Long-read PGx v2 Panel with Twist target enrichment system

Tina Han[1], Owen Hardy[1], Holly Corbitt[1], Rebecca Barnard[1], James Flynn[1], Elian Lee[1], Andreas Scherer[2], Gabe Rudy[2], Nathan Fortier[2], and Solomon Reinman[2]

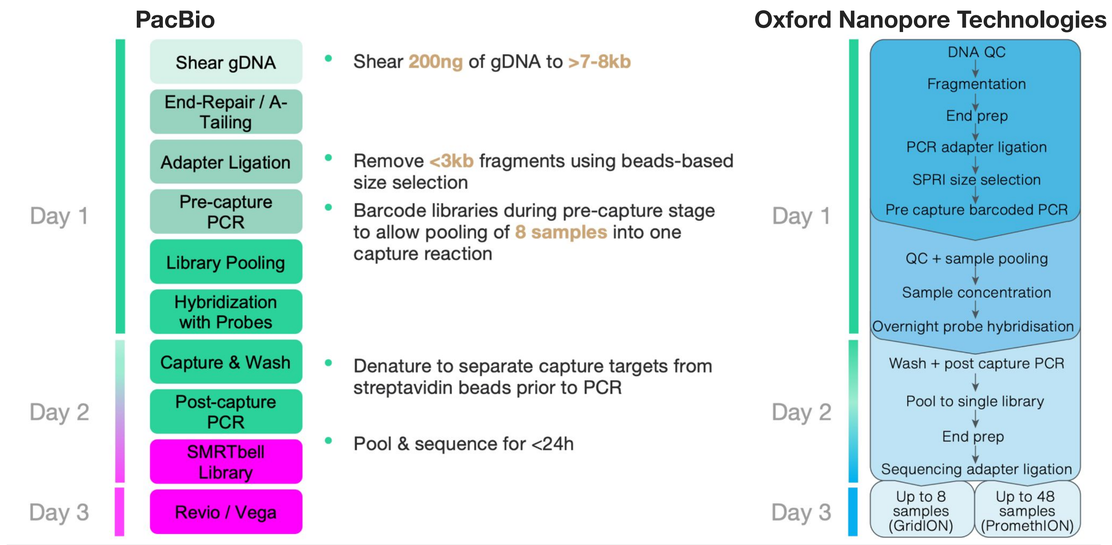
[1]Twist Bioscience, South San Francisco, California, USA; [2]Golden Helix, Bozeman, Montana, USA

1. Introduction

Many pharmacogenomics (PGx) loci, such as CYP and HLA genes, are difficult to sequence and map with short reads. While long-read sequencing provides comprehensive analysis, existing methods focus on a single SNP or gene and often produce ambiguous results. We present a scalable, cost-effective long-read workflow that simultaneously captures 30 PGx genes in one assay. The protocol and the panel are compatible with sequencers from both Oxford Nanopore Technologies (ONT) and PacBio.

2. Methods

Library preparation and sequencing. Eight Coriell GeT-RM reference samples were captured with the Twist Alliance Long-Read PGx v2 Panel: NA12878, NA18509, NA18519, NA18868, NA19109, NA19147, NA19207, and NA19226. The protocol uses 200–1000 ng of mechanically fragmented gDNA, ligated to truncated Y-adapters, and barcoded using 10-bp UDIs with either KOD Xtreme Hot Start DNA Polymerase (Merck/Millipore/Sigma) or Twist TrueAmp Polymerase. KOD libraries were prepared in duplicate (16 libraries per platform); TrueAmp libraries as singletons (8 per platform). Following overnight multiplex hybridization, post-capture libraries undergo ONT LSK114 or PacBio SMRTbell 3.0 preparation. Multiplexing capacity ranges from 40 samples on ONT GridION to 2,304 on PromethION P48, and 72 on PacBio Vega to 384 on Revio per run. GRCh38 was used as reference genome for alignment.



ONT secondary analysis. R10.4.1 flow cells; Dorado HAC v5.2.0 basecalling; minimap2 2.28 (lr:hq preset); Clair3 2.0.3 (r1041_e82_400bps_hac_v520 model) small variant calling; longphase haplotype phasing.
PacBio secondary analysis. Revio HiFi; pbmm2 1.17 (HIFI preset); DeepVariant 1.9.0 (PacBio model); HiPhase 1.7 phasing. CYP2D6, HLA-A and HLA-B diplotypes from pb-StarPhase 2.2.0 (targeted mode, database 2026-07-24).
ONT CYP2D6 caller - Chinook. ONT ran wf-pgx workflow (minimap2 2.24 alignment, Clair3 1.1.2 small variant calling, PharmCAT 2.15.5, Chinook 0.7.2) on the FASTQ files with the dna_r10.4.1_e8.2_400bps_hac@v5.0.0 basecaller configuration, Twist enrichment mode, the panel target BED file and the GRCh38 no-alt analysis set.
ONT CYP2D6 copy number (comparison only). VarSeq Coverage Statistics over the 136 panel targets, followed by the VarSeq CNV Caller on Target Regions against a reference pool of 12 two-copy ONT libraries from this cohort, with ratio thresholds calibrated to capture data (one copy 0.72, two copies 1.0, three copies 1.40). The duplicated allele was assigned from the alternate allele fraction at the allele-defining SNP in the phased VCF.
Tertiary analysis and reporting. Phased VCFs were imported into VarSeq (Golden Helix) using the PGx with CNVs template, with force-calling from BAM coverage at CPIC must-call positions. PGx report with star allele diplotypes, CPIC phenotypes and CPIC and FDA drug recommendations were assigned from Golden Helix's curated PGx annotation library (CPIC/FDA sources dated 2026-08-03 and 2026-04-17).
Concordance scoring. Calls were compared with the GeT-RM consolidated table (1-22-25-V4) for 12 genes with reference values. CYP1A2 and CYP3A4 were excluded because they carry no CPIC recommendations. HLA-A and HLA-B were excluded because GeT-RM has no values for these samples. Nomenclature identities (SLCO1B1 *1A/*1B to *1/*37; DPYD *9A to *9; UGT1A1 *60 dropped per CPIC 2017 reclassification) were applied before comparison. Every remaining disagreement was adjudicated against phased genotypes at each allele-defining position on both platforms.

3. Protocol Optimization

Figure 1. The optimized protocol (LP v2) and the updated panel design (Panel v2) offer easier operation, and the increased pre- and post-capture yields, respectively, to ensure adequate mass for downstream ONT or PacBio library preparation and sequencing.

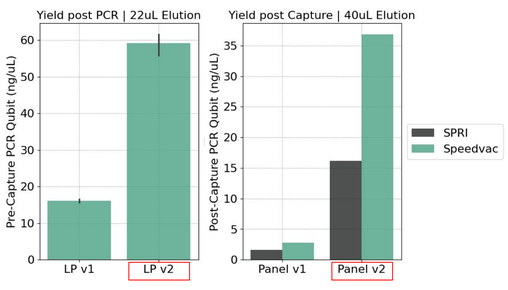
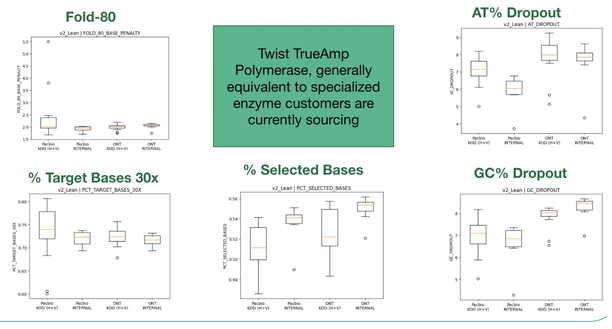


Figure 2. Twist TrueAmp Polymerase serves as a drop-in replacement for KOD Xtreme Hot Start DNA Polymerase without compromising capture performance.



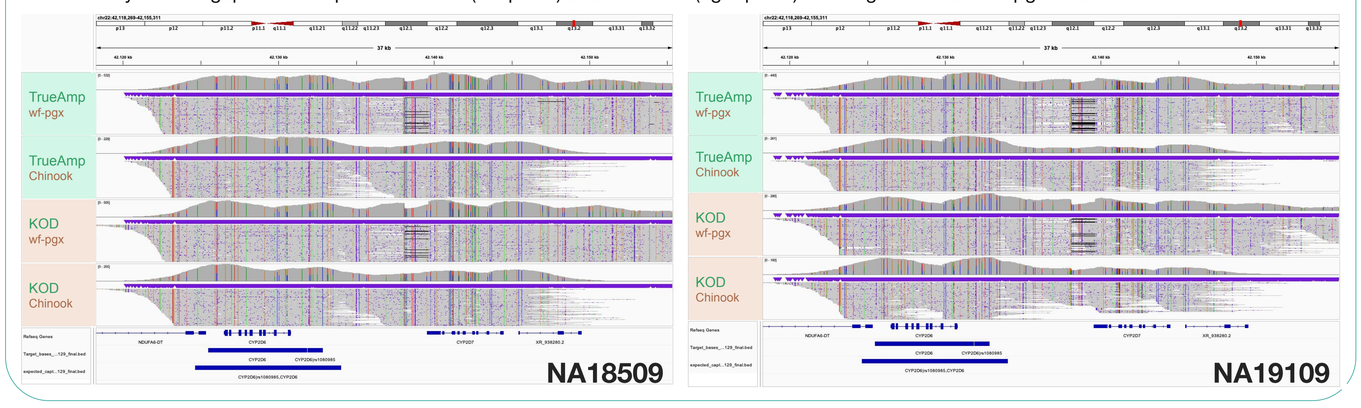
4. Updated Panel Content

The Twist Alliance Long-Read PGx v2 Panel has a 1.8 Mb footprint, covering 30 PGx genes in CPIC and DPWG with 26 sample tracking SNPs. Besides the nuclear content, variants in mitochondrial DNA are covered to enable simultaneous detection of heteroplasmy. This updated panel incorporates denser probe tiling strategy, probes to each exon, better selection of probes over longer repeats, addressing more coverage fill to the v1 Panel. *[*New in v2]*

CYP genes	HLA	Mitochondrial	Others	Sample Tracking*	Removed from the v1 Panel
CYP1A2 CYP2B6 CYP2C19 CYP2C9 CYP2D6 CYP3A4 CYP3A5 CYP4F2	HLA-A HLA-B	Entire mitochondrial genome can be targeted to cover 4 SNPs in MT-RNR1	ABCG2 CACNA1S CFTR COMT DPYD F2 F5 G6PD HTR2A* IFNL3 MTHFR NAT2 NUDT15 OPRM1 RYR1 SLC6A4 SLCO1B1 TPMT UGT1A1 VKORC1	rs2229546 rs1410592 rs10883099 rs4617548 rs7300444 rs9532292 rs2297995 rs4577050 rs2070203 rs1037256 rs9962023 rs2228611 rs497692 rs10203363 rs10373 rs4148973 rs4675 rs2819561 rs4688963 rs309557 rs2942 rs17548783 rs4735258 rs1381532 AMELX AMELY	ABCB1 ADD1 ADRA2A ANKK1 APOL1 BCHE CYP2C8 CTBP2P2 DRD2 GBA GRIK4 HLA-DQA1 HLA-DRB1 HTR2C NAGS OPRD1 OPRK1 POLG UGT2B15 YEATS4

5. Panel Performance and Coverage in IGV

Coverage was computed in VarSeq (Coverage Statistics algorithm) on the aligned BAM files from the secondary analysis, using the Twist expected capture target BED file (136 targets, 1,837,456 bp, GRCh38). Per-target mean depth and the fraction of target bases at or above 1x, 20x, 100x and 500x were exported for every library and combined with a Python script. Panel-wide values are base-weighted across all 136 targets, so a percentage at 20x is the fraction of all targeted bases, not the mean of per-target percentages. Read counts are primary mapped reads from samtools flagstat on the same BAM files. Duplicate rate is not reported because duplicates are not marked in long-read alignments, and read length was not measured. On the non-downsampled datasets, ONT libraries have a mean target depth of 650 to 700x with 98 percent of target bases at or above 20x. PacBio KOD libraries reached 237x with 98 percent of target bases at or above 20x, and PacBio TrueAmp libraries reached 66x with 90 percent at or above 20x. Coverage of CYP2D6 with TrueAmp (top track) and KOD (bottom track) showed great uniformity with no gaps in 2 samples: NA18509 (left panel) and NA19109 (right panel) with alignment from wf-pgx or Chinook.



6. Concordance of Star Alleles Calling

Across 12 genes with GeT-RM reference calls (CYP2D6, CYP2C19, CYP2C9, CYP2B6, CYP3A5, CYP4F2, SLCO1B1, TPMT, DPYD, NAT2, UGT1A1, VKORC1), star allele concordance after nomenclature reconciliation was 96-97% for ONT and 93-94% for PacBio. *[Note: Reconciled: matches GeT-RM after current PharmVar/CPIC nomenclature is applied and alleles GeT-RM did not interrogate are credited; calls contradicted by phased reads are never counted. Exact-match concordance before reconciliation: 69–70% ONT, 67% PacBio.]*

CYP2C9, CYP2C19, CYP3A5, TPMT and VKORC1 were fully concordant on every sample, polymerase and sequencing platform. Most remaining differences were attributable to GeT-RM's use of legacy allele nomenclature or to alleles GeT-RM did not interrogate, which long-read phasing resolved. In three cases, long reads contradicted a low-confidence GeT-RM assignment on both platforms. Here is a shortlist of discordant calls:

Gene	Match	Reconciled	GeT-RM not supported by reads	Mismatch	Note
CYP2B6	6	1	1	0	*17 detected in NA18519; GeT-RM (*20) in NA19226 contradicted by phase
CYP4F2	4	3	1	0	Two phase-ambiguous samples; *7 in NA19207; GeT-RM (*2) in NA18868 not present
NAT2	4	4	0	0	All differences are legacy vs PharmVar nomenclature, two resolved by phasing
DPYD	2	6 (ONT) / 5 (PacBio)	0	0 (ONT) / 1 (PacBio)	Normal-function alleles not interrogated by GeT-RM (*5, c.1218G>A, c.1349C>T)
UGT1A1	2	6	0	0	*80 found in cis with every *28; GeT-RM reports legacy *60
SLCO1B1	2	5	1	0	PharmVar sub-alleles on the *1B backbone (*20, *30, *31, *41, *42, *44)

CYP2D6 is the only gene where the two sequencing platforms and the three callers differ in substance. Three approaches were compared: read-chain inference (pb-StarPhase), depth-based copy number (VarSeq CNV caller) and local assembly (Chinook, the CYP2D6 component of ONT's wf-pgx workflow, run by ONT on the same FASTQ files). pb-StarPhase infers copy number from read-level allele chains. The panel's HiFi inserts (approximately 3.3kb) do not span the full 4.4kb gene plus downstream repeat unit, so tandem duplications are not resolved, consistent with the pb-StarPhase documentation caveat about targeted data. The deletion and the *68+*4 hybrid are detected because they produce distinctive read-level signatures. The VarSeq CNV caller takes a depth-based approach against a two-copy reference pool and recovers all three duplications, but capture depth compresses the single-copy ratio (observed 0.77 against 0.5 expected) so the *5 deletion is not called, and hybrids have no depth signature. Chinook assembles each CYP2D6 and CYP2D7 allele from the reads and realigns them to the hg38 REP region for copy number and deletion detection, so it does not depend on either insert span or reference-pool depth. On this panel it resolved all three structural allele classes on every TrueAmp library.

NA18519 is called *29/*106 by all three callers on both platforms. CYP2D6 *106 (c.1252G>A) postdates GeT-RM's characterisation of this sample and has uncertain function in CPIC, so the predicted phenotype moves from Normal Metabolizer to Indeterminate.

Sample	GeT-RM	PacBio (pb-StarPhase)	ONT (VarSeq + CNV caller)	ONT (Chinook)
NA12878	*3/(*68)+*4	*3/*68+*4	*3/*4 (hybrid not called)	*3/*68+*4
NA18509	*2/*17	*2/*17	*2/*17	*2/*17
NA18519	*1/*29	*29/*106	*29/*106	*29/*106
NA18868	*2/*5	*2/*5	*2/*2 (deletion not called)	*2/*5
NA19109	*2x2/*29	*2/*29 (duplication missed)	*2x2/*29	*2x2/*29
NA19147	*17/*29	*17/*29	*17/*29	*17/*29
NA19207	*2x2/*10	*2/*10 (duplication missed)	*2x2/*10	*2x2/*10
NA19226	*2/*2x2	*2/*2 (duplication missed)	*2/*2x2	*2/*2x2

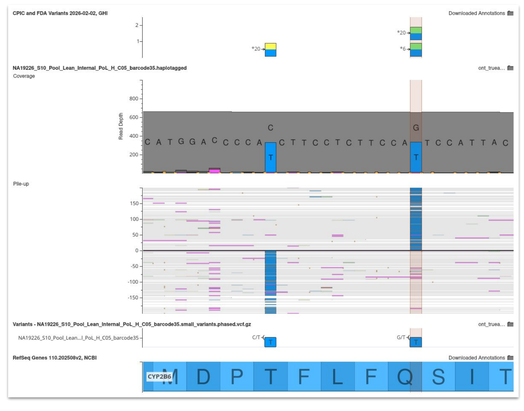
[Note: Chinook CYP2D6 diplotypes were generated by Oxford Nanopore Technologies (pre-release wf-pgx, Chinook 0.7.2).]

Why long-read matters

1. GeT-RM assignments contradicted by reads on both platforms (3 cases)
   - NA19226 SLCO1B1. GeT-RM reports *1/*1. Both platforms show c.388A>G homozygous and c.1463G>C heterozygous, giving *31/*37. *31 is a decreased-function allele, so the sample's predicted statin phenotype changes from Normal to Decreased Function.
   - NA19226 CYP2B6. A *20 call requires c.503C>T in cis with the *6-defining variants c.516G>T and c.785A>G (NM_000767.5), but read-level phasing places c.503C>T in trans to both and in cis with c.983T>C (*18). c.503 and c.516 are 13 bp apart, so the phase is read-level certain. The diplotype implied by the phasing information is *6/*18.
   - NA18868 CYP4F2. GeT-RM reports *1/(*2). The *2 variant c.34T>G is homozygous reference at high depth on both platforms. The sample instead carries c.1555C>A (*5).
2. Phase ambiguity resolved (2 cases, NAT2)
   - NA18868 NAT2. GeT-RM inferred *5/(*12). Long-read phasing places c.803G on the *5 haplotype, giving *4/*5 under PharmVar nomenclature.
   - NA19207 NAT2. GeT-RM inferred (*4)/*14. Phasing shows the second haplotype is genome reference, giving *1/*14.
3. Alleles present but not interrogated by GeT-RM
   - SLCO1B1 sub-alleles on the *1B backbone in 5 of 8 samples (*20, *30, *41, *42, *44). NA19207 carries *41, which CPIC assigns decreased function.
   - UGT1A1 *80 (rs887829) in cis with every *28 allele observed, on every library and both platforms.
   - CYP2B6 *17 in NA18519 and CYP4F2 *7 in NA19207.
   - DPYD normal-function alleles *5, c.1218G>A and c.1349C>T in 5 samples.

Phenotype impact. After reconciliation, five SLCO1B1 samples move from Normal Function to Indeterminate or Decreased Function, and NA18519 CYP2D6 moves to Indeterminate. All other reconciled calls keep the GeT-RM phenotype.

Figure 3. Long-read phasing rules out CYP2B6 *20 in NA19226. VarSeq GenomeBrowse view of the ONT haplotagged BAM and phased VCF at CYP2B6 c.503C>T and c.516G>T, 13 bp apart. The two variants never occur on the same read: c.503C>T lies in trans to the *6-defining c.516G>T and phases instead with c.983T>C (*18). Because *20 requires c.503C>T in cis with the *6 variants, the diplotype is *6/*18, not *18/*20 as reported by GeT-RM. PacBio reads show the same phase.



7. Summary

The optimized long-read protocol and the updated panel design provide higher yield and are compatible with TrueAmp Polymerase. The workflow also works with custom or fixed Twist panels and supports optional spike-ins, i.e. Twist Mitochondrial DNA Panel for MT-RNR1. Verified with GeT-RM PGx samples, the updated panel content efficiently covers targeted loci. Secondary/tertiary analysis outputs are concordant with GeT-RM. This workflow enables scalable, low-bias hybrid capture with long reads, accurately detecting diverse variant types, including structural variants and phased haplotypes difficult to resolve with short-read, qPCR or array methods.

8. Acknowledgements

The authors would like to thank Twist Application Labs and R&D teams for generating the data. The authors would also like to thank the technical assistance of Nhung Pham, Mehdi Motallebipour, and Jawahar Swaminathan at Oxford Nanopore Technologies.

Twist Long-Read Protocol

Disclaimers: For Research Use Only. Not for use in diagnostic procedures.