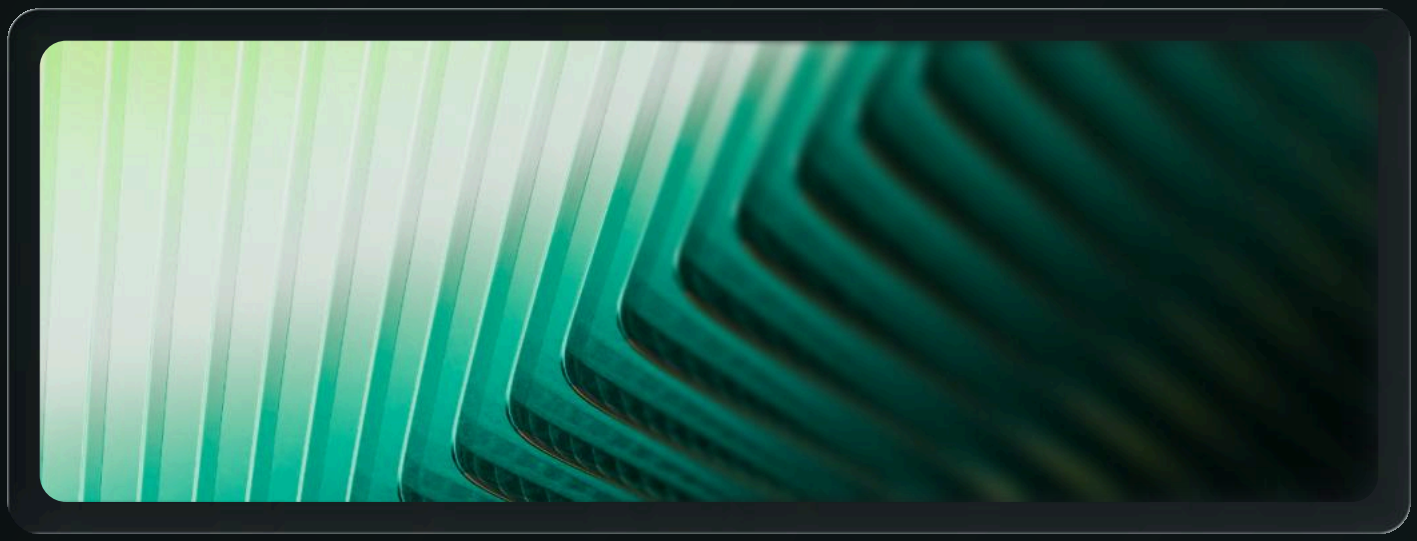
RENEW

Case Study



Long-Read Pharmacogenomics With ONT x Twist (49 Genes)

Long-Read Insights

Introduction

Decision-Grade Outputs That Scale

Full-gene PGx from a simple blood draw provides native phase, repeats, and CNVs, with no imputation and no short-read blind spots. A single long-read panel unifies UGT1A1, DPYD, CYP2D6 (CNVs/hybrids), and 46 additional genes into decision-grade outputs that scale from RUO discovery to GCP-fit exploratory covariates. (RUO; fit-for-purpose and clinical use cases available with clear validation pathways.)

Snapshot

Study Goals

Characterize ONT × Twist targeted long-read PGx (49 genes) for coverage, accuracy, and haplotype/CNV resolution to support RUO and GCP-level use in biopharma clinical trials and retrospective studies.

Approach

Sixteen cell lines (5 GIAB truth-set samples and 11 Coriell) were each split into 3 technical replicates. DNA was sheared to >10 kb, libraries were prepared with ONT × Twist capture, and standard long-read variant calling and haplotyping were performed.

Outcomes

>Q25 mean base quality; N50 >3-4 kb; mean on-target coverage >100× across targets. F1 >0.95 vs GIAB (per replicate), AUPRC >0.92; PharmCat and Chinook calls aligned with reference haplotypes.

ABCB1	ABCG2	ADD1	ADRA2A	ANKK1	APOL1	BCHE	CACNA1S
CFTR	COMT	CTBP2	CYP1A2	CYP2B6	CYP2C8	CYP2C9	CYP2C19
CYP2D6	CYP3A4	CYP3A5	CYP4F2	DPYD	DRD2	F2	F5
GBA	GRIK4	G6PD	HLA-A	HLA-B	HLA-DRB1	HLA-DQA1	HTR2C
IFNL3	MTHFR	NAGS	NAT2	NUDT15	OPRD1	OPRK1	OPRM1
POLG	RYR1	SLC6A4	SLCO1B1	TPMT	UGT1A1	UGT2B15	VKORC1
YEATS4	MT-RNR1						

Full-length gene targeted

Hot spots targeted

Oxford Nanopore workflow with Twist Bioscience covers mitochondrial DNA (mtDNA) and 49 PGx genes, comprising 39 entire gene sequencing and 10 genes covering key PGx variants.

Solving PGx's Hard Problems: Phase, Repeats, Structure

Short-read/SNP panels struggle where PGx gets hard: phase inference, repeats (e.g., UGT1A1*28), rare/ancestry-specific DPYD haplotypes, and CYP2D6 hybrids/CNVs. Long reads measure full genes end-to-end, delivering native diplotypes and structural context in one assay, so exploratory safety/PK models aren't limited by missing or mis-phased variants. For sponsors running FOLFOX/FOLFIRI/capecitabine or SN-38 ADC backbones specifically, this means cleaner exposure–toxicity covariates and fewer unexplained holds. (RUO; not a clinical diagnostic.)

Study Design & Methods

Samples & Replicates

- n = 48, 16 cell lines total
- 5 GIAB + 11 Coriell, 3x technical replicates each
- DNA sheared to > 10kb

Target Set

49 genes(full-length for 39 + key hotspots for 10;
includes UGT1A1, DPYD, CYP2D6/CYP2C family, HLA loci, etc.).

Sequencing & Analysis

ONT × Twist targeted capture; basecalling/QC; coverage by mosdepth; small variants via clair3; benchmarking via vcfeval; wf-pgx outputs; haplotypes via PharmCat; CYP2D6 structure via Chinook. For benchmarking, variants with ≥30× site coverage were included.

QC / Acceptance

QC thresholds: Mean base Q >Q25; read-length N50 >3 kb; target coverage >100× (panel-level mean); benchmarking inclusion ≥30× at sites under test.

Targeted Long Reads Deliver High-Quality, Long Inserts

Target capture yields <1M reads/sample by design, but those reads are high-quality (>Q25) and long (N50 >3–4 kb), ideal for spanning repeats and phasing across full genes. Read-length clustering reflects expected fragmentation differences by line.

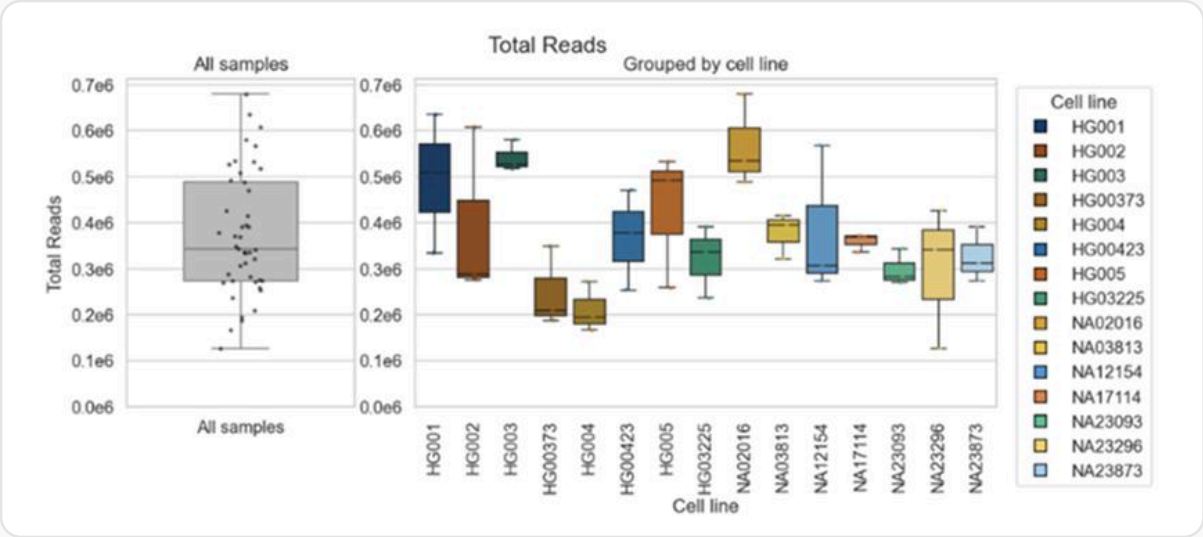


Figure 1A - Targeted Yield Per Sample (reads/barcode): Targeted capture produces fewer, richer reads by design. This plot shows total reads per sample (three technical replicates per line; log-scaled y-axis recommended). Replicates cluster tightly within each line, indicating stable library/capture performance. Note that QC is driven on-target coverage, not absolute read count; samples proceed if downstream panel-mean coverage ≥100× (reported in Fig. 2).

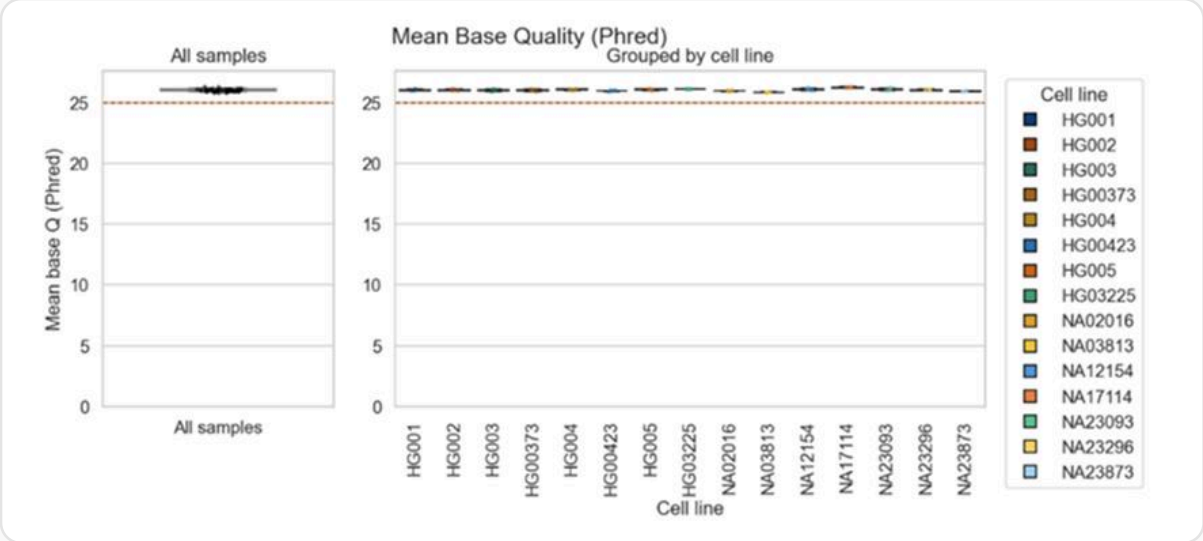


Figure 1B - Base Quality (mean Q-score per sample): Reads are consistently high quality (target: >Q25) across all lines/replicates. Points represent per-sample mean Q with IQR whiskers; a horizontal reference band at Q25 marks the acceptance gate. Stable Q supports confident small-variant calling and phasing across full genes.

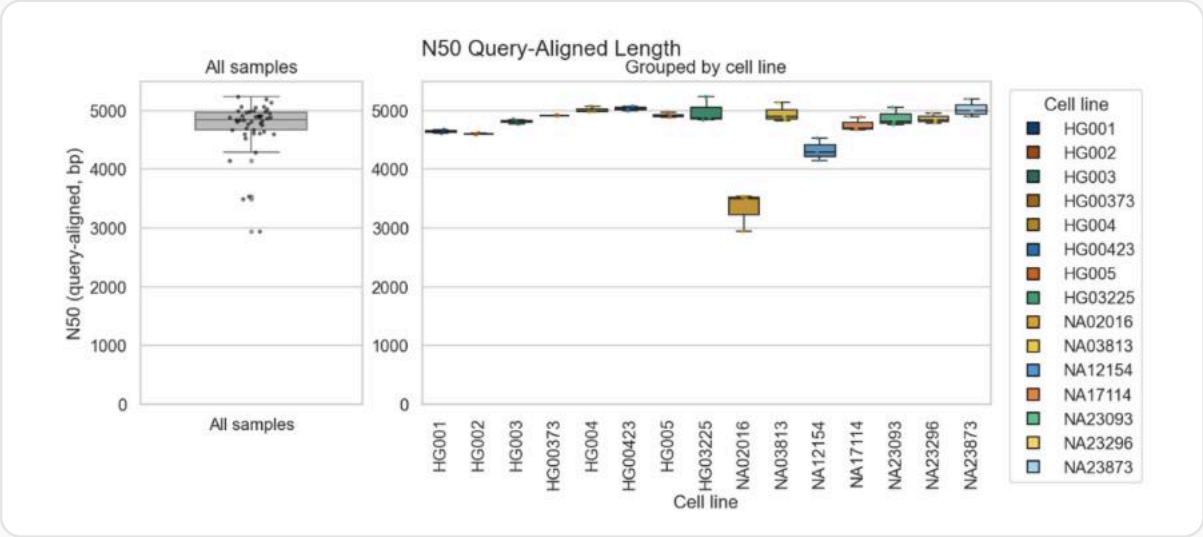


Figure 1C - Insert length (N50, query-aligned): Long inserts enable repeat sizing, phasing, and CNV/hybrid resolution. This chart shows N50 aligned read length per sample; most replicates land >3–4 kb with tight clustering by line. A horizontal reference band at 3 kb highlights the practical threshold for spanning UGT1A1*28 repeats and long haplotype blocks.

Capture is Efficient: High On-target Coverage Panel-wide

Panel mean coverage exceeds 100×, supporting confident calling across targets; a small subset of DPYD sites dip below 30× but did not impair DPYD call accuracy downstream.

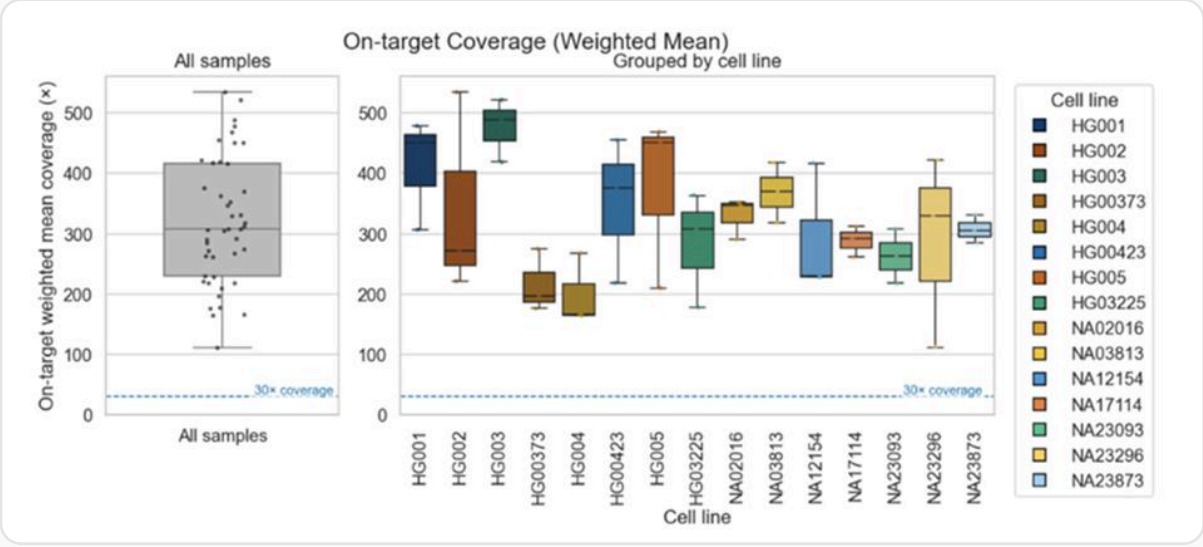


Figure 2A – Panel Mean On-Target Coverage Per Sample: Target capture delivers deep, uniform coverage across the panel. Each point (three technical replicates per line) shows the panel-mean on-target coverage for a sample; whiskers indicate IQR. The horizontal reference band at 30× marks the per-site callability/ benchmarking floor; panel means typically sit well above this band, confirming ample depth for small-variant calling and haplotyping.

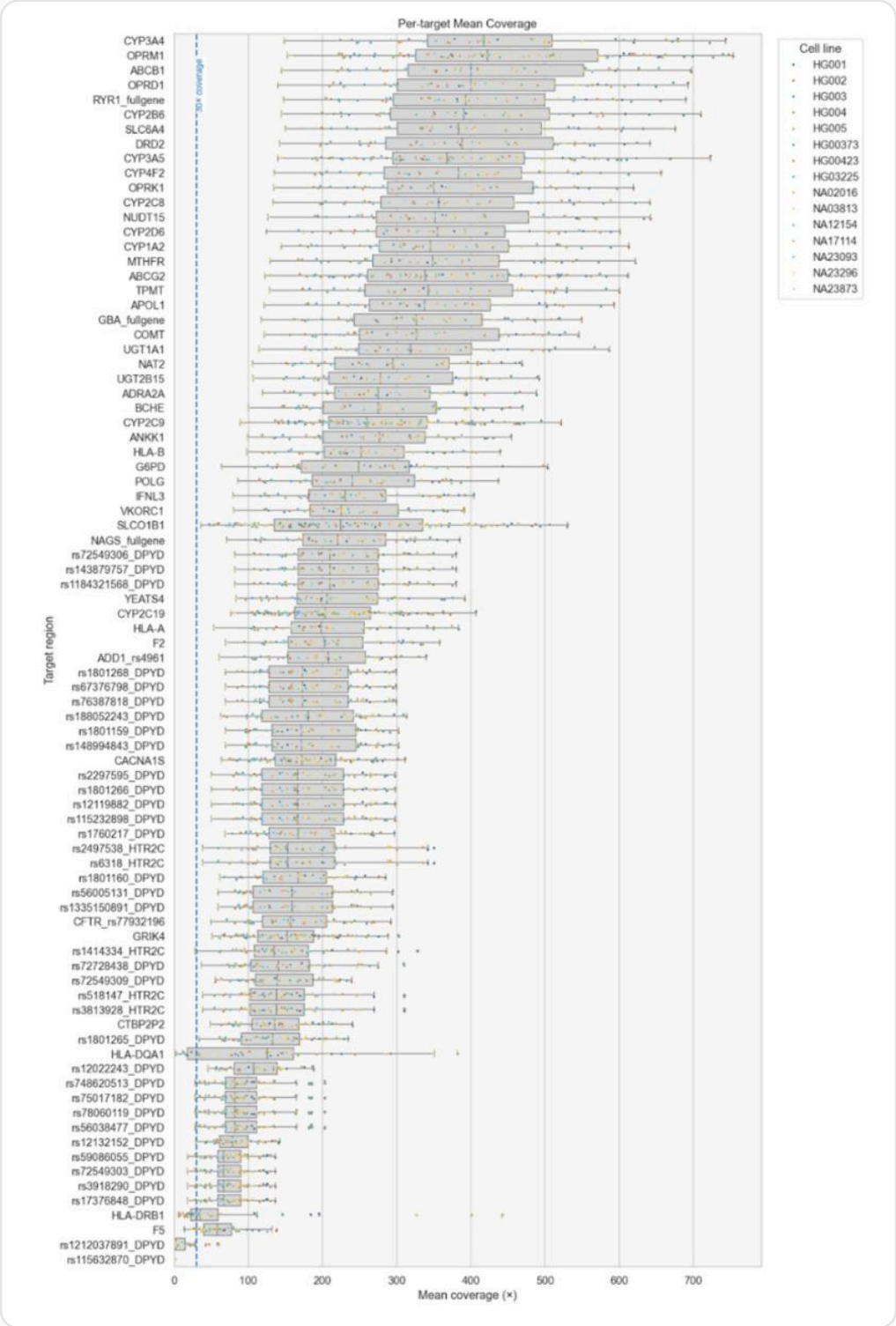


Figure 2B — Per-Target Coverage Waterfall (all genes/sites): Bars represent mean coverage per target (gene/exon/site), ordered from highest to lowest across all samples to reveal any low-depth tails. The shaded zone below 30x marks the benchmarking inclusion floor; targets falling into this band are flagged in the table (right) and annotated by gene (e.g., DPYD sites). Despite a small subset dipping below 30x, downstream accuracy for key loci remained intact (see Fig. 3).

Small-variant Accuracy vs GIAB is Decision-grade

Across GIAB truth sets (5 lines x 3 reps), precision/recall are high with F1 >0.95; AUPRC >0.92 across lines—replicates cluster tightly, indicating robust, repeatable performance.

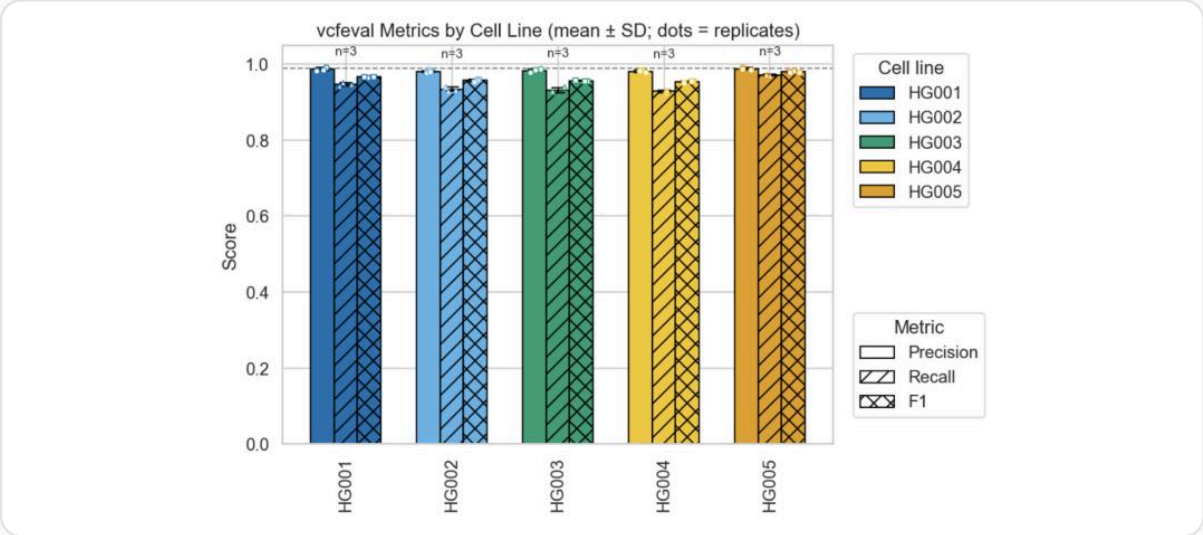


Figure 3A – Accuracy Summary By Line (F1 score): Bars (mean ± SD across 3 replicates) show F1 per GIAB cell line under the study's fixed filters and ≥30× site inclusion gate. A horizontal reference band at F1 = 0.95 marks the decision-grade target. Replicates cluster tightly within each line, indicating stable calling across runs.

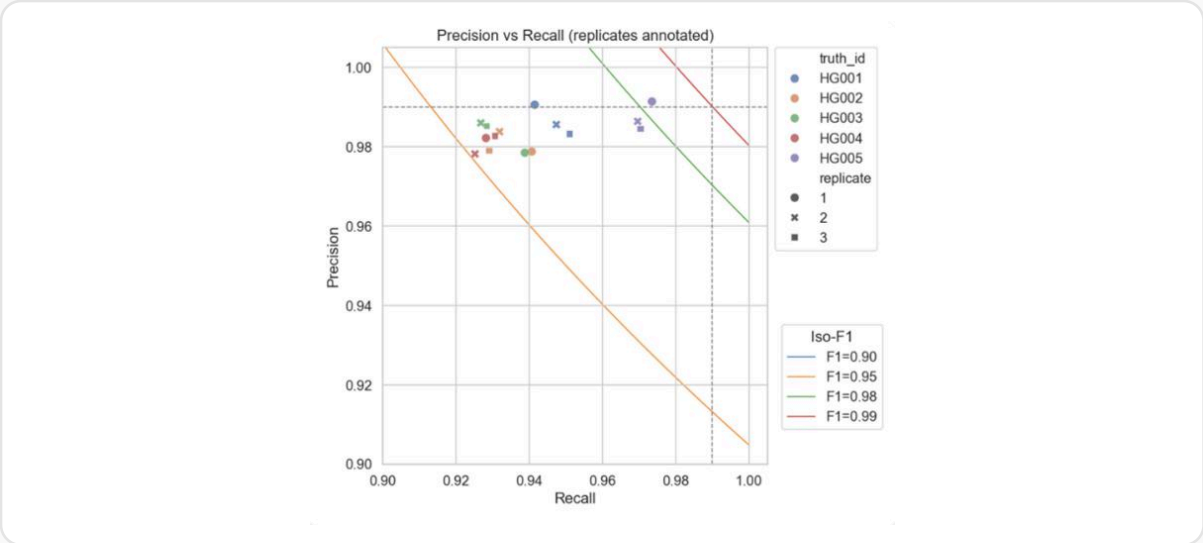


Figure 3B – Precision–Recall Trade-Off (replicate scatter): Each point is a replicate callset plotted by precision vs recall against GIAB truth; light grey iso-contours indicate constant F1. All replicates land in the high-precision/high-recall quadrant, demonstrating robustness of the pipeline at the preset thresholds (sites ≥30x; standard small-variant filters as in Methods).

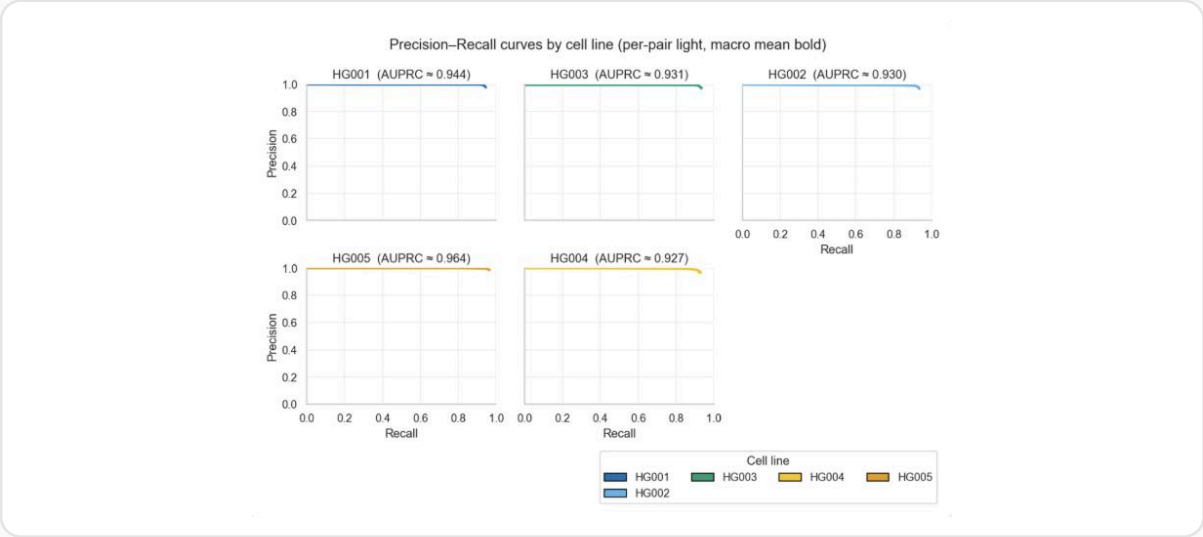


Figure 3C – PR Curves Per GIAB Sample (AUPRC): Precision–recall curves for each GIAB sample (replicates overlaid). AUPRC values (labels) remain high across lines, with minimal spread between replicates, confirming repeatable performance. Include the micro-average and macro-average curves for a panel-level view.

Small-variant Accuracy vs GIAB is Decision-grade

Across truth-set lines, PharmCat diplotypes (from wf-pgx) and Chinook-derived CYP2D6 structures are concordant with references under the study's inclusion gates (sites ≥30x). A small number of outliers occur where biology is intrinsically complex (CYP2D6 CNVs/ hybrids and rare/complex DPYD contexts), but they follow expected patterns (e.g., hybrid breakpoints, allele-specific depth skews) and are addressed via predefined review steps (split-read/assembly evidence, per-locus depth checks).



Figure 4A – CYP2D6 Structural Haplotypes (Chinook): Per-sample CYP2D6 star-allele calls with copy number and hybrid structure (e.g., *1/*4×2, *36+*10). The table pairs expected vs called haplotypes for each line; status chips indicate Match / Review.

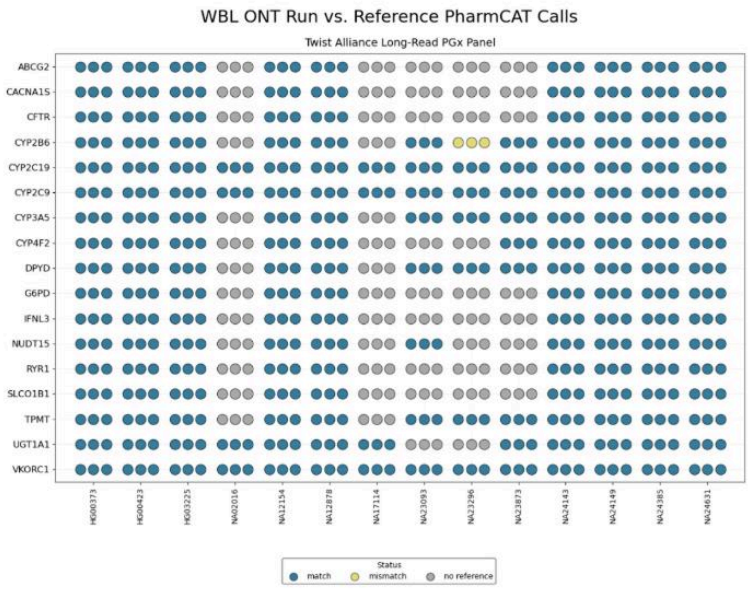


Figure 4B – PharmCat Diplotype Concordance Matrix (long-read vs reference): Tiles show diplotype concordance per sample for priority genes (UGT1A1, DPYD, CYP2D6, plus panel genes with references available). Color encodes match/mismatch. Rows = samples; columns = genes.

Conclusion

A long-read PGx panel with ONT × Twist is ideal for sponsors who need native diplotypes and structure in early development. From a routine 2 mL EDTA tube, the workflow returns full-gene UGT1A1 and DPYD with phase and repeat sizing (UGT1A1*28), captures rare/ancestry-specific DPYD alleles, and resolves CYP2D6 copy number and hybrid architectures directly–no imputation, no stitch-together CNV assays.

Consolidating these insights in one assay yields cleaner exposure–toxicity covariates, fewer “unexplained” dose holds, and faster, more interpretable AE/PK/PK-PD readouts across diverse populations and study designs. (RUO; fit-for-purpose and clinical use-cases available with clear validation pathways.)

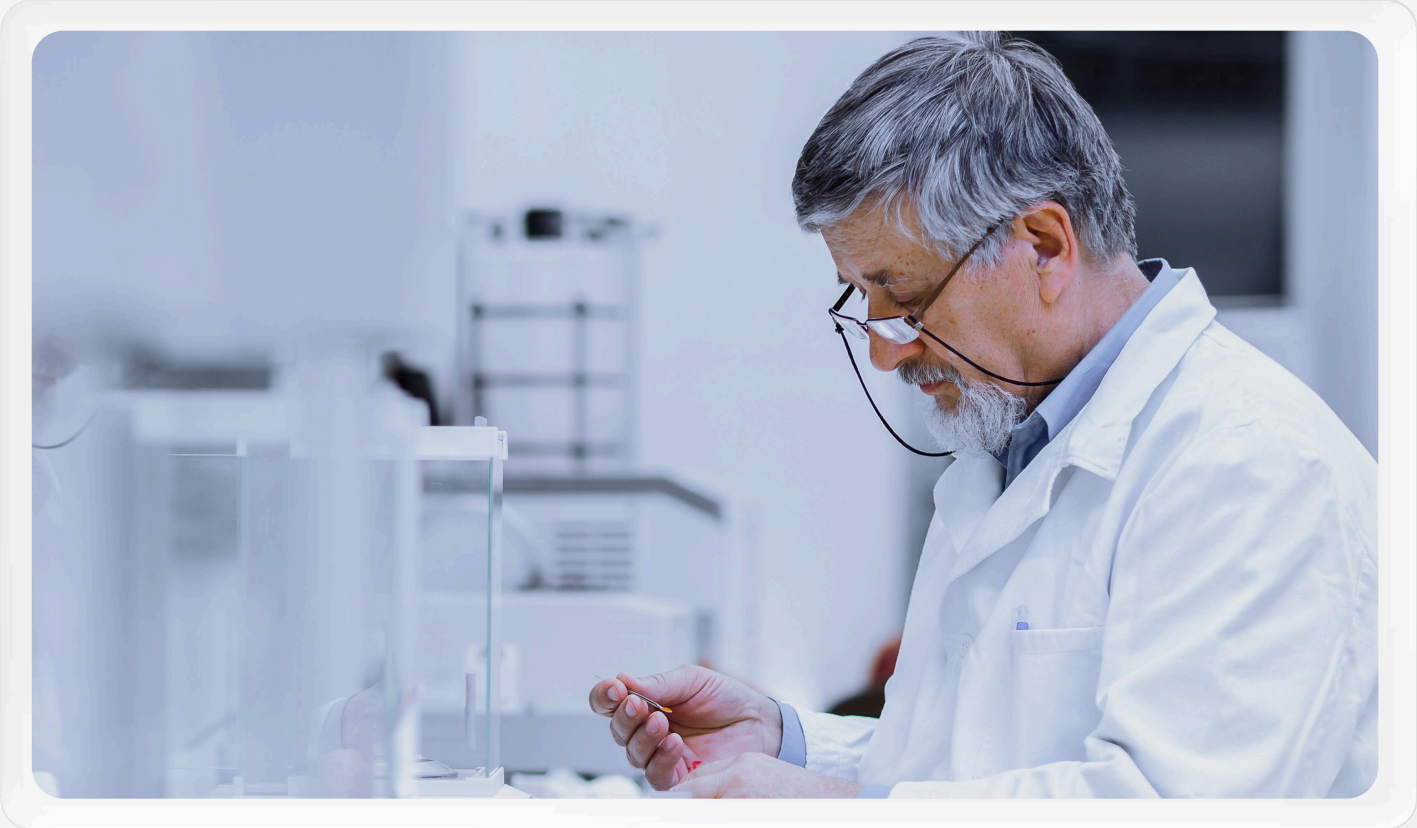
Unlock the full potential of your research with Renew

Visit renewbt.com

Contact us at info@renewbt.com

Notes

Notes



About Renew

Renew is building the next generation of sequencing-driven diagnostics and translational research solutions through native and long-read technologies, transforming genetic, epigenetic, and transcriptomic complexity into clinical products, research assays, and scalable programs designed to improve patient care.

Visit renewbt.com Contact us at info@renewbt.com