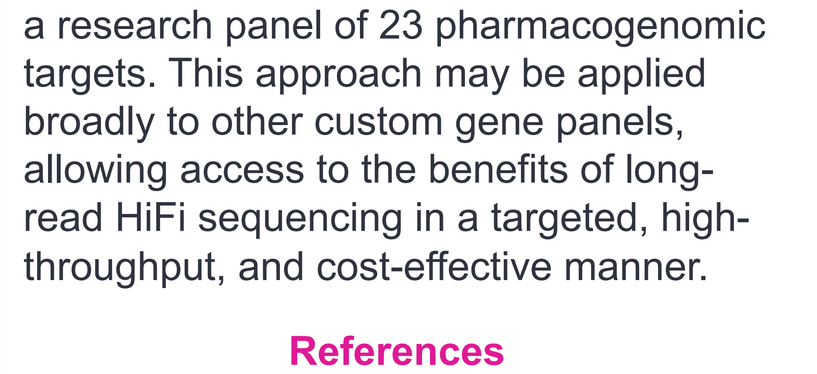

Figure 5. DPYD haplotypes of Yoruban sample NA19239, fully phased showing *1/*9 star allele diplotype, concordant with the GeT-RM consensus call. Over 50 variant positions are phased across the 50 kb region of DPYD shown here.

Conclusion

We demonstrate a long-read capture method using Twist Bioscience enrichment probes to accurately and efficiently capture a research panel of 23 pharmacogenomic targets. This approach may be applied broadly to other custom gene panels, allowing access to the benefits of long-read HiFi sequencing in a targeted, high-throughput, and cost-effective manner.

References

1. Pratt VM, et al. (2016). Characterization of 137 genomic DNA reference materials for 28 pharmacogenetic genes. *J Mol Diag,* Jan 18(1): 109-23.
2. Preparing multiplexed amplicon libraries using SMRTbell prep kit 3.0. https://www.pacb.com/wp-content/uploads/Procedure-checklist-Preparing-multiplexed-amplicon-libraries-using-SMRTbell-prep-kit-3.0.pdf
3. Numanagić I, et al. (2018). Allelic decomposition and exact genotyping of highly polymorphic and structurally variant genes. *Nat Commun* 9, 828. Version 4 (beta) available on github: https://github.com/0xTCG/aldy/tree/aldy-4