Comparison of Genome Sequencing and Clinical Genotyping for Pharmacogenes

W Yang[1], **G Wu**[2], **U Broeckel**[3], **CA Smith**[1], **V Turner**[4], **CE Haidar**[1], **S Wang**[2], **R Carter**[2], **SE Karol**[1], **G Neale**[5], **KR Crews**[1], **JJ Yang**[1], **CG Mullighan**[4], **JR Downing**[4], **WE Evans**[1], and **MV Relling**[1]

[1]Department of Pharmaceutical Sciences, St. Jude Children's Research Hospital, Memphis, Tennessee, USA

[2]Department of Computational Biology, St. Jude Children's Research Hospital, Memphis, Tennessee, USA

[3]Human and Molecular Genetics Center, Medical College of Wisconsin, Milwaukee, Wisconsin, USA

[4]Department of Pathology, St. Jude Children's Research Hospital, Memphis, Tennessee, USA

[5]Hartwell Center, St. Jude Children's Research Hospital, Memphis, Tennessee, USA

Abstract

We compared whole exome sequencing (WES, $n = 176$ patients) and whole genome sequencing (WGS, $n = 68$) and clinical genotyping (DMET array-based approach) for interrogating 13 genes with Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines. We focused on 127 CPIC important variants: 103 single nucleotide variations (SNV), 21 insertion/deletions (Indel), HLA-B alleles, and two CYP2D6 structural variations. WES and WGS provided interrogation of nonoverlapping sets of 115 SNV/Indels with call rate >98%. Among 68 loci interrogated by both WES and DMET, 64 loci (94.1%, confidence interval [CI]: 85.6–98.4%) showed no discrepant genotyping calls. Among 66 loci interrogated by both WGS and DMET, 63 loci (95.5%, CI: 87.2–99.0%) showed no discrepant genotyping calls. In conclusion, even without optimization to interrogate pharmacogenetic variants, WES and WGS displayed potential to provide reliable interrogation of most pharmacogenes and further validation of genome sequencing in a clinical lab setting is warranted.

Pharmacogenetics is being widely adopted in clinical practice to provide personalized drug dosing with the goal of reducing toxicity and improving efficacy. We have implemented *TPMT* genetically guided therapy in treating acute lymphoblastic leukemia (ALL) since the

early 1990s.[1] Recently, we moved from testing single genes to pharmacogene-directed arrays such as the Affymetrix DMET array. We have been among the first to implement preemptive genomic testing to incorporate pharmacogenetic results in the medical record to assist in patient care.[2–5] DMET arrays have been shown to accurately interrogate pharmacogenes and to be able to infer genetic ancestry.[6, 7]

Recent advances in genomic sequencing technology have made it feasible to interrogate the whole genome (WGS) or exome (WES) of an individual. The cost of genome sequencing is decreasing, and the turnaround time and accuracy have improved.[8, 9] Thus, many hypothesize that genome sequencing, possibly done for other clinical purposes, might generate pharmacogenetic data that can be useful in the clinic.[10–12] However, the utility of genome sequencing to interrogate important pharmacogenomic genes has not been comprehensively examined, and the accuracy of genotyping has not been tested or compared to clinical pharmacogenetic testing.

In research laboratories at St. Jude Children's Research Hospital, we have performed whole genome sequencing for children with cancer as part of the Pediatric Cancer Genome Project[13, 14] and generated whole exome sequencing data for hundreds of patients. We have also enrolled a subset of patients on a protocol to facilitate clinical pharmacogenetics implementation, PG4KDS,[3, 4] in which pharmacogenes were typed with clinical pharmacogenetic testing in clinical genetic laboratories. We have two main goals herein: to compare clinical pharmacogenetic testing to results from genome sequencing for actionable pharmacogenes, and to compare the concordance of our clinical genotyping array-based results with those generated by genome sequencing.

RESULTS

Genes and variants

We focused on 13 genes with Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines (Table 1),[15–28] which include the pharmacogenes tested for in a directed platform, PGRNseq.[29] CPIC guidelines specifically address 127 variants that can affect clinical decision making (termed "CPIC important variants" herein, Supplemental Table 1). These CPIC important variants include 103 single nucleotide variations (SNVs) (95 in exonic region), 21 indels (20 in exonic region), two structural variations (*CYP2D6* copy number, i.e., *CYP2D6* whole gene deletion (CN = 0 or 1) or duplication (CN>2) and *CYP2D6/CYP2D7* hybrid structure variation), and specific alleles in the *HLA-B* locus (Supplemental Table 1).

Clinical pharmacogenetic testing

All clinical testing was performed in a Clinical Laboratory Improvement Amendments (CLIA)-compliant laboratory, and many of the results were posted in patients' medical records.[4] We genotyped 2,656 patients using the Affymetrix DMET array (Figure 1). The DMET array interrogated a total of 185 variants in 10 of the 13 CPIC genes, including 75 of the 127 CPIC important variants. *CFTR, HLA-B, IFNL3* were not represented on the DMET array. DMET achieved a call rate of >98% in 73 of the 75 CPIC important variants;

exceptions were *UGT1A1*28* and *CYP2D6*29*(V136I) for which the call rates were 90.3% and 94.1%, respectively (Figure 2, Supplemental Table 2).

CYP2D6 copy number was interrogated in 2,450 patients using separate real-time quantitative polymerase chain reaction (qPCR)-based assays based on three probes.[4] Possible *CYP2D6/CYP2D7* hybrids could also be inferred from the above qPCR assay when the three probes indicated different copy number estimates. However, these possible hybrid results were not clinically validated and released into the medical record.

We compared the genotyping results derived from clinical pharmacogenetic testing via Affymetrix DMET array or qPCR vs. those obtained from genome sequencing via the Illumina platforms (Supplemental Figures 1, 2).

Whole exome sequencing (WES)

WES data in a total of 636 patients were analyzed through the standard GATK pipeline (Figure 1). The median exome-wide coverage was 75× (range: 35× to 169×). Currently, standard WES analysis pipelines do not address *CYP2D6/CYP2D7* hybrids; seven CPIC important nonexonic variants were not interrogated by WES, including *VKORC1* promoter −163 9G>A, *CYP2C19* promoter −806C>T, three intronic SNVs including *CYP2D6*59* (2291G>A), *CYP3A5*3* (6986A>G), *SLCO1B1*17* (−11187G>A), and two *IFNL3* upstream SNVs. For the remaining 117 CPIC important coding SNV/indels, we observed good call rates (>98% in 636 patients) in 115 out of 117 variants (Figure 3, Supplemental Table 2, Supplemental Figure 1). The variants with low call rates included *CYP2D6*57* (R62W) and *CYP2D6*11* (883G>C), both of which had low WES coverage (median coverage 18.5×). Both *CYP2D6*57* and *CYP2D6*11* are rare in the general population (MAF < 1% in ExAC dataset),[28] and neither variant genotype was detected among called genotypes by WES. WES also provided genotyping results for two nonexonic variants even though they are not directly targeted by the exome capture. These included *CYP2D6*41* (2988G>A) and *UGT1A1*28*, both of which are close to targeted exons (35 bps and 52 bps away, respectively).

Among the 115 variants with call rates >98%, 51 variants were polymorphic within our study population. All polymorphic variants passed the Hardy–Weinberg equilibrium test ($P > 0.016$) in 396 patients of European genetic ancestry, 95 patients of African ancestry, and 86 Hispanic patients. All of the observed allele frequencies of the variants are consistent with those reported in the ExAC population of European ancestry (Supplemental Figure 3A).

Among the 636 patients with WES, 176 patients also had DMET data available. In all, 53 SNVs and 15 indels of the 124 CPIC important SNVs/indels were interrogated with >98% call rate by both platforms (Figure 2, Supplemental Table 2). Among a total of 11,935 genotypes compared between WES and DMET, only seven genotypes in four variants showed discrepancies (Table 2A, Figure 3, Supplemental Table 2). The discordant variants were all within *CYP2D6*, including: *CYP2D6*40* (1863_1864ins, *n* = 2), *20* (1973insG, *n* = 2), *4* (1846G>A, *n* = 2), and *2* (R296C, *n* = 1). Five discordant genotypes were called heterozygous by WES; all showed biased minor allele fractions (MAFrac <16%) even

though all passed GATK's default quality control. In contrast, all the concordant heterozygous genotypes by WES exhibited MAFrac >20% (Supplemental Figure 4A). Additional orthogonal genotyping methods would be needed to resolve the discrepancies between clinical genotyping and WES.

We generated 4-digit *HLA-B* allele types from WES using Polysolver.[30] *HLA-B* was not interrogated on the DMET array; but among those samples with WES, we determined their *HLA* diplotypes using clinical molecular methods for 66 patients.[6] The concordance was 95.5% (126 out of 132 haplotypes) at 4-digit resolution. Four of the mismatches were consistent at the 2-digit resolution, while the other two mismatches involved the relatively rare alleles *78* and *81*. CPIC important *HLA-B* alleles include *57:01* and *58:01*. There were two patients with *HLA-B*57:01* and two patients with *HLA-B*58:01*, all of whom had concordant *HLA-B* results by WES and conventional clinical assays (Supplemental Figure 5A).

We applied XHMM to infer copy number for *CYP2D6* based on WES.[31] There were 105 patients with *CYP2D6* copy number determined by a validated clinical qPCR assay.[4] WES generated concordant copy number results for 98 of 105 (93.3%, confidence interval [CI]: 86.7–97.2%) patients (Figure 4A). One patient was called as *CYP2D6* gain by clinical qPCR but not by WES, while two patients were called as *CYP2D6* gain by WES but not called as such by qPCR. One patient was called as single copy *CYP2D6* by qPCR, but was called diploid by WES. Two additional patients were called *CY2D6* gain by WES, but called as possible 2D6/2D7 hybrid by qPCR. Excluding the two patients with possible *CYP2D6/2D7* hybrids by qPCR, five patients (5.9%) were called *CYP2D6* gain by WES among 85 patients of European ancestry, a frequency that is not statistically different from previously reported (4.7% in Caucasians, $P = 0.60$).[32] One benefit of WES is that for patients with *CYP2D6* duplications, one may deduce which *CYP2D6* allele was duplicated by examining the ratio of read counts between reference and alternative genotypes (Supplemental Material, Supplemental Table 3, Supplemental Figure 6). There are software tools that can be used to enhance the performance of genome sequencing for CYP2D6.[33] Additional standard samples with known allele compositions would be needed to validate genome sequencing as a method to assess *CYP2D6* copy number.[34]

For *UGT1A1*28*, the DMET call rate was low (90.3%) for 2,656 patients, and thus we did not compare DMET results to WES or WGS. However, we had ancillary genotyping results for 240 patients by a locus-specific PCR assay.[35, 36] Three results were discordant between WES and the locus-specific PCR method (concordance rate 98.8%, CI: 96.3–99.7%), and in all cases, patients with (TA)7/7 by PCR were called (TA)6/7 by WES. The WES allele fractions for (TA)6 vs. (TA)7 in these cases were <0.14, whereas in the patients with concordant (TA)6/7 genotypes the minimum MAFrac was 0.29. This suggests that by applying a more stringent quality control criterion than the defaults provided by GATK, one could eliminate all the erroneous *UGT1A1*28* calls by WES, similar to the situation described for *CYP2D6*.

The average read depth of the coding regions of the 13 genes interrogated by WES was 66× (Supplemental Table 4, Supplemental Figure 7). We defined well-interrogated regions as

those having 10 reads in more than 95% of patients. Note that the four exons shared by the genes of the *UGT1A* family were not considered herein, since a capture probe for this region is not included in the exome library kit (TruSeq from Illumina) we used in the majority of our WES. Other genes with less well-interrogated regions include *VKORC1* (77.6%), *G6PD* (90.3%), and *SLCO1B1* (93.1%); all the other genes had greater than 95% of their coding regions well interrogated (Supplemental Table 4, Supplemental Table 5).

Whole genome sequencing (WGS)

For WGS, we tested concordance for the 68 patients who also had DMET data available (Figure 1). The median genomewide coverage was 33× (range: 23× to 53×). We observed call rates >98% in 115 of 124 CPIC important coding SNVs and indels (Figure 2B, Supplemental Table 2). Among those 115 variants, 40 were polymorphic in the 68 patients examined. All the allele frequencies of the variants in 44 patients of European ancestry were consistent with those reported in ExAC (Supplemental Figure 3B).

The nine CPIC important SNV/indels with low call rates by WGS included four *CYP2D6* variants and five *G6PD* variants. *G6PD* is located on chromosome X and males have only one copy of the gene. All of the no-call genotypes were from males who had lower coverage (average read depth of *G6PD*: 13.5× in males vs. 28× in females). Based on clinical qPCR, *CYP2D6* also had copy number variation: one of the 68 patients had a homozygous *CYP2D6* deletion (CN = 0) and four patients had only one copy of *CYP2D6* (CN = 1). Among the remaining 63 patients with two or more copies of *CYP2D6* gene, only two *CYP2D6* variants, *CYP2D6*11* (883G>C) and *CYP2D6*14* (G169R), had call rates <98%.

There were 52 SNVs and 14 indel genotypes that were interrogated with >98% call rate by both WGS and DMET. Among 4,479 genotypes compared (among a total of 68 patients), only four genotypes were discordant: one *TPMT*3B* (A154T, rs1800460) genotype, one *CYP2D6*20*, and two *CYP2D6*40* genotypes (Table 2). We performed orthogonal PCR-RFLP based assays for C>T at rs1800460 and T>C at rs1142345,[37] and the results indicated heterozygosity for both single nucleotide polymorphisms (SNPs), consistent with the *TPMT*3A* result obtained with WES, but not with the DMET result. The minor allele fraction for the *CYP2D6*20* variant in the WGS patient was 12.5%, and it was the only heterozygous call generated by WGS whose MAFrac was <20% across all variants (Supplemental Figure 4B).

We inferred *HLA-B* allele types from WGS based on the Optitype program at 4-digit resolution for all 68 patients. There was no clinical *HLA* typing available for comparison for these patients assessed by WGS. It has been shown that the accuracy of *HLA* class I typing using Optitype is over 98% at 4-digit resolution based on 1,000 genome samples.[38] In the 16 patients with both WGS and WES data, we observed a concordance of 91% at 4-digit resolution (29 out of 32 haplotypes, CI: 74.9–98.0%) and 100% (CI: 89.1–100%) concordance at the 2-digit resolution (Supplemental Figure 5B). The concordant haplotypes included one patient with *HLA-B*5801* and one patient with *HLA-B*5701*, both of which are clinically actionable.[18, 19, 21, 25]

We inferred *CYP2D6* copy number based on the average read depth in the *CYP2D6* region normalized to the whole genome coverage. There were 60 patients with *CYP2D6* copy number determined by clinical qPCR using three probes. Comparing the WGS inferred *CYP2D6* copy number with clinical qPCR determined *CYP2D6* copy number, we observed concordant results in 57 of 60 patients (95%, CI: 86.1–98.9%) (Figure 4B). As in the case of the WES data, one could theoretically infer which allele was gained for patients with *CYP2D6* duplications, based on the read count ratio. However, due to the lower read depth by WGS, there was larger variation in the read count ratio, precluding estimation of the duplicated allele (data not shown).

For *UGT1A1*28*, there were only six patients available with genotypes based on locus-specific PCR, and the concordance with the WGS inferred genotype was 100% (CI: 54.1–100%). We also compared the *UGT1A1* genotypes among the 16 patients with both WES and WGS genotypes and again observed 100% concordance (CI: 79.4–100%).

There were only 16 patients that had both WES and WGS data available. We compared the genotype concordance between WES and WGS at 107 of 127 CPIC important variants (84.2%, CI: 76.7–90.1%) with passing call rates >98%. Of these 107 loci, no discordant genotyping calls were observed among the 16 patients.

The median read depth of the coding region for the 13 important CPIC genes by WGS was 34×, similar to the genomewide average read depth (Supplemental Table 4, Supplemental Figure 8). The genes with less well-interrogated regions by WGS, as defined by the percentage of the gene with read depth <10×, included *G6PD* (41.5%), *VKORC1* (64.8%), *CYP2D6* (75.2%), and *IFNL3* (92%); all the other genes had greater than 95% of their coding regions well interrogated (Supplemental Table 4, Supplemental Table 5).

Observed coding variants not included among CPIC important variants

In 636 patients with WES, we observed a total of 153 coding variants that are not included among CPIC important variants, including four stop gain variants and five frameshift indels (Table 3). All of these variants are rare except for rs200579169 (*CYP3A5*), which was observed 12 times in the cohort of 636 patients, and all are reported in the ExAC database except for the *SLCO1B1* chr12:21391951 CATCA/− frameshift deletion. Among the 144 missense variants, 12 were not observed in either the ExAC database or dbSNP (Supplemental Table 6A).

In 68 patients with WGS, we observed a total of 66 missense variants and two frameshift variants that were not included in CPIC important variants. The two frameshift variants were observed in *CYP2C9* and *CYP3A5* (rs547253411) (Table 3B). Among the 66 missense variants, six were not observed in ExAC database or in dbSNP (Supplemental Table 6B). These novel variants were not confirmed by any additional genotyping method, and additional clinical or laboratory data would be needed to establish their importance.

DISCUSSION

With the advance of genome sequencing technology, more whole genome sequencing and whole exome sequencing data are being generated in both research and in clinical settings. In the clinic, genome sequencing is being used primarily to assist in diagnosis of inherited diseases or to assess tumor genomic variation,[10–12] in which case assessments of pharmacogenetic variation might be considered as secondary findings that are worthy of return to individual patients. Given that there are a tractable number of actionable pharmacogenetic variants,[29, 39] and the inherited pharmacogenetic variants may be of use throughout a patient's lifetime, the question of whether genome sequencing-generated data can serve "double duty" as clinically actionable pharmacogenomic data is relevant. Herein, we describe for the first time the performance of clinically valid pharmacogenes assessed by arrays and WES/WGS data generated in a research lab. Comparing clinical genotyping with WGS/WES data from the same patients, we only observed discordant genotyping calls in a few samples at select loci (Table 2). These discordant calls would need to be resolved using orthogonal genotyping methods. In some of the conflicting genotypes, the unusually low minor allele fraction by genome sequencing suggested the genotyping calls were suspect despite the fact the calls passed GATK's default quality control. In addition, manually inspecting genome sequencing reads can also help correct suspect genotypes. For example, the reads carrying *CYP2D6*20* variant allele by WGS are likely due to paralogous mapping and genotyping calls could be corrected accordingly (Supplemental Figure 9).

Compared with genome sequencing technology as it exists today, the current DMET array provides an affordable method for most CPIC genes and variants. On the other hand, it does not include some of the important CPIC genes, and within some CPIC genes it does not include some of the important variants, e.g., *G6PD*. Also, DMET generates genotyping calls based on clustering of reference genotypes. Due to the limited number of reference samples with variant genotypes, DMET may be less reliable for some rare variants.

One major shortcoming of WES is that several very important noncoding regions are not interrogated, including *VKORC1* promoter variants and *CYP3A5*5*, although we found that nontargeted WES/WGS reads yielded accurate genotypes for the promoter variant *UGT1A1*28*, even though it was not in the targeted capture region. To improve next-generation coverage of pharmacogenes, efforts such as PGRNseq,[40] include a customized sequencing capture method to interrogate important noncoding regions associated with pharmacogenes.

Compared with WES, WGS provides increased coverage of all SNV/indel variants. However, in some situations that involve genes with copy number variation, such as *CYP2D6* and *G6PD*, standard analytic pipelines of WGS generated lower call rates due to gene deletions. It should be noted that the WGS and WES strategies we used herein were not optimized in a clinical laboratory, nor were they modified to target known pharmacogenes. It is possible that pipelines could be customized to take into consideration copy number status, and to increase coverage of challenging loci, and such loci (e.g., *CYP2D6, G6PD*) were identified in our analysis.

Several CPIC important genes are known to have complex genomic architecture. Some exhibit high homology within family members and also often have pseudogenes that must be distinguished for accurate genotyping (e.g., the *CYP2D6* and *CYP3A5* loci).[41, 42] Here we showed that both WES and WGS provided excellent variant calling for *CYP2C9, CYP2C19*, and *CYP3A5*. It is especially encouraging that the current state of the art algorithms, e.g., XHMM, could provide reasonably accurate copy number estimation for *CYP2D6*. However, standard default algorithms cannot infer *CYP2D6/CYP2D7* hybrids, and this will require specially designed algorithms that are specific to *CYP2D6* by incorporating knowledge from domain experts.[33, 43] Genome sequencing provides an additional advantage of clarifying which haplotype is amplified in patients with *CYP2D6* copy number gains, which can be useful because duplication of a functional vs. duplication of a nonfunctional allele can lead to clinically significant differences in phenotype assignments, and thus in prescribing decisions.[44] *HLA-B* typing via sequencing is a good example of applying this strategy. Building on the advances of analytic methods and improvement in *HLA* allele databases, *HLA-B* typing using WES or WGS should provide excellent results.

Compared with a customized array platform like DMET, genome sequencing provided interrogation of more variants in the CPIC important genes. This may be a real advantage for genes such as *G6PD* or *DPYD*, which have a large number of important rare variants. In our study population, we did not observe many novel coding variants or loss-of-function variants that have not already been reported in public databases such as ExAC (Supplemental Table 7). We observed many missense variants in our cohorts, each with a low minor allele frequency; however, their functional consequences are not clear. Systematically analyzing these rare variants using functional assays will be challenging but needed to elucidate their potential importance.[45]

In conclusion, WGS provides reliable interrogation of most clinically important pharmacogenetic variants. While WES produced similar results for coding variants, functionally significant noncoding variants are missed with current exome capture assays. With relatively modest changes to data processing and by including customized capture probes targeting these important noncoding regions, use of genome sequencing data could provide a cost-effective way to advance the clinical implementation of pharmacogenomics in the clinic.[40] In order for genome sequencing to be used in the clinical setting, additional evaluation on its reproducibility, specificity, and sensitivity is needed.

METHODS

Patients and samples

Patients in the study include St. Jude Children's Research Hospital patients enrolled in the Pediatric Cancer Genome Project (Whole Genome sequencing, $n = 68$), PGEN5 and PG4KDS protocols (Affymetrix DMET Plus array, $n = 2656$), and Total XV/XVI studies (whole exome sequencing, $n = 636$). The study was approved by the Institutional Review Board at St. Jude Children's Research Hospital. Informed consent from the parents or guardians and assent from the patients as appropriate were obtained according to Institutional Review Board guidelines for treatment and for genomic research.

Germline DNA was extracted from normal peripheral blood leukocytes. Patient genetic ancestry was inferred using principal component analysis (PCA) applied to corresponding genotyping platforms.

Clinical genotyping

All clinical testing was performed in a CLIA-compliant laboratory.[3, 4]

Affymetrix Drug Metabolizing Enzymes and Transporters (DMET) Plus array version1 (Affymetrix, Santa Clara, CA) assays were performed at the Medical College of Wisconsin.[4, 6]

CYP2D6 copy number was determined using a supplemental assay based on quantitative PCR as previously described.[4] This is a clinical test run in a clinical, CLIA-certified laboratory. The test has been validated as required by CLIA for laboratory-developed tests (42 CFR 493.1253 – Standard: Establishment and verification of performance specifications). This assay utilized three probes; when concordant in copy number estimates, they indicate likely *CYP2D6* duplication or deletion, and when discordant (e.g., 2/3/3N, 2/3/2N, etc), signal a possible *CYP2D6/2D7* hybrid. Such possible hybrids were not assigned a final diplotype for clinical purposes, as further testing is needed to confirm complex *CYP2D6* hybrid structures.[41]

Whole exome sequencing and genotyping

Exome capture and enrichment were performed using TruSeq Exome Enrichment and Nextera Exome Enrichment kits according to standard protocols. Paired-end whole-exome sequencing was performed using Illumina HiSeq2500 instruments. The sequence reads were aligned to the human reference genome hg19 using the BWA. All whole exome sequencing assays were performed in the Hartwell Center at St. Jude Children's Research Hospital.

Genotypes were generated using Genome Analysis Toolkit v. 3.2 according to the best practice research laboratory workflow (https://www.broadinstitute.org/gatk/guide/best-practices). Genotypes with quality scores lower than 20 were considered as no calls.

For the purposes of this analysis, additional workflows were applied to the WES data: *CYP2D6* copy number was inferred using XHMM based on WES data,[31] and *HLA-B* alleles were inferred using PolySolver software.[30] Clinical *HLA*-typing was performed using clinical molecular methods as described.[6]

Whole genome sequencing and genotyping

Whole genome data were generated in research laboratories as part of the St. Jude–Washington University Pediatric Cancer Genome Project as previously described.[14] Variant calls were generated using GATK as described above.

Germline CNVs for WGS were identified using "CONSERTING."[46] The segmentation files were then mapped to genes with the R package "cghMCR," which provides the copy number status of *CYP2D6*. The average read depth in the *CYP2D6* gene region was normalized against the average whole genome coverage for each patient. For purposes of

this analysis, the additional workflow applied to the WGS data was that *HLA-B* alleles were inferred using OptiType.[38]

Additional genotyping

UGT1A1 genotyping was performed based on a locus-specific PCR assay as a research test as described previously.[35, 36]

Additional *TPMT*3B* genotyping was performed based on PCR-restriction fragment length polymorphism (RFLP)-based analysis.[39]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Evans WE, et al. Preponderance of thiopurine S-methyltransferase deficiency and heterozygosity among patients intolerant to mercaptopurine or azathioprine. J. Clin. Oncol. 2001; 19:2293–2301. [PubMed: 11304783]

2. Hicks JK, et al. A clinician-driven automated system for integration of pharmacogenetic interpretations into an electronic medical record. Clin. Pharmacol. Ther. 2012; 92:563–566. [PubMed: 22990750]

3. Bell GC, et al. Development and use of active clinical decision support for preemptive pharmacogenomics. J. Am. Med. Inform. Assoc. JAMIA. 2014; 21:e93–99. [PubMed: 23978487]

4. Hoffman JM, et al. PG4KDS: a model for the clinical implementation of pre-emptive pharmacogenetics. Am. J. Med. Genet. C Semin. Med. Genet. 2014; 166C:45–55. [PubMed: 24619595]

5. Dunnenberger HM, et al. Preemptive clinical pharmacogenetics implementation: current programs in five US medical centers. Annu. Rev. Pharmacol. Toxicol. 2015; 55:89–106. [PubMed: 25292429]

6. Fernandez CA, et al. Concordance of DMET plus genotyping results with those of orthogonal genotyping methods. Clin. Pharmacol. Ther. 2012; 92:360–365. [PubMed: 22871999]

7. Jackson JN, et al. A comparison of DMET Plus microarray and genome-wide technologies by assessing population substructure. Pharmacogenet. Genomics. 2016 e-pub ahead of print.

8. KA, W. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). 2016. <www.genome.gov/sequencingcosts>

9. Pirooznia M, et al. Validation and assessment of variant calling pipelines for next-generation sequencing. Hum. Genomics. 2014; 8:14. [PubMed: 25078893]

10. Kingsmore SF, Saunders CJ. Deep sequencing of patient genomes for disease diagnosis: when will it become routine? Sci. Transl. Med. 2011; 3:87ps23.

11. Tabor HK, et al. Pathogenic variants for Mendelian and complex traits in exomes of 6,517 European and African Americans: implications for the return of incidental results. Am. J. Hum. Genet. 2014; 95:183–193. [PubMed: 25087612]

12. Biesecker LG, Green RC. Diagnostic clinical genome and exome sequencing. N. Engl. J. Med. 2014; 371:1170.

13. Downing JR, et al. The Pediatric Cancer Genome Project. Nat. Genet. 2012; 44:619–622. [PubMed: 22641210]

14. Zhang J, et al. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. Nature. 2012; 481:157–163. [PubMed: 22237106]

15. Relling MV, et al. Clinical pharmacogenetics implementation consortium guidelines for thiopurine methyltransferase genotype and thiopurine dosing: 2013 update. Clin. Pharmacol. Ther. 2013; 93:324–325. [PubMed: 23422873]

16. Relling MV, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines for rasburicase therapy in the context of G6PD deficiency genotype. Clin. Pharmacol. Ther. 2014; 96:169–174. [PubMed: 24787449]

17. Wilke RA, et al. The clinical pharmacogenomics implementation consortium: CPIC guideline for SLCO1B1 and simvastatin-induced myopathy. Clin. Pharmacol. Ther. 2012; 92:112–117. [PubMed: 22617227]

18. Leckband SG, et al. Clinical Pharmacogenetics Implementation Consortium guidelines for HLA-B genotype and carbamazepine dosing. Clin. Pharmacol. Ther. 2013; 94:324–328. [PubMed: 23695185]

19. Clancy JP, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines for ivacaftor therapy in the context of CFTR genotype. Clin. Pharmacol. Ther. 2014; 95:592–597. [PubMed: 24598717]

20. Muir AJ, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines for IFNL3 (IL28B) genotype and PEG interferon-alpha-based regimens. Clin. Pharmacol. Ther. 2014; 95:141–146. [PubMed: 24096968]

21. Martin MA, et al. Clinical Pharmacogenetics Implementation Consortium guidelines for HLA-B genotype and abacavir dosing: 2014 update. Clin. Pharmacol. Ther. 2014; 95:499–500. [PubMed: 24561393]

22. Gammal RS, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for UGT1A1 and Atazanavir Prescribing. Clin. Pharmacol. Ther. 2015 e-pub ahead of print.

23. Birdwell KA, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guidelines for CYP3A5 Genotype and Tacrolimus Dosing. Clin. Pharmacol. Ther. 2015; 98:19–24. [PubMed: 25801146]

24. Hicks JK, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for CYP2D6 and CYP2C19 Genotypes and Dosing of Selective Serotonin Reuptake Inhibitors. Clin. Pharmacol. Ther. 2015; 98:127–134. [PubMed: 25974703]

25. Saito Y, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines for human leukocyte antigen B (HLA-B) genotype and allopurinol dosing: 2015 update. Clin. Pharmacol. Ther. 2015 e-pub ahead of print.

26. Johnson JA, et al. Clinical Pharmacogenetics Implementation Consortium Guidelines for CYP2C9 and VKORC1 genotypes and warfarin dosing. Clin. Pharmacol. Ther. 2011; 90:625–629. [PubMed: 21900891]

27. Scott SA, et al. Clinical Pharmacogenetics Implementation Consortium guidelines for CYP2C19 genotype and clopidogrel therapy: 2013 update. Clin. Pharmacol. Ther. 2013; 94:317–323. [PubMed: 23698643]

28. Caudle KE, et al. Clinical Pharmacogenetics Implementation Consortium guidelines for dihydropyrimidine dehydrogenase genotype and fluoropyrimidine dosing. Clin. Pharmacol. Ther. 2013; 94:640–645. [PubMed: 23988873]

29. Rasmussen-Torvik LJ, et al. Design and anticipated outcomes of the eMERGE-PGx project: a multicenter pilot for preemptive pharmacogenomics in electronic health record systems. Clin. Pharmacol. Ther. 2014; 96:482–489. [PubMed: 24960519]

30. Shukla SA, et al. Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. Nat. Biotechnol. 2015; 33:1152–1158. [PubMed: 26372948]

31. Fromer M, et al. Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. Am. J. Hum. Genet. 2012; 91:597–607. [PubMed: 23040492]

32. Beoris M, Amos, Wilson J, Garces JA, Lukowiak AA. CYP2D6 copy number distribution in the US population. Pharmacogenet. Genomics. 2016; 26:96–99. [PubMed: 26551314]

33. Twist GP, et al. Constellation: a tool for rapid, automated phenotype assignment of a highly polymorphic pharmacogene, CYP2D6, from whole-genome sequences. Npj Genomic Med. 2016; 1:15007.

34. Fang H, et al. Establishment of CYP2D6 reference samples by multiple validated genotyping platforms. Pharmacogenomics J. 2014; 14:564–572. [PubMed: 24980783]

35. Te HS, et al. Donor liver uridine diphosphate (UDP)-glucuronosyltransferase-1A1 deficiency causing Gilbert's syndrome in liver transplant recipients. Transplantation. 2000; 69:1882–1886. [PubMed: 10830226]

36. Kishi S, et al. Effects of prednisone and genetic polymorphisms on etoposide disposition in children with acute lymphoblastic leukemia. Blood. 2004; 103:67–72. [PubMed: 12969965]

37. Ameyaw MM, Collie-Duguid ES, Powrie RH, Ofori-Adjei D, McLeod HL. Thiopurine methyltransferase alleles in British and Ghanaian populations. Hum. Mol. Genet. 1999; 8:367–370. [PubMed: 9931345]

38. Szolek A, et al. OptiType: precision HLA typing from next-generation sequencing data. Bioinformatics. 2014; 30:3310–3316. [PubMed: 25143287]

39. Relling MV, Evans WE. Pharmacogenomics in the clinic. Nature. 2015; 526:343–350. [PubMed: 26469045]

40. Gordon AS, et al. PGRNseq: a targeted capture sequencing panel for pharmacogenetic research and implementation. Pharmacogenet. Genomics. 2016 e-pub ahead of print.

41. Gaedigk A, et al. Identification of novel CYP2D7-2D6 hybrids: nonfunctional and functional variants. Front. Pharmacol. 2010; 1:121. [PubMed: 21833166]

42. Kuehl P, et al. Sequence diversity in CYP3A promoters and characterization of the genetic basis of polymorphic CYP3A5 expression. Nat. Genet. 2001; 27:383–391. [PubMed: 11279519]

43. Numanagic I, et al. Cypiripi: exact genotyping of CYP2D6 using high-throughput sequencing data. Bioinformatics. 2015; 31:i27–34. [PubMed: 26072492]

44. Ramamoorthy A, Skaar TC. Gene copy number variations: it is important to determine which allele is affected. Pharmacogenomics. 2011; 12:299–301. [PubMed: 21449666]

45. Ormond KE, et al. Challenges in the clinical application of whole-genome sequencing. Lancet. 2010; 375:1749–1751. [PubMed: 20434765]

46. Chen X, et al. CONSERTING: integrating copy-number analysis with structural-variation detection. Nat. Methods. 2015; 12:527–530. [PubMed: 25938371]

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

☑ Pharmacogenetics is being widely adopted in clinical practice to provide personalized drug dosing. Most pharmacogenetics tests performed today are based on pharmacogene-directed assays.

WHAT QUESTION DID THIS STUDY ADDRESS?

☑ This study addressed the question of whether genome sequencing could be effectively used to provide good interrogation of pharmacogenes.

WHAT THIS STUDY ADDS TO OUR KNOWLEDGE

☑ By comparing clinical genotyping (here, primarily Affymetrix DMET array) vs. genome sequencing for their coverage of all CPIC genes, this study provides benchmarks for genome sequencing in the interrogation of pharmacogenes.

HOW THIS MIGHT CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE

☑ The results of this study provided support for the adoption of genome sequencing in clinical practice of pharmacogenetics and will encourage more rigorous evaluation of genome sequencing in clinical labs to evaluate its sensitivity and specificity.

Figure 1.
Number of patients interrogated using different platforms.

Figure 2.
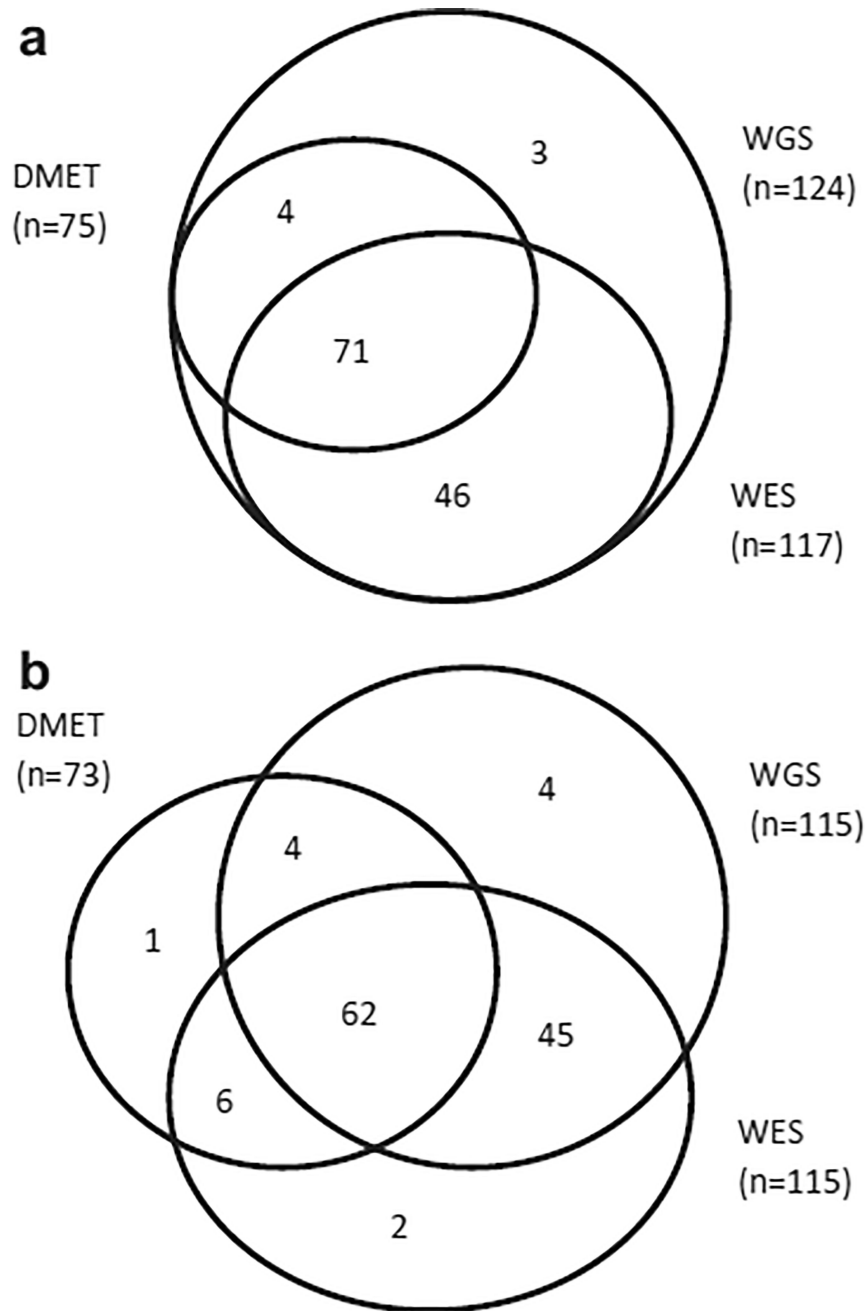Numbers of CPIC clinically important variants interrogated by DMET, WES, and WGS. (**a**)
All 124 interrogated SNV/indels are included; excludes *CYP2D6* copy number variation,
CYP2D6/2D7 hybrids, and *HLA-B* ($n = 3$). (**b**) SNV/indels ($n = 124$ variants) with call rate
greater than 98% in any platform are included; excludes *CYP2D6* copy number variation,
CYP2D6/2D7 hybrids, *or HLA-B* ($n = 3$ variants).
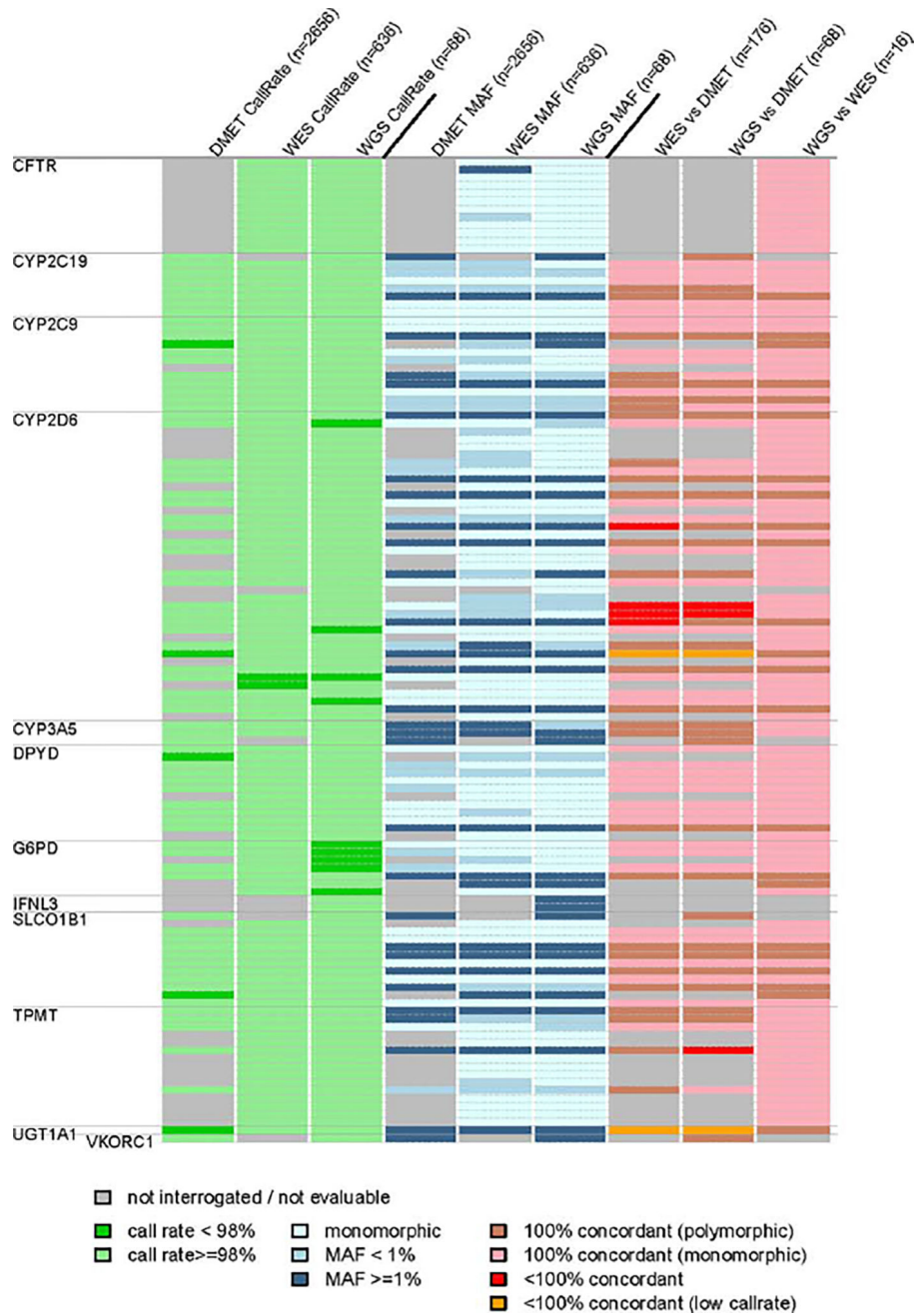
Figure 3.
Genotyping call rates, minor allele frequencies, and concordance among platforms for CPIC important SNVs (*n* = 124). Each row is a genomic variant; the order of the variants (top to bottom) is the same as in Supplemental Table 2.
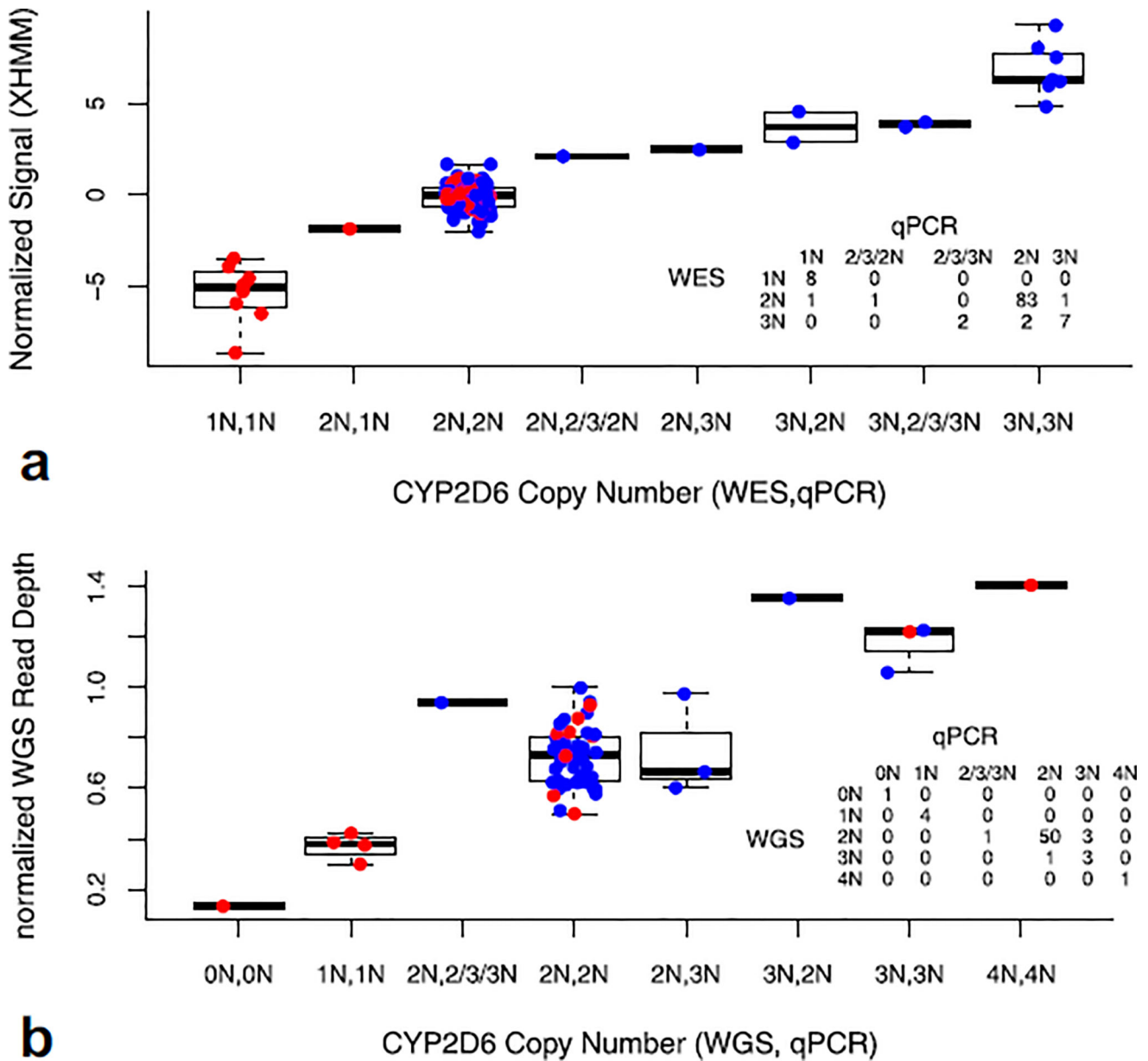
Figure 4.
Comparison of *CYP2D6* copy number determined by **(a)** WES or **(b)** WGS vs. clinical qPCR. Plotted are normalized signals generated by XHMM for WES or normalized read depth in *CYP2D6* gene region for WGS. Patients were grouped according to their combined copy number estimates determined by WES or WGS and qPCR (x-axis). For example, in **(a)**, "3N,2N" indicates the patient was genotyped as "CN = 3" by WES but "CN = 2" by qPCR. Discordance in the three probes as ascertained by qPCR, e.g., 2/3/3N vs. 2/3/2N, etc., indicates possible CYP2D6/2D7 hybrids. Red dots in this figure represent patients who are homozygous for the reference allele or homozygous for the same variant alleles, having identical reads in all positions interrogated by WES or WGS. This would be consistent with the situation in which a person either has a single copy of the gene or has multiple copies of *CYP2D6*, each with identical sequence. Blue dots indicate patients with any heterozygous

call at any interrogated *CYP2D6* SNV/indel loci. The table insets depict the concordance between clinical qPCR and WES/WGS.

Table 1

CPIC important genes included a total of 127 CPIC important variants

Gene	Number of CPIC important variants				Affymetrix DMET and add-on assays	Whole exome sequencing	Whole genome sequencing
	SNV (exonic)	Indel (exonic)	Other	Total			
CFTR	10 (1N0)	2 (2)		12	Not interrogated	12	12
CYP2C19	8 (7)	0		8	8	7	8
CYP2C9	10 (10)	2 (2)		12	10	12	12
CYP2D6	26 (24)	13 (13)	2 structural variations	41	23 and two structural variations	36 and CYP2D6 copy number (missing CYP2D6/2D7 hybrid)	35 and CYP2D6 copy number (missing CYP2D6/2D7 hybrid, lower call rate due to copy number variation)
CYP3A5	2 (1)	1 (1)		3	3	2 (missing important variants)	3
DPYD	10 (10)	2 (2)		12	9 (missing important variants)	12	12
G6PD	7 (7)	0		7	4 (missing important variants)	7	2 (lowercall rate due to copy number variation)
HLA-B	0	0	1 haplotype	1	Not interrogated	1	1
IFNL3	2 (0)	0		2	Not interrogated	Missing important variants	2
SLCO1B1	12 (11)	0		12	10	11	12
TPMT	15 (15)	0		15	5	15	15
UGT1A1	0	1 (0)		1	Low call rate	1	1
VKORC1	1 (0)	0		1	1	Missing important variants	1
Total	103 (95)	21 (20)	3	127			

Genotyping performance by DMET, WES, and WGS is summarized as the number of variants interrogated with call rate >98%.

Table 2

Observed discordant genotypes comparing DMET CPIC important SNV/indel variants to A) WES (*n* = 176 patients), or to B) WGS (*n* = 68 patients)

A)

Gene	Allele	dbSNP	DMET call	WES call	Reference allele count	Alternative allele count	Minor allele fraction	Comment
CYP2D6	*20 (1973insG)	rs72549354	T/T	T/TC	364	57	13.5%	WES low minor allele fraction
CYP2D6	*20 (1973insG)	rs72549354	T/T	T/TC	278	51	15.5%	WES low minor allele fraction
CYP2D6	*40 (1863_1864ins)	rs72549356	–/18bps	–/–	138	0	0.0%	Reason for discrepancy unclear
CYP2D6	*40 (1863_1864ins)	rs72549356	–/18bps	–/–	292	0	0.0%	Reason for discrepancy unclear
CYP2D6	*4 (1846G>A)	rs3892097	T/T	C/T	10	83	10.8%	WES low minor allele fraction
CYP2D6	*4 (1846G>A)	rs3892097	T/T	C/T	15	133	10.1%	WES low minor allele fraction
CYP2D6	*2 (R296C)	rs16947	A/A	A/G	71	9	11.3%	WES low minor allele fraction

B)

Gene	Allele	dbSNP	DMET call	WES call	Reference allele count	Alternative allele count	Minor allele fraction	Comment
TPMT	*3B (A154T)	rs1800460	C/C	C/T	24	25	49.0%	WGS call agreed with orthogonal PCR-RFLP method for TPMT
CYP2D6	*20 (1973insG)	rs72549354	T/T	T/TC	42	6	12.5%	WGS low minor allele fraction
CYP2D6	*40 (1863_1864ins)	rs72549356	–/18bps	–/–	30	0	0.0%	Reason for discrepancy unclear
CYP2D6	*40 (1863_1864ins)	rs72549356	–/18bps	–/–	32	0	0.0%	Reason for discrepancy unclear

Table 3

Loss of function variants not included as CPIC important variants that were discovered by A) WES (n = 636 patients) and B) WGS (n = 68 patients) in our cohort and frequency of those variants reported in ExAC (http://exac.broadinstitute.org/ version 0.3)

A)

Platform	Gene	Chromosome position	Ref	Alt	Function	Amino acid change	dbSNP(138)	Observed (n = 636)	ExAC freq White	ExAC freq African
WES	CFTR	Chr7:117232266	-	A	frameshift insertion	p.T682fs	rs121908746	1	9.123e-05	0
WES	CFTR	Chr7:117227832	G	T	stopgain SNV	p.G542*	rs113993959	1	0.0003522	9.628e-05
WES	CYP2D6	Chr22: 42523532	-	CA	frameshift insertion	p.Q364fs	Not found	1	0	0.002234
WES	CYP3A5	Chr7:99247736	AC	A	frameshift deletion	p.V458fs	rs547253411	2	0.002	0.0002
WES	CYP3A5	Chr7:99273810	-	C	frameshift insertion	p.G31fs	rs200579169	12	0.01098	0.001353
WES	CYP3A5	Chr7:99247731	G	A	stopgain SNV	p.Q460*	rs149664815	2	0	0.00125
WES	DPYD	Chr1:98293683	G	A	stopgain SNV	p.R74*	rs189768576	1	0	0
WES	SLCO1B1	Chr12:21391951	CATCA	-	frameshift deletion	p.S635fs	Not Found	1	Notobserved	Notobserved
WES	SLCO1B1	Chr12:21375289	C	T	stopgain SNV	p.R580*	rs71581941	1	0.001384	0.0009684

B)

Platform	Gene	Chromosome position	Ref	Alt	Function	Amino acid change	dbSNP(138)	Observed (n = 68)	ExAC freq White	ExAC freq African
WGS	CYP3A5	chr7:99247736	AC	A	frameshift deletion	p.V458fs	rs547253411	1	0.002	0.0002
WGS	CYP2C9	chr10:96740948	CA	C	frameshift deletion	p.Q324fs	Not Found	1	1.0e-04	0