Executive Summary
Pharmacogenomics

- 90-99% of the population has at least one actionable variant in an established pharmacogenomics (PGx) gene.\(^1\)\(^-\)\(^4\)
- An estimated 64.8% of medical home patients in the United States would have been exposed to a medication with a known PGx-association within a 5 year period.\(^5\)
- Variations in PGx genes can influence the absorption, distribution, metabolism and/or excretion of certain medications, leading to variations in efficacy or risk of side-effects.

Color’s Medication Response Test

- Color has developed a high-sensitivity, low-cost, next generation sequencing assay that identifies well-established alleles in 14 PGx related genes.
- In a validation study, using established cell-line controls and DNA derived from peripheral blood and saliva, the Color test accurately identified 534 out of 534 reportable alleles and variants.

Introduction

About 50% of all Americans have taken a prescription drug within the last 30 days. Additionally, 23% of the population have taken three or more prescription drugs in the last month.\(^6\) However, it is estimated that only 50% of patients respond favorably to their medications.\(^7\)\(^,\)\(^8\) Due to the large number of prescriptions and polypharmacy (the use of multiple drugs concurrently), a significant subset of the population is at an increased risk for therapy failure and/or adverse drug reactions (ADRs).\(^5\) While the exact number of ADRs is not precisely known, it has been estimated that over 2 million serious ADRs occur each year, making it the fourth leading cause of death in the United States with an estimated annual cost of $136-177 billion.\(^9\)\(^-\)\(^12\)

In a clinical setting, the goal of pharmacogenomics (PGx) is to characterize genetic variation that impacts the response to pharmacotherapy, with the aims of maximizing therapeutic benefit, reducing therapy “trial and error,” and reducing ADRs. It has been determined that between 90-99% of the population has at least one actionable variant in an established PGx related gene.\(^1\)\(^-\)\(^4\)

Established guidelines exist to support the effective use of PGx information in practice. The Clinical Pharmacogenetics Implementation Consortium (CPIC) was established with the goal of assisting physicians in understanding how to utilize PGx information. CPIC has built a standardized grading system to evaluate the levels of evidence linking genotypes to phenotypes, assigning phenotypes to clinical genotypes, prescribing recommendations based on genotype/phenotype, and a standardized method for assigning strength to each prescribing recommendation.\(^13\)\(^-\)\(^15\) CPIC guidelines have been endorsed by the following professional societies: The Association for Molecular Pathology (AMP), American Society for Clinical Pharmacology and Therapeutics (ASCPT), The American Society of Health-System Pharmacists (ASHP).

Due to variability in drug response rate and the risk of ADRs, the FDA has implemented PGx recommendations on over 120 drug labels.\(^16\) In addition, the FDA has also created guidance detailing how to incorporate PGx data into new drug submissions.\(^17\) Other professional medical organizations and international institutions have created additional frameworks for utilizing PGx information and inclusion of PGx information in a manufacturer’s label including: the Canadian Pharmacogenomics Network for Drug Safety (CPNDS), the Royal Dutch Association for the Advancement of Pharmacy Pharmacogenetics Working Group (DPWG), Health Canada (Santé Canada) (HCSC), the European Medicines Agency (EMA), and the Pharmaceuticals and Medical Devices Agency, Japan (PMDA).

PGx analysis and reporting focuses on the identification of previously described haplotypes, called “star-alleles”. Star-alleles are characterized by a set of observable genetic variants, some of which are functional and some of which are markers for a larger haplotype. \(^*1\) indicates that none of the interrogated alleles are definitely present, and is assay and analysis dependent; \(^*1\) does not eliminate the possibility that an unanalyzed star-allele, or that a novel loss or gain of function variant is present.\(^18\)

Traditional PGx analysis has been performed using
various targeted genotyping technologies. The Color test derives PGx analysis from NGS data, in four steps: (1) identification of genotypes at predetermined positions, (2) deriving the most likely diplotype to explain the observed genotypes, (3) converting the diplotype into an established metabolizer phenotype, and (4) based on phenotype, providing recommendations from sources such as CPIC.

Color has built upon its existing clinical grade, quality-controlled sequencing platform to analyze the most recognized drug-related genes. Here, we describe the validation approach and data supporting Color’s Medication Response Test.

Definitions

Genotype: The nucleotide sequence at a particular chromosomal position (i.e. AA, AG, or GG).

Haplotype: A large region of the genome inherited from one parent in a block. In this context, haplotype refers to a “version” of a gene or DNA sequence.

Diplotype: A pair of haplotypes on homologous chromosomes.

Star-allele: A previously described haplotype of a pharmacogenomic gene that has an established effect on the protein’s enzymatic activity. For each gene, a person has two star-alleles, one for each haplotype.

Drug-metabolizing Phenotypes:\textsuperscript{14}

- **Ultra-Rapid Metabolizer**: Increased enzyme activity compared to rapid metabolizers
- **Rapid Metabolizer**: Increased enzyme activity compared to normal metabolizers but less than ultrarapid metabolizers
- **Normal Metabolizer**: Fully functional enzyme activity
- **Intermediate Metabolizer**: Decreased enzyme activity (activity between normal and poor metabolizer)
- **Poor Metabolizer**: Little to no enzyme activity
- **Indeterminate**: Allele contains a variant that has not yet been characterized

Transporter Phenotype:\textsuperscript{14}

- **Increased**: Increased transporter function compared to normal function.
- **Normal**: Fully functional transporter function
- **Decreased**: Decreased transporter function (function between normal and poor function)
- **Poor function**: Little to no transporter function

Methods

Laboratory procedure

The Color Medication Response Genetic Test uses the same, proven technology utilized in the Color Hereditary Cancer Genetic Test and the Color Hereditary High Cholesterol Genetic Test. Our laboratory, certified by CLIA (05D2081492) and accredited by CAP (8975161), has developed a systematic process of automated laboratory protocols and tailored bioinformatics analyses to achieve sensitive, accurate, and reliable results. This process is based on chemistry from industry leaders such as Agilent, Roche, IDT, and Illumina, and automation solutions from Perkin Elmer, Hamilton, and HighRes robotics. All laboratory procedures and chemistries have been optimized for this application. The laboratory process starts with the isolation of high-quality genomic DNA purified from blood or saliva (Chemagen Magnetic Bead Technology, Perkin Elmer), followed by fragmentation and attachment of sequencing adaptors tagged with dual-unique molecular barcodes (Kapa Hyper Plus Kit, Roche). Barcoded samples are enriched for the assayed genomic regions of interest using a hybrid-capture procedure (SureSelect, Agilent) followed by paired-end sequencing (NovaSeq 6000, Illumina). The generated sequencing data are transferred to our bioinformatic pipeline for quality control and analysis.

Data Analysis - Calling Diplotypes from NGS

The bioinformatics pipeline aligns sequence data and calls variants using a suite of well established algorithms such as BWA-MEM, SAMtools, Picard and GATK. Copy number variants (CNVs), insertions and inversions are detected using CNVkit, LUMPY, and internally developed algorithms for read-depth analysis and split-read alignment detection. The assay’s minimum coverage requirements are 20X for each base of the reportable range. Median coverage typically ranges between 400-700X.

Diplotype calls are computed using an implementation
of the Aldy software\textsuperscript{19} and an internally developed tool, Diplo. Aldy uses data directly from sequence alignment files (.bam), enumerates possible solutions, and solves them using a combinatorial optimization framework. There are two customizations used for the Color implementation. First, CYP2D6 CNV status is inferred from read depth at representative regions based on data from CNVkit and then used as input to Aldy. \textsuperscript{20} Second, the gene configuration files used by Aldy are modified to match the canonical alleles included in Color’s reportable list.

For non-CYP2D6 diplotype calls, Diplo, a second caller designed in-house, is implemented to independently assess diplotypes. Using variant calls from the bioinformatics pipeline and a set of allele definitions focusing on defining SNPs, Diplo enumerates all possible diplotypes and their expected genotypes, returning diplotypes where the sample data matches expected genotypes.

Clinical Interpretation of Results

The functional status of alleles (haplotypes) is established based on CPIC consensus terms and recommendations.\textsuperscript{14} Alleles are then summarized into diplotypes and associated with the predicted phenotype using data from CPIC, PharmGKB and PharmVar. For genes that encode drug-metabolizing enzymes such as CYP2C19 and CYP2D6 the possible phenotypes are: ultra-rapid, rapid, normal, intermediate and poor metabolizers. For transporter and other protein genes such as SLCO1B1, the phenotypes are defined as increased, normal, decreased, or poor function. Once a phenotype is defined, clinical decision support information for drug selection and dosing is generated according to CPIC guidelines, Dutch guidelines, FDA drug package insert, or related literature. Performance of the full implementation is actively monitored, and results are verified and approved by licensed clinical molecular geneticists.

<table>
<thead>
<tr>
<th>Table 1: Included Genes &amp; Alleles</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CYP3A4</strong></td>
</tr>
<tr>
<td><strong>CYP1A2</strong></td>
</tr>
<tr>
<td><strong>CYP3A5</strong></td>
</tr>
<tr>
<td><strong>CYP4F2</strong></td>
</tr>
<tr>
<td><strong>DPYD</strong></td>
</tr>
<tr>
<td><strong>F5</strong></td>
</tr>
<tr>
<td><strong>IFNL3</strong></td>
</tr>
<tr>
<td><strong>NUDT15</strong></td>
</tr>
<tr>
<td><strong>SLCO1B1</strong></td>
</tr>
<tr>
<td><strong>TPMT</strong></td>
</tr>
<tr>
<td><strong>VKORC1</strong></td>
</tr>
</tbody>
</table>

Test Validation

Sample Selection

To validate the sensitivity, specificity, and precision of the Color Medication Response Genetic Test, a total of 117 independent and blinded samples were compared to previously characterized results. The validation includes DNA derived from saliva (n = 19), peripheral blood (n = 10), and cell lines (n = 88). The cell lines had consensus diplotypes reported by numerous studies.\textsuperscript{19,21–24} DNA from Blood and saliva samples were sent to an independent CLIA-certified laboratory for confirmation.

The validation consisted of samples with a diplotype status falling into one of the following groups:

- Known “negative” or reference allele samples (n = 31) with a “normal” metabolizer status. e.g. (*1/*1)
- Known “positive” samples (n = 86) with reportable, non-normal diplotypes.

All samples were blinded to the operators and treated under identical experimental conditions.
Data Analysis

As described above, *1 indicates the absence of any tested allele, and is assay and analysis dependent; additionally, specific reporting of certain alleles depends on the inclusion or exclusion of other related refining alleles. It is therefore possible that analytically equivalent results can be reported as different diplotypes by different laboratories. Because published documentation for cell lines often only includes diplotypes without sufficient information about the set of tested alleles or the underlying genotypes, analysis of validation results followed a two-step process. Diploype matches were counted as concordant. In cases of discordance at the diplotype level, a comparison of underlying contributing genotypes was made. Cases where all overlapping underlying genotypes were consistent were also counted as concordant.

Validation Study

Table 2: Validation Samples

<table>
<thead>
<tr>
<th>Gene</th>
<th># of Unique Samples with reference genotype or diplotype</th>
<th># DNA from Blood</th>
<th># DNA from Cell line</th>
<th># DNA from Saliva</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP1A2</td>
<td>47</td>
<td>10</td>
<td>18</td>
<td>19</td>
</tr>
<tr>
<td>CYP2C19</td>
<td>58</td>
<td>10</td>
<td>29</td>
<td>19</td>
</tr>
<tr>
<td>CYP2C9</td>
<td>48</td>
<td>10</td>
<td>19</td>
<td>10</td>
</tr>
<tr>
<td>CYP2D6</td>
<td>117</td>
<td>10</td>
<td>88</td>
<td>19</td>
</tr>
<tr>
<td>CYP3A4</td>
<td>33</td>
<td>10</td>
<td>4</td>
<td>19</td>
</tr>
<tr>
<td>CYP3A5</td>
<td>32</td>
<td>10</td>
<td>3</td>
<td>19</td>
</tr>
<tr>
<td>CYP4F2</td>
<td>7</td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>F5</td>
<td>29</td>
<td>10</td>
<td></td>
<td>19</td>
</tr>
<tr>
<td>IFNL3</td>
<td>29</td>
<td>10</td>
<td></td>
<td>19</td>
</tr>
<tr>
<td>SLC01B1</td>
<td>45</td>
<td>10</td>
<td>16</td>
<td>19</td>
</tr>
<tr>
<td>TPM7</td>
<td>30</td>
<td>10</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td>VKORC1</td>
<td>59</td>
<td>10</td>
<td>30</td>
<td>19</td>
</tr>
<tr>
<td>Total</td>
<td>534</td>
<td>110</td>
<td>215</td>
<td>209</td>
</tr>
</tbody>
</table>

The Color Medication Response Test showed 100% concordance across all genes in all tested samples. In this dataset, 505 results were compared, with no false positives called in any of the 117 samples. In addition, all samples submitted for an independent CLIA-certified were confirmed and no additional variants of relevance were detected.

Table 3: Summary of validation studies

<table>
<thead>
<tr>
<th>Metric</th>
<th>Color Assay Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>&gt; 99%</td>
</tr>
<tr>
<td>Specificity</td>
<td>&gt; 99%</td>
</tr>
</tbody>
</table>

CYP2D6 copy number

Fig 1: An initial subset of client samples were examined for copy number in representative sites within CYP2D6. We observe distinct separation of copy numbers.

CYP2D6 imparts extra complexity. In addition to being adjacent to two highly homologous pseudogenes, CYP2D7 and CYP2D8, it has over 100 reported alleles that vary in frequency by ethnicity. These allelic variants are composed of single nucleotide polymorphisms (SNPs), insertions and deletions, copy number variants, larger rearrangements, and hybrid gene conversion events. In particular, copy number changes are quite common. An estimated 12.6% of the US population has zero, one, or three or more copies.

To derive a clear signal amidst these homology complications, CYP2D6 copy number is assessed by an analysis of a region including exon 6 and intron 6. To confirm that homology does not confound copy number assessment, observed copy number across a
set of 6266 samples was evaluated. A clear separation of integer copy numbers is observed. In addition, the validation set included 11 known copy number variants, all accurately detected.

Conclusions
Actionable PGx variants are common and one is likely to be prescribed a medication with an established PGx interaction over the course of their lifetime. Color has created an accurate and affordable preemptive PGx testing platform on NGS technology to guide medication selection and dose adjustment as early as the first fill.

References


20. Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: Genome-Wide Copy Number Detection and


