Clinical Technical Validation of Pharmacogenomic STAR Allele and *CYP2B6/CYP2D6* Targeted Callers for Whole Genome Sequencing

Diana Toledo[1], Katie Larkin[1], Beena Thomas[1], Edyta Malolepsza[1], Christopher Kachulis[1], Kingshuk Mukherjee[2], Eric Roller[2], Michael DaSilva[1], Tom Howd[1], Marina DiStefano[1], Heidi Rehm[1,3], Niall Lennon[1], Sean Hofherr[1]
[1]The Broad Institute of MIT and Harvard, Broad Clinical Labs, Burlington, MA; [2]Illumina, Inc. [3]Massachusetts General Hospital, Boston, MA

Background

- Broad Clinical Labs (BCL) is a CLIA-licensed and CAP-accredited clinical laboratory that provides a clinical technical whole genome sequencing (cWGS) product to partner laboratories and clinical research programs.
- Here, we present the clinical validation of the Pharmacogenomic (PGx) outputs using the Illumina DRAGEN (**D**ynamic **R**ead **A**nalysis for **GEN**omics) v4.3.6 post-analytical pipeline including the STAR Allele Caller and the *CYP2B6/CYP2D6* Targeted Caller.
- Pharmacogenomics (PGx) evaluates how genetic variation — "star alleles" — affects drug metabolism, guiding dosing and drug selection to reduce adverse reactions.
- These callers generate PGx genotypes and predicted metabolizer phenotypes directly from cWGS data for 24 pharmacogenes included in the FDA PGx recommendations or designated as CPIC Level A genes.

Intended Use

- Detect star alleles and metabolizer phenotypes across 24 pharmacogenes from blood, saliva, buccal, or genomic DNA input, and provide as a clinical technical deliverable from cWGS.
- 22 genes via the Star Allele Caller (SNV/indel-based); CYP2D6 & CYP2B6 via the Targeted Caller (CNV/SV + small variants).

24 Pharmacogenes Assessed

Gene	Type	Caller	In control samples	In clinical samples
CACNA1S	SNP	Star	N	Y
CFTR	SNP	Star	N	Y
CYP2C19	SNP	Star	Y	Y
CYP2C9	SNP	Star	Y	Y
CYP3A4	SNP	Star	Y	Y
CYP3A5	SNP	Star	Y	Y
CYP4F2	SNP	Star	Y	Y
IFNL3	SNP	Star	N	Y
RYR1	SNP	Star	N	Y
NUDT15	SNP	Star	Y	Y
SLCO1B1	SNP	Star	Y	Y
TPMT	SNP	Star	Y	Y

Gene	Type	Caller	In control samples	In clinical samples
UGT1A1	SNP	Star	Y	Y
VKORC1	SNP	Star	Y	Y
DPYD	SNP	Star	Y	Y
G6PD	SNP	Star	N	Y
MT-RNR1	SNP	Star	N	N
BCHE	SNP	Star	N	N
ABCG2	SNP	Star	N	Y
NAT2	SNP	Star	Y	Y
F5	SNP	Star	N	N
UGT2B17	SNP	Star	Y	N
CYP2D6	CNV	Target	Y	N
CYP2B6	CNV	Target	Y	N

Sample Cohorts

Cohort	Source Material	N	Comparator
Control (Coriell)	Cell lines	25	GeT-RM dataset (12/30/24_Version 2)
Clinical samples	Blood	9	Infinium GDA + Enhanced PGx
Clinical samples	Saliva	12	Infinium GDA + Enhanced PGx
Clinical samples	Buccal	4	Infinium GDA + Enhanced PGx

- The validation cohort was chosen to preferentially test tier 1 alleles defined by the Association for Molecular Pathology, and represent PGx star alleles in the following 21 genes:
  - *ABCG2, CACNA1S, CFTR, CYP2B6, CYP2C9, CYP2C19, CYP2D6, CYP3A4, CYP3A5, CYP4F2, DPYD, G6PD, IFNL3, NAT2, NUDT15, RYR1, SLCO1B1, TPMT, UGT1A1, UGT2B17, VKORC1*
- Control (Coriell) samples were chosen to test the following alleles for genes with no AMP tier 1 alleles:
  - CYP4F2*3, SLCO1B1*5, SLCO1B1*15, UGT1A1*6, UGT1A1*27, UGT1A1*28, UGT1A1*36, UGT1A1*37, and UGT2B17*2
- Star allele calls from the following three genes were not included in this validation cohort:
  - *BCHE, MT-RNR1* and *F5*.

Validation Assessments & Results

Assessment	Design	Result
Analytical repeatability	Same sample, 3 reprocessing runs	100% identical alleles
Intra-run precision	3 blood samples × 3 replicates, same batch	100% concordant
Inter-run precision	3 blood samples, 3 independent batches	100% concordant
Input titration (LOD)	100 / 250 / 350 / 500 / 700 ng inputs	≥250 ng passed, 100ng below LOD

Pass criteria: ≥95% concordance for precision studies; titration must meet deliverable QC metrics.

- Intra-run and inter-run precision showed 100% of the samples had identical alleles returned (three samples run in triplicate for each).
- Analytical repeatability testing showed that 100% of samples had identical alleles returned when reanalyzed using the same pipeline.
- Input titration showed that an input of 250ng of gDNA and above had acceptable technical outputs, while an input of 100ng was too low to generate acceptable technical results.

Analytical Accuracy — Concordance

99.2%
Clinical genotype concordance vs orthogonal GDA array (25 clinical samples, 18 genes)

100%
Control allele concordance vs GeT-RM truth data (25 Coriell samples)



Pass criteria: ≥95% agreement overall; ≥80% per individual sample and per individual gene — all met.

Sources of Residual Discordance

- Array mis-calls at low-confidence SNPs (e.g., CYP2C19, CYP4F2); reassigned PGx nomenclature; and two CYP2D6 calls later confirmed correct by long-read PacBio sequencing.
- A DRAGEN patch (4.3.6-8-g38d9df55) corrected erroneous null phenotype calls found during final testing; all criteria remained met after re-test.

DRAGEN Outputs and STAR Allele JSON Output Example

Output Type	File Name Pattern	Caller
JSON with targeted callers and PGx caller results	<prefix>.targeted.json	PGx
TSV with gene/diplotype calls	<prefix>.star_allele.tsv	PGx
JSON with PGx results	<prefix>.star_allele.json	PGx
gVCF with PGx-relevant variants	<prefix>.select.gvcf	PGx
TSV with metabolism status for CYP2D6	<prefix>.cyp2d6.tsv	CYP2D6
TSV with metabolism status for CYP2B6	<prefix>.cyp2b6.tsv	CYP2B6


```
{
  "gene": "CYP3A5",
  "geneId": "HGNC:2638",
  "starAlleleDatabaseSource": "PharmGKB Database Version: Snapshot-2022.01.01",
  "genotype": "*3/*3",
  "genotypeQuality": 43,
  "phenotypeDatabaseAnnotation": "Poor Metabolizer",
  "supportingVariants": [
    {
      "alleleId": "*3",
      "chrom": "chr7",
      "pos": 99672916,
      "ref": "T",
      "alt": "C,<NON_REF>",
      "gt": "1/1",
      "quality": 43
    }
  ],
  "variantStarAllelesFound": "*3",
  "missingVariantSites": []
}
```


Conclusions

PASSED. The Illumina DRAGEN v4.3.6 Star Allele Caller and CYP2D6/CYP2B6 Targeted Caller met all pre-defined acceptance criteria for precision, repeatability, and analytical accuracy — supporting clinical use of this technical PGx calling method for clinical WGS.

Limitations

- **Orthogonal comparator was lower-sensitivity than WGS**: Clinical samples lacked known PGx truth data, so results were compared against the Infinium GDA array (less sensitive than the genome), and array-based genotyping was unavailable for UGT2B17, CYP2D6, and CYP2B6.
- **Incomplete truth data coverage**: GeT-RM lacked truth data for nine genes (CACNA1S, CFTR, IFNL3, RYR1, G6PD, MT-RNR1, BCHE, ABCG2, F5); BCHE, MT-RNR1, and F5 could not be assessed at all due to no available orthogonal data.
- **DPYD handling deviates from CPIC guidance**: DRAGEN matches DPYD to defined haplotypes rather than using the CPIC activity-score approach, causing it to fail to call a haplotype (return "./.") in 5 of 24 validation samples.
- **Minor nomenclature inaccuracies with no clinical impact**: Outdated/imprecise allele naming was seen for NAT2 (NAT212 reported as NAT24), G6PD (Mediterranean haplotype mislabeled), and MT-RNR1 (*930G>A reported as reference) — all functionally equivalent under current annotations.
- **Technical validation only:** Clinical interpretive reporting requires separate validation.
- **Practical/scope caveats**: This clinical validation was run using Illumina-sequenced data on the DRAGEN pipeline v4.3.6, which is not the newest or most updated version. Newer and upcoming DRAGEN versions have acknowledged and corrected many of the limitations noted above. Future directions for our clinical WGS assay + PGx involve re-validation using a newer DRAGEN version, and validation of PGx calling off of our cBGE (clinical Blended Genome Exome) assay, see poster 2038F.

Contact BCL

Learn more at https://broadclinicallabs.org/ Contact us at https://broadclinicallabs.org/contact-us/