



Targeted Long-Read Sequencing and Pharmacometabolic Profiling in 75 Breast Cancer Patients: Oxford Nanopore Adaptive Sampling and PacBio Cas9 Targeted Sequencing

Monica J. Wolfe¹, Chloe Ziemelis¹, Summer Salman¹, Anika M. Patel¹, Haoran Li¹, Pranjal Srivastava¹, Rachel Case¹, Nina Gill¹, Yisang Moon¹, Nideesh Srivastavan¹, Gregory A. Farnum¹, Isaac P. Farnum¹, Zachary T. Freeman³, Balram Chowbay², Gilbert S. Omenn¹, Brian D. Athey¹

¹Gilbert S. Omenn Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor

²National Cancer Centre Singapore, Singapore

³Transgenic Animal Core, University of Michigan, Ann Arbor

Abstract

Accurate pharmacogenomic profiling of CYP2D6 is limited by conventional array-based and short-read methods, which fail to resolve structural variation and haplotype phase. We applied targeted long-read sequencing using Oxford Nanopore adaptive sampling and PacBio Cas9 enrichment to characterize CYP2D6 variation in breast cancer patients treated with tamoxifen. Long-read sequencing identified a high prevalence of structural variation, including hybrid alleles, duplications, and deletions, many of which are not detectable using standard genotyping approaches. These structural differences impact diplotype assignment and metabolizer phenotype classification. Our results demonstrate that long-read sequencing provides improved resolution of complex pharmacogenes and supports more accurate pharmacogenomic interpretation in clinically relevant contexts.

Introduction

Pharmacogenomic testing is used to predict drug response and reduce adverse drug reactions (ADRs), but structurally complex loci such as CYP2D6 remain difficult to resolve using array-based genotyping (e.g., PharmacoFocus) and short-read sequencing. These approaches frequently miss structural variants and hybrid alleles, leading to incorrect diplotype and phenotype assignments.

Long-read sequencing technologies, including Oxford Nanopore Technologies (ONT) and PacBio, enable direct resolution of structural variation and haplotype phase across complex genomic regions. In this study, we apply targeted long-read sequencing to evaluate CYP2D6 variation in tamoxifen-treated breast cancer patients and assess its impact on pharmacogenomic inference.

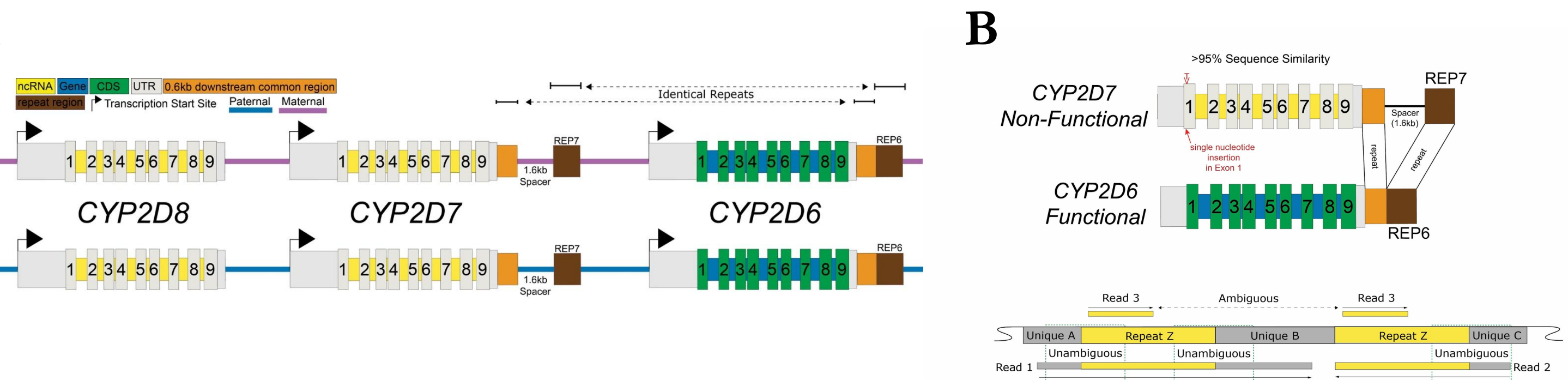


Figure 1. CYP2D6 Structural Complexity and the Power of Long-Read Sequencing. (A) CYP2D6 Locus Complexity: Schematic representation of the CYP2D6 gene and its highly homologous pseudogenes CYP2D7 and CYP2D8. The locus contains repeated elements (REP6 and REP7) and near-identical exonic regions, making short-read sequencing prone to misalignment. Despite CYP2D6 being only 4.4 kb in length, accurate genotyping is challenging due to its close proximity and sequence similarity (>95%) to neighboring pseudogenes.

(B) Advantage of Long-Read Sequencing: Visualization of how long-read technologies span the CYP2D6–CYP2D7 region, resolving repetitive elements and distinguishing true CYP2D6 alleles from pseudogene-derived sequences. Long reads provide unambiguous phasing across unique regions and repeats, enabling confident diplotype assignment and structural variant detection.

Sequencing Workflow

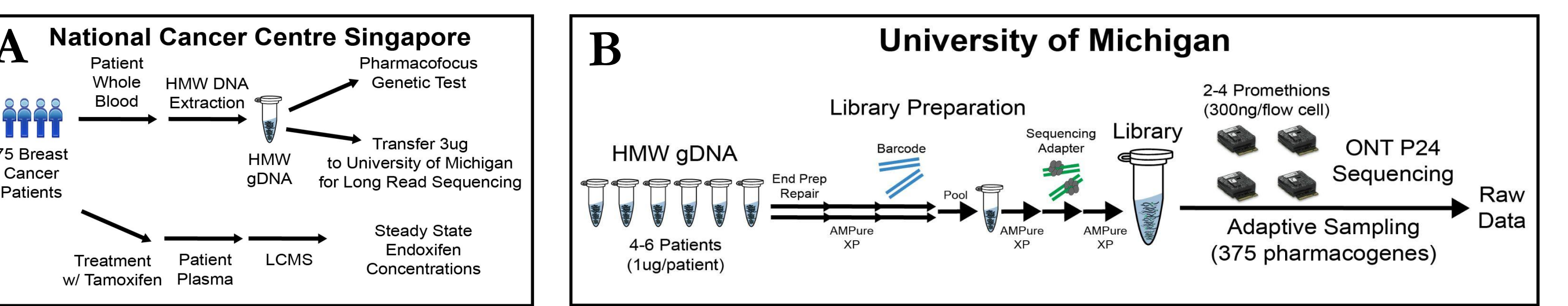
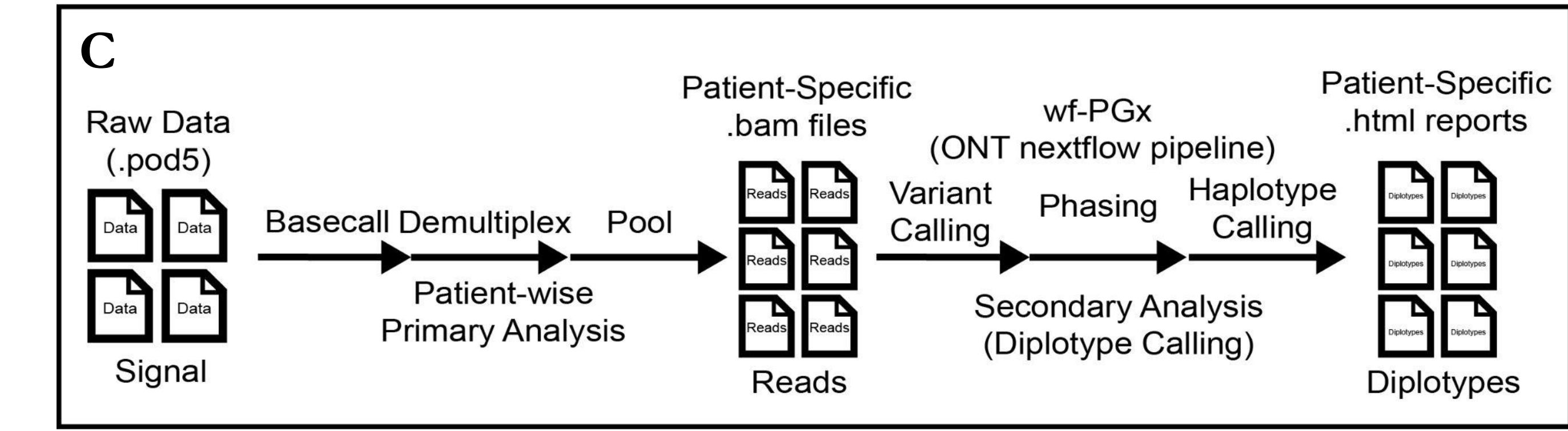


Figure 2. Study design and targeted long-read sequencing workflow in tamoxifen-treated ER+ breast cancer patients.

(A) High-molecular-weight genomic DNA was collected from patients at the National Cancer Centre Singapore during active tamoxifen therapy and processed through parallel workflows: clinical pharmacogenomic testing (PharmacoFocus) and targeted long-read sequencing. Plasma endoxifen levels were measured by LC–MS as a pharmacokinetic phenotype.

(B) At the University of Michigan, DNA underwent Oxford Nanopore library preparation and sequencing with adaptive sampling for enrichment of 375 pharmacogenes.

(C) Sequencing data were basecalled, demultiplexed, and analyzed using a pharmacogenomics pipeline for alignment, variant calling, phasing, and diplotype assignment (PharmCAT).

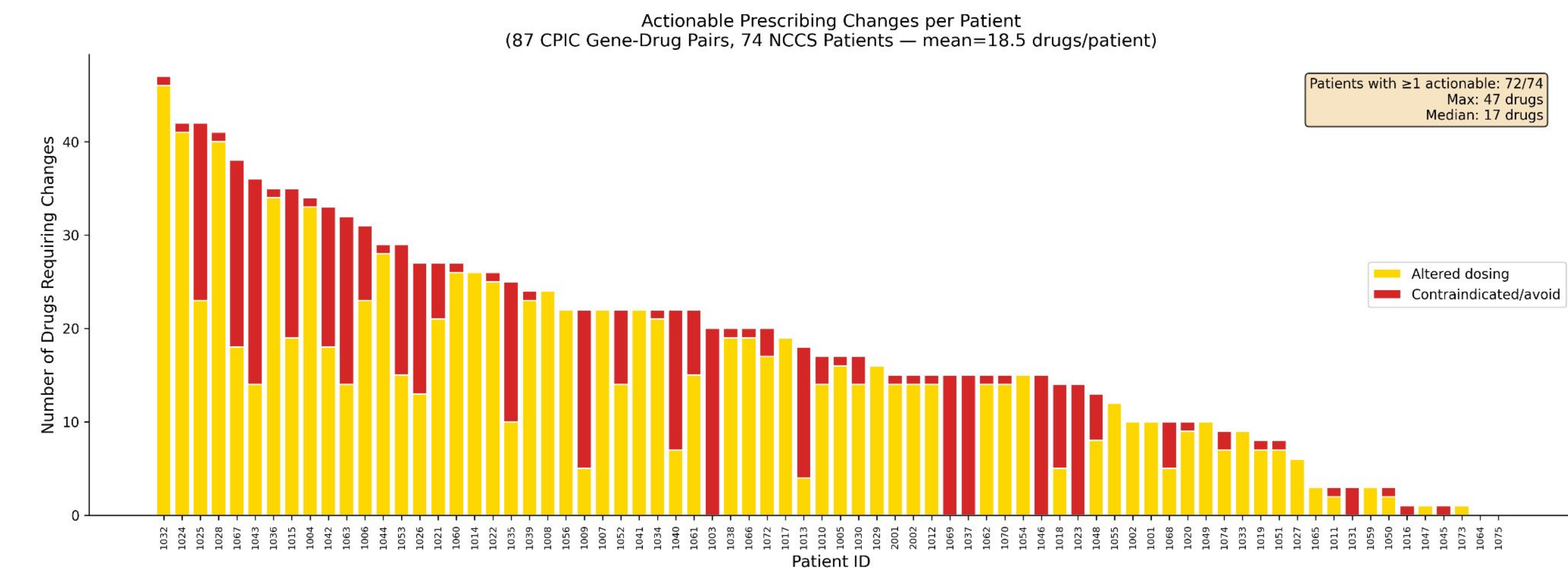


Figure 3. Number of actionable pharmacogenomic results per patient (n = 74). The distribution is right-skewed with mean 18.5 actionable drugs per patient (range: 4-31). Seventy-two of 74 patients (97%) carried at least one actionable result, and 89% carried results across ≥ 3 pharmacogenes.

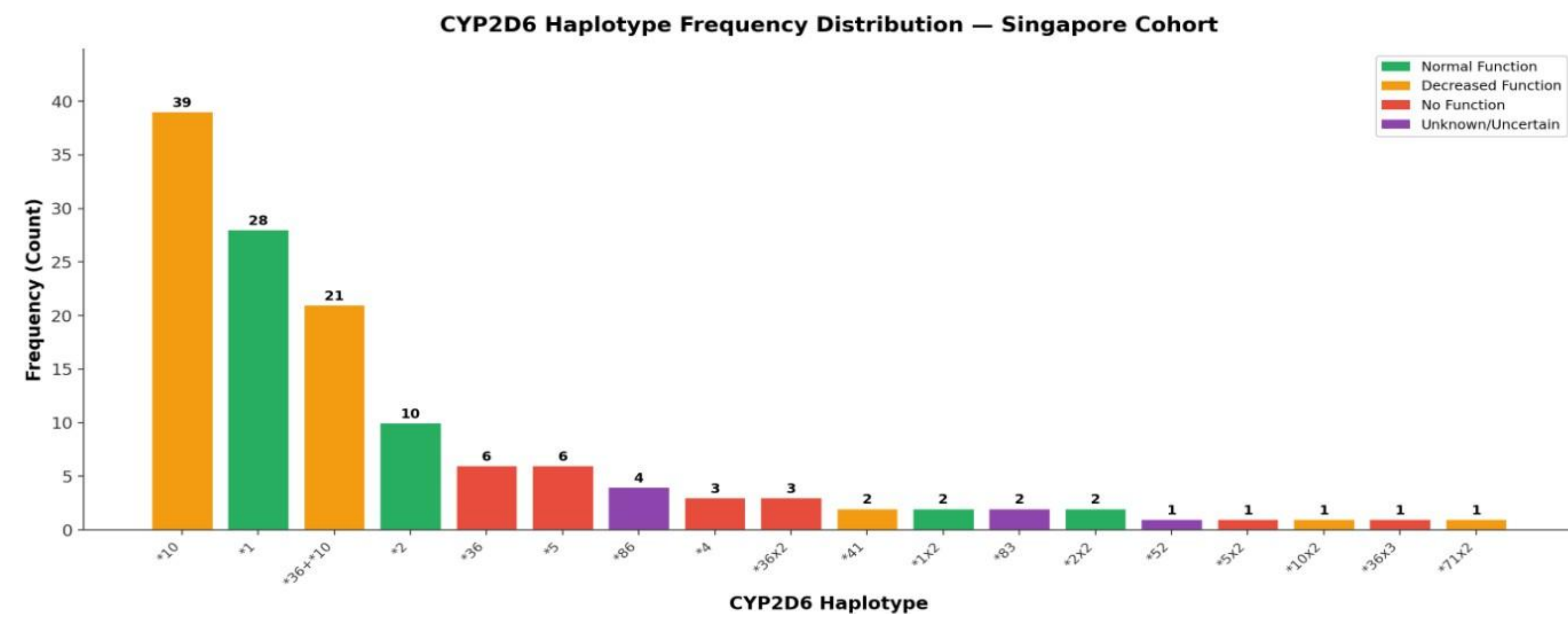


Figure 4. Distribution of CYP2D6 haplotypes across the patient cohort. The haplotype distribution reflects population-specific genetic variation. The predominance of the *10 haplotype – commonly enriched in East and Southeast Asian populations – dominates over *1.

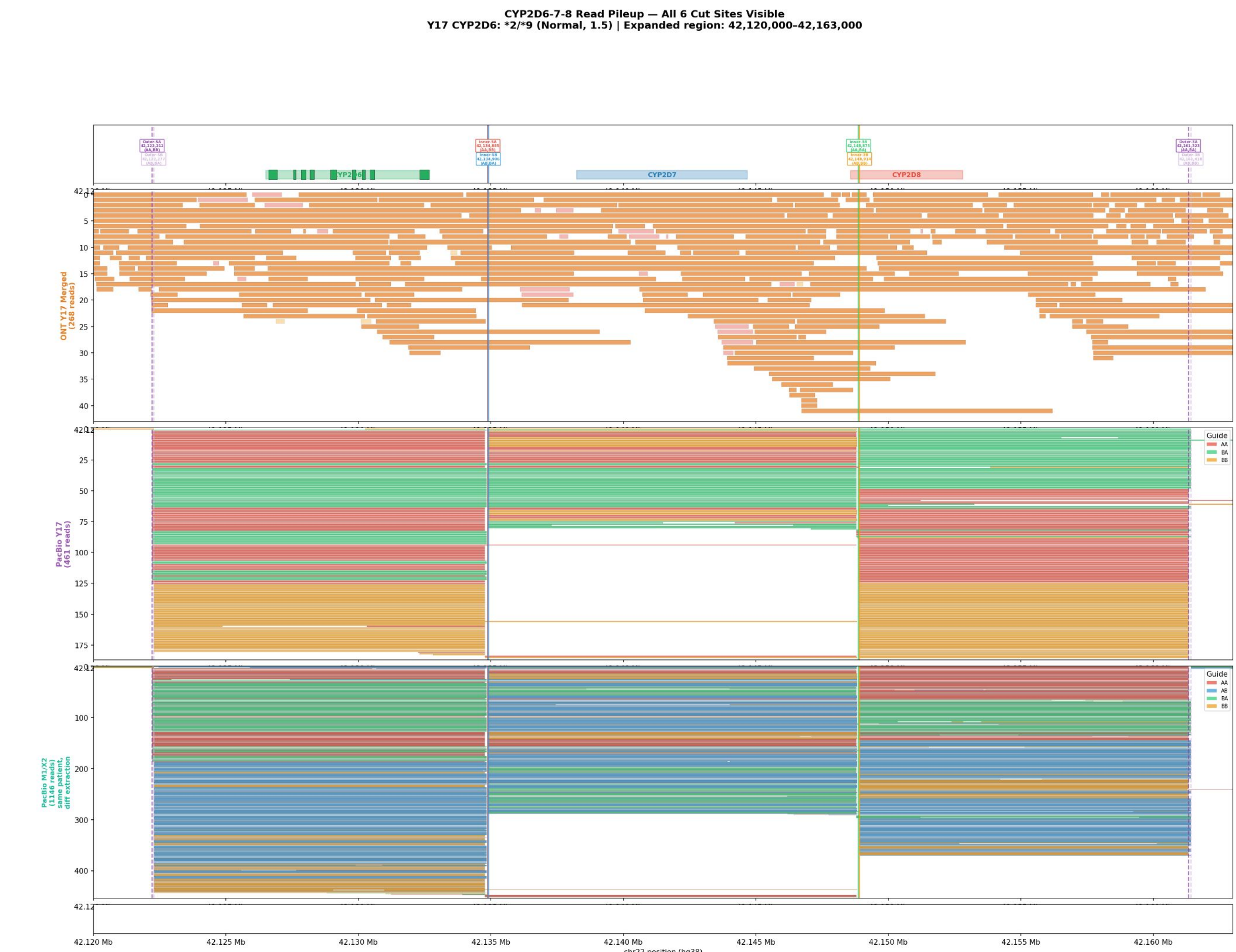


Figure 5. IGV alignment of Pacbio PureTarget enrichment over the CYP2D6 Locus. Combined read pileup across the enriched region, illustrates the coverage profile generated by overlapping fragments from multiple guide RNA pairs. The top panel, orange, demonstrates 20 x coverage using ONT. This is compared to 100 x using Pacbio PureTarget (bottom panels).

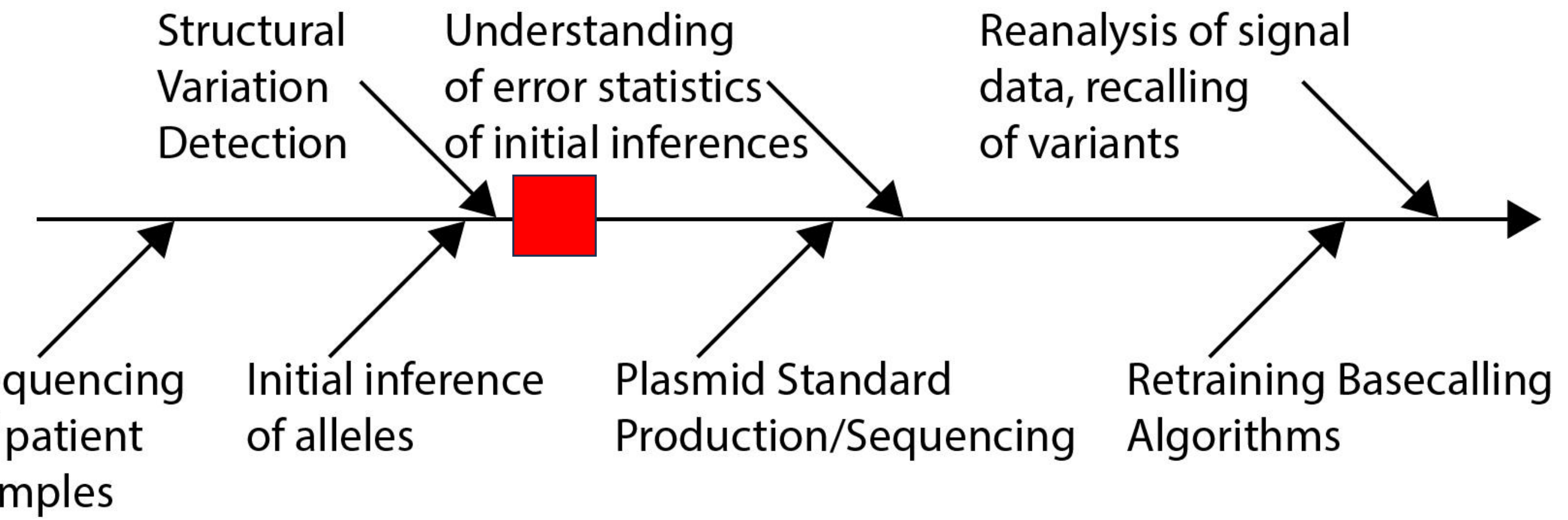


Figure 6. Iterative refinement workflow for CYP2D6 diplotype calling using long-read sequencing.

Initial sequencing of patient samples enables detection of structural variants and preliminary diplotype assignment. Results are evaluated to identify uncertain or low-confidence regions, followed by refinement of basecalling, alignment, and variant calling parameters. Reanalysis of sequencing data improves variant detection and phasing accuracy. This iterative process enhances confidence in diplotype assignment and overall pharmacogenomic interpretation.

Discussion and Conclusions

- Structural variation in CYP2D6 is common and clinically relevant
- Array-based and short-read methods miss key structural features
- ONT Long-read sequencing resolves copy number, hybrid alleles, and phase
- Accurate structural resolution is required for correct diplotype assignment
- Misclassification may impact metabolizer status and treatment decisions
- Pacbio Cas9 Targeted Sequencing of CYP2D6 provides ~100X coverage of the locus in control samples and will be applied to all 75 patients to further resolve diplotypes