White Paper

DMET™ Plus allele translation reports: Summary of comprehensive drug disposition genotyping into commonly recognized allele names

Abstract

Background

The DMET™ Plus Solution is a comprehensive drug metabolism profiling assay capable of genotyping common and complex functional variants in more than 225 ADME-related genes simultaneously. Interpretation of the genetic profiles generated by the DMET Plus Solution is greatly aided by summarizing results and translating genotyping data into common pharmacogenetic allele names.

Purpose

We describe a report-generation tool in DMET™ Console Software that generates fully annotated marker reports. The reports include commonly recognized, haplotype-based allele calls commonly cited in Medline reference studies. We describe how a look-up table was created that predicts phenotypes. We describe workflow improvements to permit users to replace missing data. Finally, we describe differences in DMET genotype profiles among ethnic groups.

Results summary

The DMET™ Plus allele translation software produces four reports:

- A tabular comprehensive genotyping report containing pharmacogenomic reference data on all probes
- A variant summary in similar format, where defining SNPs are systematically retrieved
- A phenotype report on a subset of higher-visibility genes
- An uncalled report detailing the missing genotypes within translated genes

The uncalled report may be used as a template to feed externally derived genotype data back into the analysis pipeline to supplement study results for samples with "no call" or "possible rare allele" calls (probe signals that fall out of the range of the training data). Affymetrix has processed DNA from the HapMap extended diversity panel to provide population-based allele frequencies across the full pharmacogenomic assay panel, which confirmed many known population-based allele frequency differences. Novel associations found in this survey are summarized at the end of this white paper.

Conclusions

Conversion of genotype results in the DMET Core genes to standardized names, e.g., the star allele nomenclature of the CYP450 genes, facilitates interpretation of DMET Plus genotyping results. Among the ADME genes represented on the DMET Plus platform, VKORC1, CYP3A4, and PPARG show the greatest allele frequency differences worldwide.

Introduction

A recent survey by the FDA found that approximately 25 percent of all prescriptions filled by a major US healthcare provider contained pharmacogenomic labeling recommendations.[1] Variation in genes responsible for the metabolism of these drugs is very common. Genetic profiling of the most common alterations in drug metabolism genes, a critical component of personalized medicine, should improve efficacy and safety of drug therapy.[2,3] Adverse drug reactions occur most frequently when a drug has a narrow therapeutic index. The blood thinner warfarin (Coumadin) and chemotherapeutic agents such as Irinotecan are two examples. Discovery and application of biomarkers resulting from pharmacogenetic associations with these two drugs will be used to illustrate the utility of the DMET™ Plus platform in pharmacology clinical research.

Drug discovery and development timelines are long and costly, and the process remains a high-risk and high-stake problem for the pharmaceutical industry. Indeed, the attrition of drug candidates in preclinical and development stages is a major problem in drug design: approximately 30 percent of candidates fail.[4] This attrition is primarily due to poor pharmacokinetics and toxicity, two risks that can be strongly influenced by genetically determined variation. Thus, pharmaceutical companies are actively reevaluating strategies used during drug discovery and development to prevent these failures. The process of dissecting environmental factors, physiological variables, and genetic characteristics (pharmacogenetics) is complicated by the large numbers of drug-metabolizing enzymes that contribute to the processing of most drugs. Genetic variants in these metabolic pathways and environmental factors, such as concurrent medication, age, and gender, alter the relationship between the absolute dose and individual blood concentration–time profiles.

Increasingly, a focus in drug development is to design targets that circumvent polymorphisms in drug-metabolizing genes such as CYP2D6, CYP2C19, or UGT1A1, and in drug transporter genes such as human ABC transporters ABCB1 and ABCG2.[5] The successful prediction of anticancer drug disposition and toxicity with pharmacogenetics of ABCG2 lays the foundation for similar efforts in other therapeutic arenas. High-activity alleles in these genes can lead to AGCG2-associated multidrug resistance to cancer drugs,[6] whereas low-activity alleles may increase bioavailability of chemotherapeutic agents such as irinotecan[7] and topotecan.[8] Individualizing drug dose to the ABCG2 genotype has been proposed as a way to minimize adverse side effects or toxicity of these drugs.[9]

To improve the therapeutic index of anticancer drugs, a variety of strategies, including pharmacogenetic genotyping, are now being evaluated in cancer patients.[10] To date, cancer pharmacogenetic profiling has mainly focused on variants in genes encoding the drug-metabolizing enzymes thiopurine S-methyltransferase (TPMT),[11] Dihydropyrimidine dehydrogenase (DYPD),[12] members of the cytochrome P450 family, including the CYP2B, -2C, -2D, and -3A subfamily genes,[13] and members of the UDP glucuronosyltransferase family (primarily UGT1A1 and UGT1A9).[14] The ATP-binding cassette transporters ABCB1 (P-glycoprotein) and ABCG2 (breast cancer resistance protein) are additional genes of interest in clinical pharmacology. All of these genes are among the ADME Core genes in the DMET Plus Product. Several investigator groups are recommending the implementation of more widespread genotyping of ADME gene variants.[9,10] Dangerous drug interactions[15-18] or, more recently, nutritional supplement-related hepatic toxicity and liver failure have also been traced to functional variants in the same group of genes.[19]

By analyzing a comprehensive set of targeted mutations, the DMET™ Plus product, an array-based assay, may reveal significant aspects of patient-to-patient pharmacokinetic variability in drug response.[20] One such example is the recent identification using the DMET Plus product of a novel genetic variant (CYP4F2*3, defined by rs2108622; V433M) that alters response to the blood thinner warfarin.[21] Interestingly, this common nonfunctional CYP4F2 allele was entirely missed in a genome-wide scan using a linkage analysis approach with tagged single nucleotide polymorphisms (SNPs).[22] Nevertheless, the CYP4F2*3 association was replicated in three independent populations when the DMET Plus product was used as the genotyping assay platform.[21] Here, we describe a new software tool to help facilitate interpretation of genetic determinants involved in warfarin dosing or selected oncology drugs as examples of clinical application areas.

Highly multiplexed assay designed for pharmacogenetics by pharmacogenetic professionals

Affymetrix scientists and pharmaceutical specialists participated in the PharmaADME consortium (www.pharmaadme.org/joomla/), an initiative of academic, pharmaceutical industry, and genomic technology representatives. The primary objective of this collaborative group was to create a consensus list of the known and likely functional variations present in key genes involved in the absorption, distribution, metabolism, and excretion (ADME) of commonly prescribed medication. More than 9,000 SNPs and other complex mutations (triallelic markers, small insertion/deletion mutations, gene conversions, and/or whole gene deletion alleles) were ranked for clinical research utility. The goal in further refining the DMET™ Panel from the prioritized marker set was to focus on mutations associated with known metabolic differences in this collection and to enrich for coverage in drug transporter genes and important transcription regulators (PPARD, PPARG, NR1I2, etc.). Care was taken to remove redundant or misannotated markers from the recommended panel.

Additional content was also added at the recommendation of the pharmaceutical industry and other Affymetrix users beyond the original putative biomarkers from the PharmaADME consortium. The criteria used in updating DMET Plus content was to specifically select markers with the highest likelihood of adding biological utility to the multiplex assay panel. Among the additional genes are novel drug transporters, transcription regulators, and genes known to induce other ADME genes. The design criteria, product benefit, and example markers in each category are listed in Figure 1. Thus, the DMET Plus product was designed by domain experts specifically for use in clinical pharmacology research studies, pharmaceutical product development, and clinical trials.

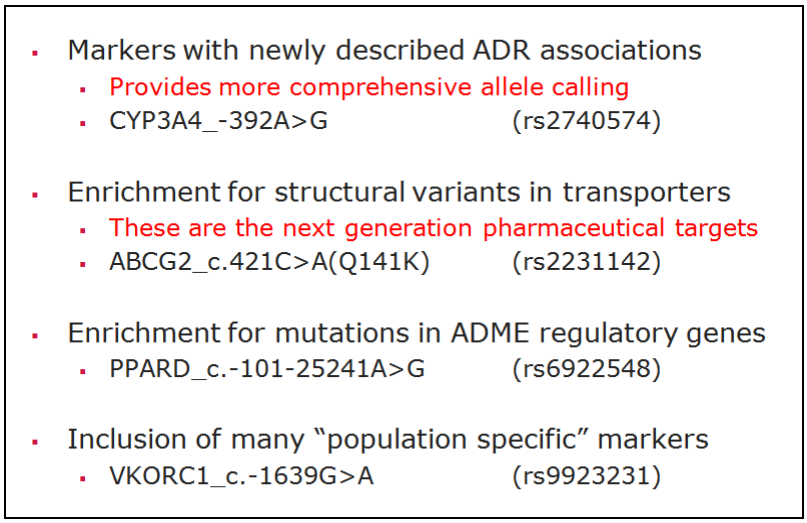


Figure 1: Content design guidelines used to enhance the DMET Plus Assay.

Standardized nomenclature for common variation in drug metabolism

The utility of DMET™ Plus analysis is further enhanced by the use of a standardized nomenclature to track important known clinical variants. One example is support for the star allele nomenclature, which supports genetic polymorphism annotation for the CYP450 genes.[23,24] As clinical pharmacogenetic testing becomes more widespread, this system for describing common functional and nonfunctional haplotypes becomes a mechanism to interpret genotype and predicted clinical phenotype.[23] DMET allele translation reporting provides a valuable mechanism for structured interpretation of these alleles across a large body of literature characterizing known genetic variants.

Methods applied in building the DMET Plus translation database

Affymetrix and collaborators curated data from reference databases and primary literature for a core set of genes represented in the DMET™ Plus Product to annotate mutations ranked by the PharmaADME consortium to be of primary importance in drug metabolism. The gene tables include primary literature reference citations, genomic locations, mRNA positions, and notation of coding changes that result when variants are detected at allele-defining probe sets. The translation reports are flexible to accommodate multiple allele naming conventions defined for the CYP450 (www.cypalleles.ki.se/) and other gene families as reported in PharmGKB (www.pharmgkb.org/) or in PubMed. This information is integrated with the genotyping calls made in DMET Console Software to help interpret whether important genetic variants have been identified. DMET Plus genotype data across more than 1,200 unique genomes, including 597 individuals from the extended HapMap population data, were run during product development verification and validation. The haplotype names and definitions in the translation table adhere as closely as possible to the literature definitions of the haplotypes. To confirm accuracy of the information in the DMET Plus translation library file, an independent DMET Scientific Advisory Consortium of seven external laboratories reviewed the content of these gene tables.

Domain experts review translation reporting

An abbreviated database containing the CYP2D6, VKORC1, NAT2, and CYP3A4 genes was generated to test user acceptance of the pharmacogenomic data reporting. These genes are among the most widely studied and contain the largest number of documented haplotypes; they

also contain representatives of complex genetic markers (trialleles, insertion/deletion mutations, and copy number variants), as well as SNP markers.

DMET Scientific Advisory Consortium: external review of biological annotations

To confirm accuracy of the information in the DMET™ Plus allele translation database, an independent DMET Scientific Advisory Consortium of seven external laboratories reviewed the content of key gene tables. Members of this advisory group included:

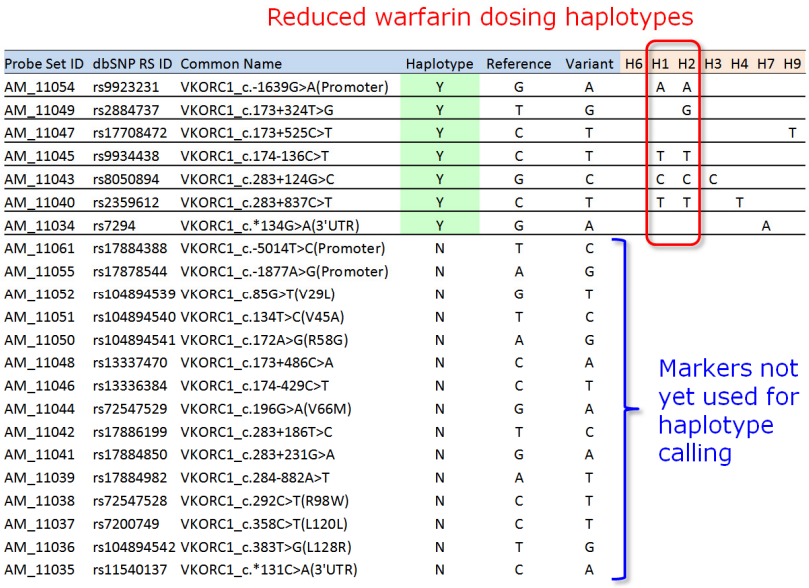
- Dr. William Douglas Figg, Senior Chairman, National Cancer Institute
- Dr. John Deeken, Georgetown University
- Dr. David Flockhart, University of Indiana
- Dr. Howard McLeod, University of North Carolina
- Dr. Alex Sparreboom, St. Jude Children's Hospital
- Dr. Teri Klein, PharmGKB/Stanford University

Additional contributions were made by:

- Dr. Tristan Sissung, National Cancer Institute
- Dr. Doug Price, National Cancer Institute
- Dr. Libusha Kelly, University of California, San Francisco

Additional markers increase haplotype accuracy and differentiation

Figure 2 shows an example of a gene table used in the translation process. Selected genes, such as VKORC1, show extensive genetic variation and many novel haplotypes. Multiple groups have recently reported population variations in VKORC1 haplotype structure and allele frequencies.[25-27] However, different naming conventions and partially overlapping sets of five to nine SNP markers were used in these studies. Generally, the most widely referenced study reporting VKORC1 haplotypes used nine SNPs across this gene describing five common haplotypes that fall into two groups: Group A (H1 and H2 haplotypes), with high frequency in Asian populations, and Group B (H7, H8, and H9 haplotypes), with high frequency in studied African American and Caucasian patient groups.[28] The A- and B-haplotype groups correlated with mRNA expression levels and with low- or high-dose requirements for effective control of clotting with warfarin. Finer granularity of the gene structure and the haplotype groups is observed when additional SNP markers are included across the gene. However, the additional markers will not be assigned to existing or new haplotypes until the scientific community chooses to define new or modify existing haplotype definitions.



Probe Set ID	dbSNP RS ID	Common Name	Haplotype	Reference	Variant	H6	H1	H2	H3	H4	H7	H9
AM_11054	rs9923231	VKORC1_c.-1639G>A(Promoter)	Y	G	A		A	A				
AM_11049	rs2884737	VKORC1_c.173+324T>G	Y	T	G			G				
AM_11047	rs17708472	VKORC1_c.173+525C>T	Y	C	T							T
AM_11045	rs9934438	VKORC1_c.174-136C>T	Y	C	T		T	T				
AM_11043	rs8050894	VKORC1_c.283+124G>C	Y	G	C		C	C	C			
AM_11040	rs2359612	VKORC1_c.283+837C>T	Y	C	T		T	T		T		
AM_11034	rs7294	VKORC1_c.*134G>A(3'UTR)	Y	G	A						A	
AM_11061	rs17884388	VKORC1_c.-5014T>C(Promoter)	N	T	C							
AM_11055	rs17878544	VKORC1_c.-1877A>G(Promoter)	N	A	G							
AM_11052	rs104894539	VKORC1_c.85G>T(V29L)	N	G	T							
AM_11051	rs104894540	VKORC1_c.134T>C(V45A)	N	T	C							
AM_11050	rs104894541	VKORC1_c.172A>G(R58G)	N	A	G							
AM_11048	rs13337470	VKORC1_c.173+486C>A	N	C	A							
AM_11046	rs13336384	VKORC1_c.174-429C>T	N	C	T							
AM_11044	rs72547529	VKORC1_c.196G>A(V66M)	N	G	A							
AM_11042	rs17886199	VKORC1_c.283+186T>C	N	T	C							
AM_11041	rs17884850	VKORC1_c.283+231G>A	N	G	A							
AM_11039	rs17884982	VKORC1_c.284-882A>T	N	A	T							
AM_11038	rs72547528	VKORC1_c.292C>T(R98W)	N	C	T							
AM_11037	rs7200749	VKORC1_c.358C>T(L120L)	N	C	T							
AM_11036	rs104894542	VKORC1_c.383T>G(L128R)	N	T	G							
AM_11035	rs11540137	VKORC1_c.*131C>A(3'UTR)	N	C	A							

Figure 2: The translation table for VKORC1, which defines the callable haplotypes. The DMET™ Plus Assay Panel includes 29 VKORC1 markers. Variation across HapMap population DNA was observed at 18 of these markers. However, only 7 were described by Rieder *et al.,*[28] who defined the haplotype names H1 through H9. Therefore, only 7 markers are used to call these haplotypes. The reference or variant status of all 29 markers is reported in addition to the haplotype calls.

The allele translation algorithm identifies all possible pairs of defined haplotypes (diplotypes) that are consistent with the pattern of marker-level genotypes. In compound heterozygous samples, more than one diplotype can be consistent with the genotypes. All possible diplotypes are reported, and they are not ranked by likelihood.

Genes with larger numbers of polymorphic sites, in particular multiple SNPs with high minor allele frequencies or many recent mutations, may have large numbers of haplotypes at very low frequency. Consequently, the gene tables may not fully describe all genetic diversity. The DMET™ Console Software translation engine therefore assumes that these undefined haplotypes may indeed exist as unknown alleles (reported as "UNK"). Diplotypes with UNK calls are reported as alternative results in a separate field of the translation reports. Classification codes, useful for sorting and interpreting the comprehensive or summary translation output, are also reported.

Report generation is integrated into DMET Console Software

Analysis of DMET™ Plus Array results is conducted in an integrated software package, DMET Console Software. The software provides a familiar user interface and staged workflow as implemented in other Affymetrix® Genotyping Software products. For example, the menu structure and tool bar contain common operations and icons (Figure 3) for defining project workspace, importing data, and signal processing (creation of DMET CHP files) followed by analysis of genotyping results. Equally importantly, DMET Console Software includes specialized analysis routines for processing non-SNP markers in the DMET Plus product or for displaying triallelic marker plots. The software features additional tools designed to facilitate the analysis of the DMET™ Plus Panel, including the ability to define, analyze, and report on specific subsets of markers. Concordance checking utilities for the DMET Plus genomic and plasmid control samples are also included. DMET Plus allele translations are performed from genotype result sets with the

use of a single, intuitive dialog box. Genotyping results may also be exported in different report formats for downstream analysis.

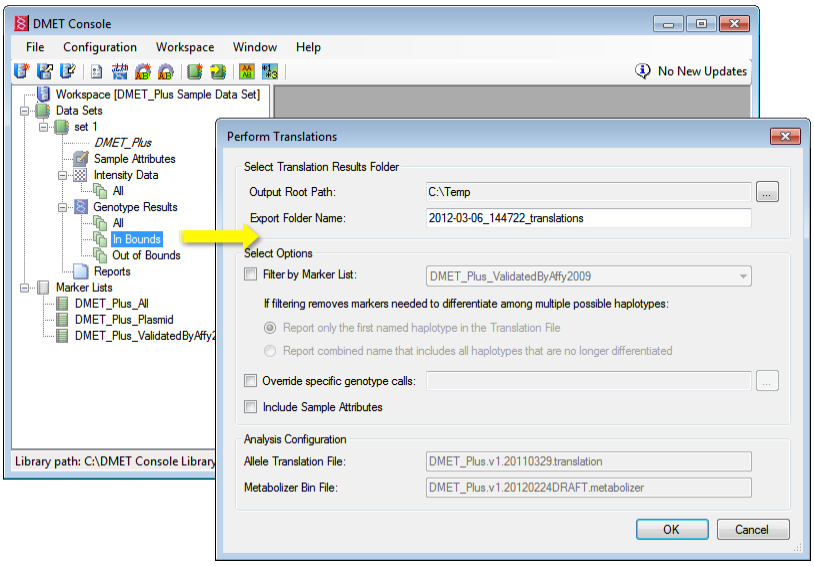


Figure 3: Allele translation dialog box within DMET™ Console Software permits all relevant analysis parameters to be set by users at the time of analysis.

Output of allele translation reports

Four reports are generated at the completion of DMET™ Plus allele translation:

- Comprehensive report – contains detailed annotations of the core gene results
- Variant summary report – summarizes genetic variants identified in the test samples across markers of interest
- Phenotype report – contains the predicted phenotype for some genes in the test samples
- Uncalled report – a template that may be used to replace missing data

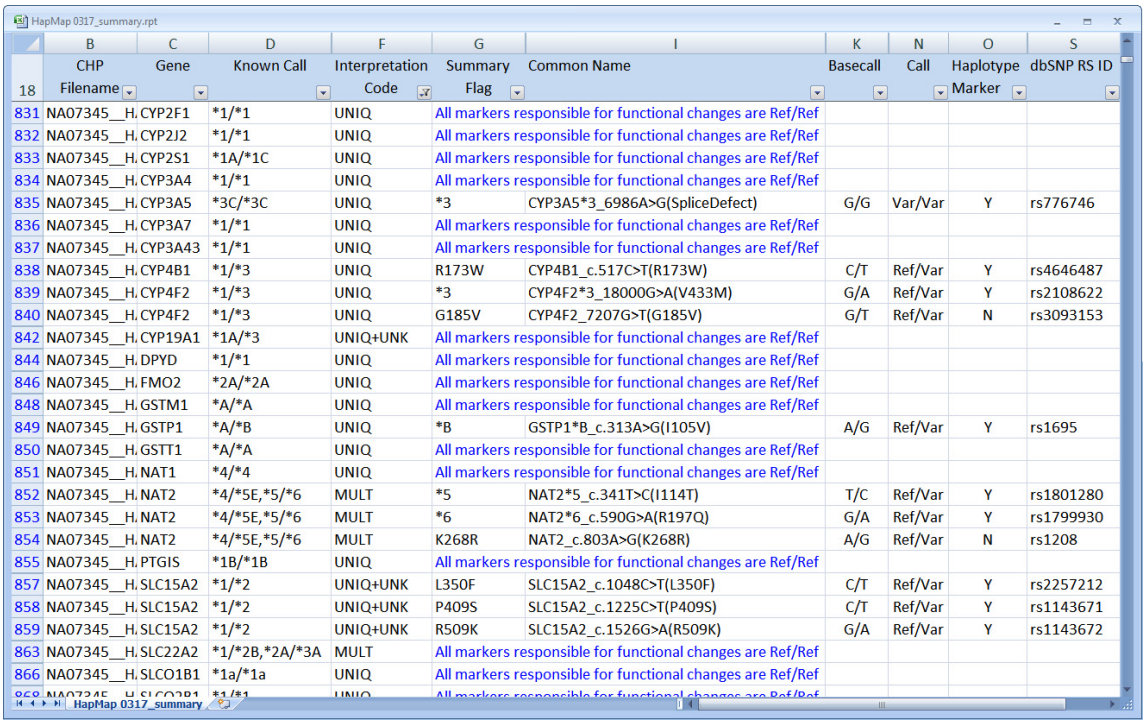
The data is saved to user-defined folders in tab-delimited files, where they can be imported into external databases or processed by downstream software.

Variant summary report summarizes allele profile of all test samples

The variant summary report is formatted in a similar manner as the comprehensive translation report, but is much more abbreviated. All defining variant alleles (in either homozygous or heterozygous states) are reported. Additional biological annotations extracted from the DMET Plus allele translation tables are included along with each marker. The marker level details provide evidence-based reporting for the structural or functional variant allele forms of each gene identified.

Table 1 shows an example of a translation summary. When more than one diplotype (haplotype pair) is consistent with a sample's diploid genotypes, all allelic pairs are reported and annotated as a multiple haplotype. One example of this is shown at the NAT2 gene; the allele call reads *4/*5E,*5/*6. This indicates that the alternative allele pairs cannot be distinguished. This is because the sample contains two heterozygous calls and the NAT2*5E allele contains both variants that define the *5 and the *6 haplotypes, while the NAT2*4 allele is the reference sequence. The file includes additional fields for interpreting both reported and unknown haplotype

calls (data not shown). Thus, a sample with the NAT2 4/*5E,*5/*6 allele calls is encoded as a "MULT" call, indicating that multiple haplotype pairs are possible due to phase ambiguity.



HapMap 0317_summary.rpt

	B	C	D	F	G	I	K	N	O	S
18	CHP Filename	Gene	Known Call	Interpretation Code	Summary Flag	Common Name	Basecall	Call	Haplotype Marker	dbSNP RS ID
831	NA07345__H	CYP2F1	*1/*1	UNIQ	All markers responsible for functional changes are Ref/Ref					
832	NA07345__H	CYP2J2	*1/*1	UNIQ	All markers responsible for functional changes are Ref/Ref					
833	NA07345__H	CYP2S1	*1A/*1C	UNIQ	All markers responsible for functional changes are Ref/Ref					
834	NA07345__H	CYP3A4	*1/*1	UNIQ	All markers responsible for functional changes are Ref/Ref					
835	NA07345__H	CYP3A5	*3C/*3C	UNIQ	*3	CYP3A5*3_6986A>G(SpliceDefect)	G/G	Var/Var	Y	rs776746
836	NA07345__H	CYP3A7	*1/*1	UNIQ	All markers responsible for functional changes are Ref/Ref					
837	NA07345__H	CYP3A43	*1/*1	UNIQ	All markers responsible for functional changes are Ref/Ref					
838	NA07345__H	CYP4B1	*1/*3	UNIQ	R173W	CYP4B1_c.517C>T(R173W)	C/T	Ref/Var	Y	rs4646487
839	NA07345__H	CYP4F2	*1/*3	UNIQ	*3	CYP4F2*3_18000G>A(V433M)	G/A	Ref/Var	Y	rs2108622
840	NA07345__H	CYP4F2	*1/*3	UNIQ	G185V	CYP4F2_7207G>T(G185V)	G/T	Ref/Var	N	rs3093153
842	NA07345__H	CYP19A1	*1A/*3	UNIQ+UNK	All markers responsible for functional changes are Ref/Ref					
844	NA07345__H	DPYD	*1/*1	UNIQ	All markers responsible for functional changes are Ref/Ref					
846	NA07345__H	FMO2	*2A/*2A	UNIQ	All markers responsible for functional changes are Ref/Ref					
848	NA07345__H	GSTM1	*A/*A	UNIQ	All markers responsible for functional changes are Ref/Ref					
849	NA07345__H	GSTP1	*A/*B	UNIQ	*B	GSTP1*B_c.313A>G(I105V)	A/G	Ref/Var	Y	rs1695
850	NA07345__H	GSTT1	*A/*A	UNIQ	All markers responsible for functional changes are Ref/Ref					
851	NA07345__H	NAT1	*4/*4	UNIQ	All markers responsible for functional changes are Ref/Ref					
852	NA07345__H	NAT2	*4/*5E,*5/*6	MULT	*5	NAT2*5_c.341T>C(I114T)	T/C	Ref/Var	Y	rs1801280
853	NA07345__H	NAT2	*4/*5E,*5/*6	MULT	*6	NAT2*6_c.590G>A(R197Q)	G/A	Ref/Var	Y	rs1799930
854	NA07345__H	NAT2	*4/*5E,*5/*6	MULT	K268R	NAT2_c.803A>G(K268R)	A/G	Ref/Var	N	rs1208
855	NA07345__H	PTGIS	*1B/*1B	UNIQ	All markers responsible for functional changes are Ref/Ref					
857	NA07345__H	SLC15A2	*1/*2	UNIQ+UNK	L350F	SLC15A2_c.1048C>T(L350F)	C/T	Ref/Var	Y	rs2257212
858	NA07345__H	SLC15A2	*1/*2	UNIQ+UNK	P409S	SLC15A2_c.1225C>T(P409S)	C/T	Ref/Var	Y	rs1143671
859	NA07345__H	SLC15A2	*1/*2	UNIQ+UNK	R509K	SLC15A2_c.1526G>A(R509K)	G/A	Ref/Var	Y	rs1143672
863	NA07345__H	SLC22A2	*1/*2B,*2A/*3A	MULT	All markers responsible for functional changes are Ref/Ref					
866	NA07345__H	SLCO1B1	*1a/*1a	UNIQ	All markers responsible for functional changes are Ref/Ref					

HapMap 0317_summary

Table 1: Example output of the variant summary. The sample's genotype (column D) is reported along with a notation of the position and modification underlying the gene change (columns G to S). Thus, the sample summary provides evidence-based reporting of potential mutations of particular importance throughout each core gene.

Phenotype report provides predicted gene activities for some genes

Genetic differences can lead to differences in the expression and activity of genes. In most cases, an individual has two copies of a gene, one from each parent. The combined activity of both copies of the gene can be used to predict the phenotype, which is the metabolizing or transporting rate of the associated enzyme for the substrates that interact with it. The Phenotype report further translates the reported diplotypes (star allele pairs) from a subset of genes in the Comprehensive report into one of several phenotype categories. As the software reads the comprehensive.rpt file, it will try to match Known Call diplotype values for each gene of each sample to one row of the metabolizer library file table. If a match is found, the associated phenotype and allele activities are written to the phenotype.rpt. If a match is not found, a Phenotype Call of "unknown" is reported. A portion of a phenotype report is shown in Table 2.

Index	CHP File	Gene	Phenotype Call	Gene Activity	Known Call	Unknown Call	Interpretation Code
0001-0020	test_01.c	CYP1A2	EM	normal/normal	*1F/*1F		UNIQ
0001-0022	test_01.c	CYP2A6	EM	normal/normal	*1/*1		UNIQ
0001-0024	test_01.c	CYP2B6	EM	normal/normal	*1/*1		UNIQ
0001-0029	test_01.c	CYP2D6	PM	none/none	*5/*5		UNIQ
0002-0020	test_02.c	CYP1A2	EM_or_IM	normal/reduced	*1A/*1L,*1C/*1F		MULT
0002-0022	test_02.c	CYP2A6	EM_or_IM	normal/reduced	*1/*17		UNIQ
0002-0024	test_02.c	CYP2B6	EM_or_IM	normal/reduced	*1/*6	*4/UNK	UNIQ+UNK
0002-0029	test_02.c	CYP2D6	EM_or_IM	normal/reduced	*2/*29	*2/UNK,*29/UNK	NC/PRA/NA
0004-0020	test_04.c	CYP1A2	EM	normal/normal	*1A/*1A		UNIQ
0004-0022	test_04.c	CYP2A6	EM	normal/normal	*1/*1		UNIQ
0004-0024	test_04.c	CYP2B6	IM	normal/none	*1/*18		UNIQ
0004-0029	test_04.c	CYP2D6	EM	normal/normal	*1/*1		UNIQ
0005-0020	test_05.c	CYP1A2	EM	normal/normal	*1A/*1F		UNIQ
0005-0022	test_05.c	CYP2A6	EM	normal/normal	*1/*1		UNIQ
0005-0024	test_05.c	CYP2B6	IM	reduced/reduced	*6/*6		UNIQ
0005-0029	test_05.c	CYP2D6	IM	normal/none	*2/*4	*1/UNK,*2/UNK,*4/I	NC/PRA/NA
0009-0020	test_09.c	CYP1A2	EM	normal/normal	*1F/*1F		UNIQ
0009-0022	test_09.c	CYP2A6	EM	normal/normal	*1/*1		UNIQ
0009-0024	test_09.c	CYP2B6	IM	reduced/reduced	*6/*6		UNIQ
0009-0029	test_09.c	CYP2D6	IM	normal/none	*2/*4	*1/UNK,*10/UNK,UN	UNIQ+UNK

Table 2: Example output of the phenotype report for 4 genes and 5 samples. The Phenotype Call is made based on the supplied Known Call from the comprehensive report and the reference metabolizer library file. If multiple Known Calls all predict the same pair of Gene Activities, then only a single Phenotype Call value will be reported.

Content of the metabolizer library file used for reporting phenotypes

NOTE: The description that follows is specific to the library file *DMET_Plus.v1.20120316_Affymetrix.metabolizer*.

Genes for which phenotypes are reported			
CYP1A2	CYP2C9	CYP3A7	SLCO1B1
CYP2A6	CYP2D6	GSTM1	TPMT
CYP2B6	CYP2E1	GSTP1	UGT1A1
CYP2C19	CYP3A4	NAT1	UGT2B7
CYP2C8	CYP3A5	NAT2	VKORC1

Table 3: Genes for which phenotypes are reported

All efforts have been made to ensure that the allele names assigned to the pattern of marker-level genotypes are consistent with literature definitions. A literature review was then performed to determine the relative gene activity levels of each allele. An example that captures the findings is given in Table 4. For alleles for which little or no evidence was found, the metabolizer library file contains an "unknown" activity value. *In vivo* studies that measure activity differences across different alleles of the gene provide the best evidence. In cases of rarer alleles or less well-studied genes, the only evidence available to assign an activity category comes from *in vitro* studies, which do not always predict activity *in vivo*. The metabolizer library file includes the level of evidence found to support the stated activity of each allele, as well as references to the literature reviewed to support the stated activity. Literature references are in the form of PubMed publication identifier numbers (http://www.ncbi.nlm.nih.gov/pubmed).

gene	allele	activity	evidence	PubMed references
UGT2B7	*1a	normal	in_vivo	
UGT2B7	*1g	normal	in_vivo	12695358
UGT2B7	*2a	increased	in_vivo	8423545|17178267|21658222|12920168|12811366
UGT2B7	*2c	increased	in_vivo	21846671|21862974|12695358
UGT2B7	*2e	unknown		
UGT2B7	*3	normal	in_vivo	17965522|15319348|21856293

Table 4: Portion of an activity table used to create the metabolizer library file.

A given allele's activity may depend on the substrate and dosage. For example, one paper noted that "CYP2D6*17 is generally considered as an allele with reduced function, but it displays remarkable variability in its activity towards substrates such as dextromethorphan, risperidone, codeine and haloperidol."[55] Another paper noted that "the most significant difference observed was a consistent but substrate-dependent decrease in the catalytic efficiencies of cDNA-expressed CYP2D6*10 and CYP2D6*17 compared with CYP2D6*1, yielding 1.32 to 27.9 and 7.33 to 80.4% of the efficiency of CYP2D6.1, respectively."[56] Affymetrix used the decision table shown in Table 5 to settle on a single activity value in the metabolizer library file when clear substrate-specific activity differences were reported.

Source 1 or substrate 1 allele activity	Source 2 or substrate 2 allele activity	Reported activity for allele in metabolizer library file
normal	reduced	reduced
normal	increased	increased
increased	reduced	unknown

Table 5: Resolving differences in allele activity among literature sources in order to create a generic metabolizer file.

Once the activity values were assembled for every reportable allele among the high visibility genes, the next step was to assign a phenotype given a pair of allele activities. As there is no consistent terminology in the literature across all genes for either relative activities or reported phenotypes, multiple lookup tables were selected to cover the genes of interest. Tables 6-10 were used to assign a pair of gene activities to a phenotype.

allele 1\2	increased	normal	reduced	none	unknown
increased	UM	UM_or_EM	Not_PM	Not_PM	Not_PM
normal		EM	EM_or_IM	IM	Not_PM
reduced			IM	IM_or_PM	unknown
none				PM	unknown
unknown					unknown

Table 6: The generic gene activity pair to phenotype lookup table, used for most genes. UM = ultrarapid metabolizer, EM = extensive metabolizer, IM = intermediate metabolizer, PM = poor metabolizer

One phenotype lookup table was not adequate to cover all of the genes. Tables 7-10 are for specific genes.

UGT allele 1\2	increased	normal	reduced	unknown
increased	UM	UM_or_EM	Not_PM	Not_PM
normal		EM	IM	Not_PM
reduced			PM	unknown
unknown				unknown

Table 7: UGT gene activity pair to phenotype lookup table. UM = ultrarapid metabolizer, EM = extensive metabolizer, IM = intermediate metabolizer, PM = poor metabolizer.

NAT allele 1\2	rapid	slow	absent	unknown
rapid	RA	IA	IA_or_SA	Not_Off
slow		SA	SA	IA_or_SA
absent			Off	Not_RA
unknown				unknown

Table 8: NAT gene activity pair to phenotype lookup table. RA = rapid acetylator, IA = intermediate acetylator, SA = slow acetylator, Off = absent expression (protein not produced).

SLCO1B1 allele 1\2	increased	normal	reduced	unknown
increased	UT_or_ET	UT_or_ET	ET_or_IT	Not_PT
normal		ET	IT	Not_PT
reduced			PT	unknown
unknown				unknown

Table 9: SLCO1B1 gene activity pair to phenotype lookup table. UT = ultrarapid transporter, ET = extensive transporter, IT = intermediate transporter, PT = poor transporter

VKORC1 allele 1\2	resistant	normal	sensitive	unknown
resistant	Resistant++	Resistant+	unknown	unknown
normal		Normal	Sensitive-	unknown
sensitive			Sensitive--	unknown
unknown				unknown

Table 10: VKORC1 gene activity pair to phenotype lookup table. The phenotypes describe the relative resistance of VKORC1 to inhibition by coumarin-like drugs.

A master metabolizer table was created that combined the information from individual allele activities and the phenotype lookup tables. This is the library file used by DMET Console Software, a portion of which is shown in Table 11.

gene	allele_1	allele_2	phenotype	activity_1	activity_2	evidence_1	evidence_2	pubmed_uid_1	pubmed_uid_2
UGT2B7	*2a	*2a	UM	increased	increased	in_vivo	in_vivo	8423545|1717826	8423545|1717826
UGT2B7	*2a	*2c	UM	increased	increased	in_vivo	in_vivo	8423545|1717826	21846671|218629
UGT2B7	*2c	*2c	UM	increased	increased	in_vivo	in_vivo	21846671|218625	21846671|218629
UGT2B7	*2a	*1a	UM_or_EM	increased	normal	in_vivo	in_vivo	8423545|17178267	21658222|1292
UGT2B7	*2a	*1g	UM_or_EM	increased	normal	in_vivo	in_vivo	8423545|1717826	12695358
UGT2B7	*2a	*3	UM_or_EM	increased	normal	in_vivo	in_vivo	8423545|1717826	17965522|153193
UGT2B7	*2c	*1a	UM_or_EM	increased	normal	in_vivo	in_vivo	21846671|21862974	12695358
UGT2B7	*2c	*1g	UM_or_EM	increased	normal	in_vivo	in_vivo	21846671|218625	12695358
UGT2B7	*2c	*3	UM_or_EM	increased	normal	in_vivo	in_vivo	21846671|218625	17965522|153193
UGT2B7	*1a	*1a	EM	normal	normal	in_vivo	in_vivo		
UGT2B7	*1a	*1g	EM	normal	normal	in_vivo	in_vivo		12695358
UGT2B7	*1a	*3	EM	normal	normal	in_vivo	in_vivo		17965522|153193
UGT2B7	*1g	*1g	EM	normal	normal	in_vivo	in_vivo	12695358	12695358
UGT2B7	*1g	*3	EM	normal	normal	in_vivo	in_vivo	12695358	17965522|153193
UGT2B7	*3	*3	EM	normal	normal	in_vivo	in_vivo	17965522|153193	17965522|153193
UGT2B7	*2a	*2e	Not_PM	increased	unknown	in_vivo		8423545|17178267	21658222|1292
UGT2B7	*2c	*2e	Not_PM	increased	unknown	in_vivo		21846671|21862974	12695358
UGT2B7	*1a	*2e	Not_PM	normal	unknown	in_vivo			
UGT2B7	*1g	*2e	Not_PM	normal	unknown	in_vivo		12695358	
UGT2B7	*3	*2e	Not_PM	normal	unknown	in_vivo		17965522|15319348	21856293
UGT2B7	*2e	*2e	unknown	unknown	unknown				

Table 11: Portion of the metabolizer library file. Each row represents one possible allele pair, with rows colored by reported phenotype. Supporting information for each allele is provided by its associated activity, evidence, and PubMed citation fields.

As DMET™ Console Software reads the comprehensive.rpt file it has just generated as part of the Allele Translation operation, it will try to match allele name pairs (a diplotype) for each gene of each sample to one row of the metabolizer table. If a match is found, the associated phenotype and allele activities are written to the phenotype.rpt. If a match is not found, an "unknown" phenotype is reported.

Phenotype reporting guidance

It is important to note that even if users were to agree on the relative activities of the alleles for a given substrate, they might still assign a given pair of activities to a different phenotype than is used by the metabolizer library file supplied by Affymetrix. For example, some users might prefer a model where a **normal/none** allele activity pair presents the same phenotype as a **normal/normal** allele activity pair, such as the CYP2D6 phenotypes predicted by the AmpliChip™ CYP450 test.[57] Therefore, instructions are available for editing a copy of the metabolizer file to suit a specific application. Instructions can be found from the Help menu within DMET Console Software, and also in the PDF version:
http://www.affymetrix.com/Auth/support/downloads/manuals/dmet_console1_user_manual.pdf

Affymetrix recognizes that there is no "one size fits all" solution to assigning phenotypes. Therefore, the metabolizer library file provided by Affymetrix should be considered as a template. **Users are responsible for reviewing the metabolizer file for accuracy, and are free to add to, modify, or remove content from a copy of the file as needed.**

Incorporating missing data into DMET™ Plus studies

Missing genotypes, in particular from SNPs that resolve multiple haplotypes across a gene, is another reason that multiple haplotype pairs may be reported in DMET Plus translation reports. Haplotypes containing both the variant and reference base are reported when data are missing

("no call" or "possible rare allele" calls). This ambiguity may occasionally be resolved by reviewing the DMET™ Console Software marker clusters. Alternatively, critical markers may be individually genotyped with alternative technologies at the user's discretion. With this objective in mind, the analysis workflow has been streamlined, as shown in Figure 4.

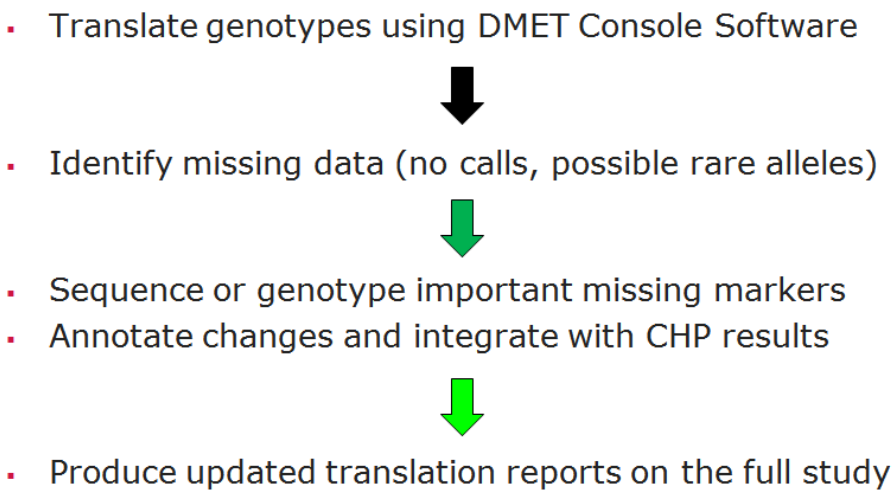


Figure 4: Workflow used to minimize missing data in haplotype calls.

The allele translation operation creates an "uncalled.rpt" file that includes a list of all missing genotyping data that can be used to direct viewing of genotyping clusters for these markers. Probe signals may be very close to, but not within, the cluster boundaries used for genotyping, so the actual call can be adjudicated by using forced calls in the extended genotype reports, by manually reviewing the marker clusters, or by using external information. NoCalls are conveniently marked using a separate symbol (X) in the cluster graphs. NoCalls in the *uncalled.rpt* can be replaced with the expected call, and this modified file can be used as an "override" file. During allele translation, the software will accept genotypes from both the CHP files and the override file, with the override file content replacing information supplied from the CHP files. In this manner, missing calls in the CHP files can be filled in at the point of allele translation (see Figure 4).

Pharmacogenetics as an emerging focus in cancer therapy

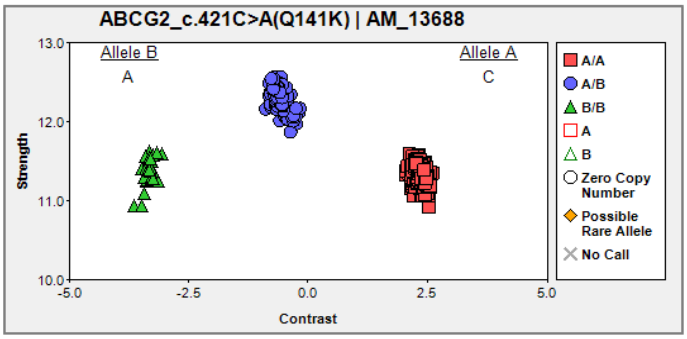
Several genes in the DMET™ Plus Assay Panel detoxify carcinogens and other environmental toxins and are thought to contribute to cancer susceptibility (GSTT1, GSTM1, CYP2D6, CYP1A1, and CYP2E1, among others).[30] Unraveling the genetic basis of toxic or other adverse reactions to chemotherapeutic agents is complex because most agents are metabolized via several genes.[31] Novel SNP-based biomarkers are being characterized with increasing frequency, such as determining the role of CYP2D6 variation in activation of tamoxifen.[32] Additionally, roles for CES2 and CES3, ABCG2,[7] and UGT1A1 mutations that alter the pharmacokinetics and toxicity of Irinotecan[14] have been replicated in multiple clinical research studies. Functional variants at these and more than 200 additional ADME genes are represented in the multiplex DMET™ Plus Assay. These genetic biomarker associations are sufficiently strong that prospective screening of patients prior to chemotherapy selection has been advocated to reduce the frequency of severe toxicities.

Two widely studied drug transporters include ABCC1, also known as multidrug resistance protein 1 (MRP1),[34] and ABCG2, also known as breast cancer resistance protein 2 (BCRP).[35] Researchers

are increasingly focusing on ABCG2 because it modulates the transport of several chemotherapeutic agents from the target tissues, including topoisomerase inhibitors daunorubicin, mitoxantrone, etoposide,[36] and 6-mercaptopurine,[37] among other drugs. In addition, ABCG2 variants have been implicated in several adverse drug reactions, including toxic reactions to erlotinib,[38] sulfasalazine,[39] and gefitinib.[40] A common genetic nonfunctional variant of ABCG2 in the DMET™ Plus Product, c.421C>A(Q141K), has been implicated in several of these adverse events.[40] The risk of occurrence and rate of survival of other cancers have also been found to be associated with this mutation[41] due to its role in the transport of cytotoxins.

ABCG2 421C>A genotypes can easily be distinguished in the DMET™ Plus Assay Panel (Figure 5A) and the allele variants are common in selected populations. Interestingly, frequency of the 421C>A allele also differs by population (Figure 5B). Notice that among the HapMap DNA samples tested, 421C>A was not observed in the African populations. In contrast, this variant is the one with the highest frequency across the three Asian populations, consistent with previously published studies.[7]

A.



B.

AM_13688 Call	Population							
	CEU	CHB	CHD	JPT	LWK	MEX	YRI	Total
C/C	72	45	40	37	86	59	140	479
A/C	16	38	41	50		28		173
A/A	1	8	8	7		3		27
Total	89	91	89	94	86	90	140	679

Figure 5: ABCG2 421C>A genotype clustering and allele frequency distribution across HapMap populations. YRI: Yoruba population of Africa; LWK: Luhya in Webuye, Kenya (African); CEU: CEPH Utah residence with ancestry from northern and western Europe (Caucasian); MEX: Mexican in Mexico City; CHB: Han Chinese in Beijing, China (Asian); CHD: Chinese in Metropolitan Denver, CO, USA; and JPT: Japanese in Tokyo, Japan. Additional information about the HapMap human diversity project is available at http://hapmap.ncbi.nlm.nih.gov.

Worldwide population diversity reflected in HapMap controls

During development of the DMET™ Plus Product, Affymetrix performed validation and verification studies on the DMET™ Plus Assay, including runs at external laboratory sites to assess genotyping reproducibility and accuracy. More than 5,300 arrays profiling 1,235 distinct DNA samples were processed by 82 operators with the goal of capturing as much nonbiological information as possible in the training data. Analysis of 597 unique DNA samples from the HapMap extended diversity panel was included in this study to provide population-based allele frequencies across the comprehensive pharmacogenomic assay panel and to test genotype accuracy (agreement with reference data) and reproducibility.

In addition to verifying population-based allele frequency differences such as in ABCG2 421C>A, analysis of the HapMap diversity panel DNA confirmed appropriate interpretation of large allele

frequency differences observed in several key ADME genes including VKORC1 variants,[42-44] high-activity CYP2A6 alleles,[45] GSTT1*0,[46-48] and the CYP2C8*5 frameshift mutation,[49] as well as several alleles at CYP1A1[47,50] and CYP1B1.[51,52] Many of these same genes are also considered to impart increased cancer disposition,[53,54] increasing the importance of better understanding them for clinical research.

To identify additional markers with extensive differences in allele frequencies across the major racial groups, variance in minor allele frequencies was determined (average of variance for each of the A, B, or C alleles for each marker in the assay panel). Table 12 summarizes the population-based allele frequencies at the 25 markers with greatest worldwide genetic diversity among the HapMap controls. It is clear that allele frequency differences are both pronounced and common across the ADME genes.

Probe ID	Common Name/ Population >	YRI (N=119	LWK (N=89)	CEU (N=59)	MEX (N=60)	CHB (N=90)	CHD (N=87)	JPT (N=91)
AM_11045	VKORC1_c.174-136C>T	0.03	0.06	0.38	0.45	0.95	0.94	0.90
AM_11054	VKORC1_c.-1639G>A(Promoter)	0.03	0.06	0.38	0.45	0.95	0.93	0.90
AM_12590	SULT1C4_c.15C>G(D5E)	0.04	0.07	0.81	0.84	0.95	0.94	0.80
AM_11032	PRSS53_c.89A>G(Q30R)	0.00	0.00	0.02	0.30	0.83	0.82	0.76
AM_14846	CYP3A4_-392A>G	0.23	0.17	0.98	0.93	1.00	1.00	1.00
AM_13800	ADH1B_c.143G>A(R48H)	0.00	0.00	0.00	0.09	0.72	0.75	0.73
AM_11040	VKORC1_c.283+837C>T	0.15	0.19	0.38	0.45	0.95	0.93	0.90
AM_14111	PPARD_c.-102+15103T>C	0.24	0.24	0.92	0.95	0.94	0.98	0.96
AM_12327	SLC5A6_c.1442T>C(F481S)	0.16	0.15	0.71	0.43	0.89	0.89	0.95
AM_14116	PPARD_c.-102+24108C>T	0.24	0.25	0.92	0.95	0.94	0.98	0.96
AM_14120	PPARD_c.-101-25241A>G	0.24	0.25	0.92	0.95	0.94	0.98	0.96
AM_11043	VKORC1_c.283+124G>C	0.23	0.17	0.38	0.48	0.95	0.93	0.90
AM_11734	GSTM3_c.468+24_26delAGG	0.26	0.32	0.90	0.87	1.00	1.00	1.00
AM_12461	CYP1B1*3_4326C>G(L432V)	0.12	0.22	0.55	0.70	0.92	0.83	0.89
AM_14759	CYP3A5*3_6986A>G(SpliceDefect)	0.16	0.13	0.94	0.74	0.68	0.76	0.74
AM_14814	CYP3A4_20230G>A	0.14	0.13	0.92	0.59	0.74	0.76	0.73
AM_13646	DCK_c.*165C>T(3'UTR)	0.26	0.38	0.95	0.90	0.95	0.96	0.93
AM_13741	ADH4_c.-711A>G	0.23	0.17	0.13	0.09	0.68	0.75	0.72
AM_13760	ADH6_c.-1992G>T	0.30	0.20	0.28	0.13	0.77	0.79	0.80
AM_14234	GSTA2_c.629A>C(E210A)	0.23	0.28	0.93	0.93	0.83	0.83	0.77
AM_14859	CYP3A43_21503T>C(N198N)	0.41	0.30	0.93	0.91	0.99	1.00	1.00
AM_14127	PPARD_c.-101-15916T>C	0.32	0.36	0.92	0.95	0.93	0.97	0.96
AM_10490	SLCO1B3_c.1833A>G(G611G)	0.85	0.72	0.10	0.14	0.26	0.18	0.29
AM_11730	GSTM3_c.670G>A(V224I)	0.03	0.14	0.39	0.40	0.74	0.75	0.73
AM_11740	SLC16A1_c.*1942T>C	0.38	0.35	0.73	0.92	0.99	0.99	1.00

Table 12: DMET™ Plus markers with greatest genetic diversity. Observed frequency by population is reported using the allele with lowest frequency in the Yoruba population of Africa.

Interestingly, with the exception of the VKORC1 markers in this table, the alleles with low frequency in the African populations are the major allele in both Asian and Caucasian groups. Indeed, selected alleles that are rare in the African populations have become fixed in other selected populations, in particular among the three Asian groups. Examples of this include CYP3A4_-392A>G (rs2740574) and GSTM3_c.468+24_26delAGG (rs1799735). Thus, global differences in allele frequency are reflected at both the individual marker level and across full gene haplotype calls such as the CYP star allele. Indeed, principal component analysis across the full DMET™ Plus Assay Panel predicts dramatic differences in drug metabolism profiles across the major racial groups (Figure 6).

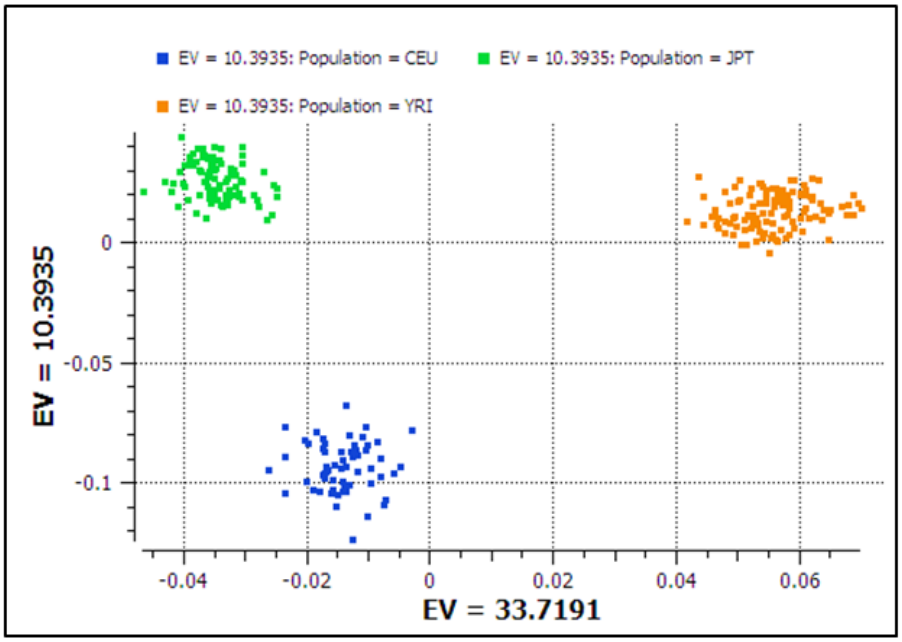


Figure 6: Principal component analysis of DMET™ Plus marker frequencies across three major racial groups. The principal component analysis (PCA) tool in the SNP & Variation Suite (Golden Helix) was used to identify major genotypic covariates in drug metabolism. A plot of the first two eigenvalues computed by principal component analysis shows complete segregation of these three racial groups across HapMap. Populations used in the study are described in the legend of Figure 5 and include independent DNA samples from Japan (JPT; green), Eastern European Caucasians (CEU; royal blue), and Yoruba, Africa (YRI; orange).

References

1. Frueh F. W., *et al*. Pharmacogenomic biomarker information in drug labels approved by the United States Food and Drug Administration: prevalence of related drug use. *Pharmacotherapy* **28**(8):992-8 (2008).
2. Lacana E., *et al*. The emerging role of pharmacogenomics in biologics. *Clin Pharmacol Ther* **82**(4):466-71 (2007).
3. Frueh F. W. and D. Gurwitz. From pharmacogenetics to personalized medicine: a vital need for educating health professionals and the community. *Pharmacogenomics* **5**(5):571-9 (2004).
4. Singh S. S. Preclinical pharmacokinetics: an approach towards safer and efficacious drugs. *Curr Drug Metab* **7**(2):165-82 (2006).
5. Yoshikawa M., *et al*. Novel camptothecin analogues that circumvent ABCG2-associated drug resistance in human tumor cells. *Int J Cancer* **110**(6):921-7 (2004).
6. Bram E. E ., *et al*. C421 allele-specific ABCG2 gene amplification confers resistance to the antitumor triazoloacridone C-1305 in human lung cancer cells. *Biochem Pharmacol* **74**(1):41-53 (2007).
7. de Jong F. A., *et al*. ABCG2 pharmacogenetics: ethnic differences in allele frequency and assessment of influence on irinotecan disposition. *Clin Cancer Res* **10**(17):5889-94 (2004).
8. Sparreboom A., *et al*. Effect of ABCG2 genotype on the oral bioavailability of topotecan. *Cancer Biol Ther* **4**(6):650-8 (2005).

9. Deeken J. F., *et al*. Toward individualized treatment: prediction of anticancer drug disposition and toxicity with pharmacogenetics. *Anticancer Drugs* **18**(2):111-26 (2007).
10. Bosch T. M. Pharmacogenomics of drug-metabolizing enzymes and drug transporters in chemotherapy. *Methods Mol Biol* **448**:63-76 (2008).
11. Fakhoury M., *et al*. Should TPMT genotype and activity be used to monitor 6-mercaptopurine treatment in children with acute lymphoblastic leukaemia? *J Clin Pharm Ther* **32**(6):633-9 (2007).
12. Maitland M. L., *et al.* TPMT, UGT1A1 and DPYD: genotyping to ensure safer cancer therapy? *Trends Pharmacol Sci* **27**(8):432-7 (2006).
13. Ekhart C., *et al*. Influence of polymorphisms of drug metabolizing enzymes (CYP2B6, CYP2C9, CYP2C19, CYP3A4, CYP3A5, GSTA1, GSTP1, ALDH1A1 and ALDH3A1) on the pharmacokinetics of cyclophosphamide and 4-hydroxycyclophosphamide. *Pharmacogenet Genomics* **18**(6):515-23 (2008).
14. Toffoli G., *et al*. The role of UGT1A1*28 polymorphism in the pharmacodynamics and pharmacokinetics of irinotecan in patients with metastatic colorectal cancer. *J Clin Oncol* **24**(19):3061-8 (2006).
15. Flockhart D. A. and J. E. Tanus-Santos. Implications of cytochrome P450 interactions when prescribing medication for hypertension. *Arch Intern Med* **162**(4):405-12 (2002).
16. Shin J. G., *et al*. A. Potent inhibition of CYP2D6 by haloperidol metabolites: stereoselective inhibition by reduced haloperidol. *Br J Clin Pharmacol* **51**(1):45-52 (2001).
17. Flockhart D. A. and J. R. Oesterheld. Cytochrome P450-mediated drug interactions. Child *Adolesc Psychiatr Clin N Am* **9**(1):43-76 (2000).
18. Ketter T. A., *et al*. The emerging role of cytochrome P450 3A in psychopharmacology. *J Clin Psychopharmacol* **15**(6):387-98 (1995).
19. Foti R. S., *et al*. Metabolism and related human risk factors for hepatic damage by usnic acid containing nutritional supplements. *Xenobiotica* **38**(3):264-80 (2008).
20. Daly T. M., *et al*. Multiplex assay for comprehensive genotyping of genes involved in drug metabolism, excretion, and transport. *Clin Chem* **53**(7):1222-30 (2007).
21. Caldwell M. D., *et al*. CYP4F2 genetic variant alters required warfarin dose. *Blood* **111**(8): 4106-12 (2008).
22. Cooper G. M., *et al*. A genome-wide scan for common genetic variants with a large influence on warfarin maintenance dose. *Blood* **112**(4):1022-7 (2008).
23. Robarge J. D., *et al*. The star-allele nomenclature: retooling for translational genomics. *Clin Pharmacol Ther* **82**(3):244-8 (2007).
24. Bolt H. M. and J. G. Hengstler. A new series of review articles on drug metabolizing enzymes: nomenclature of isoenzyme families, genetic organization, polymorphisms, substrate specificities, clinical relevance and role in carcinogenesis. *Arch Toxicol* **82**(7):413-4 (2008).
25. Marsh S., *et al*. Population variation in VKORC1 haplotype structure. *J Thromb Haemost* **4**(2):473-4 (2006).
26. Limdi N. A., *et al*. VKORC1 polymorphisms, haplotypes and haplotype groups on warfarin dose among African-Americans and European-Americans. *Pharmacogenomics* **9**(10):1445-58 (2008).
27. Flockhart D. A., *et al*. Pharmacogenetic testing of CYP2C9 and VKORC1 alleles for warfarin. *Genet Med* **10**(2):139-50 (2008).
28. Rieder M. J., *et al*. Effect of VKORC1 haplotypes on transcriptional regulation and warfarin dose. *N Engl J Med* **352**(22):2285-93 (2005).
29. Geisen C., *et al*. VKORC1 haplotypes and their impact on the inter-individual and inter-ethnical variability of oral anticoagulation. *Thromb Haemost* **94**(4):773-9 (2005).
30. Aydin-Sayitoglu M., *et al*. Role of CYP2D6, CYP1A1, CYP2E1, GSTT1, and GSTM1 genes in the susceptibility to acute leukemias. *Am J Hematol* **81**(3):162-70 (2006).

31. Owen R. P., *et al.* PharmGKB and the International Warfarin Pharmacogenetics Consortium: the changing role for pharmacogenomic databases and single-drug pharmacogenetics. *Hum Mutat* **29**(4):456-60 (2008).
32. Borges S., *et al*. Quantitative effect of CYP2D6 genotype and inhibitors on tamoxifen metabolism: implication for optimization of breast cancer treatment. *Clin Pharmacol Ther* **80**(1):61-74 (2006).
33. Bellott R., *et al*. Functional study of the 830C>G polymorphism of the human carboxylesterase 2 gene. *Cancer Chemother Pharmacol* **61**(3):481-8 (2008).
34. Morrow C. S., *et al*. Multidrug resistance protein 1 (MRP1, ABCC1) mediates resistance to mitoxantrone via glutathione-dependent drug efflux. *Mol Pharmacol* **69**(4):1499-505 (2006).
35. Krishnamurthy P. and J. D. Schuetz. Role of ABCG2/BCRP in biology and medicine. *Annu Rev Pharmacol Toxicol* **46**:381-410 (2006).
36. Schellens J. H., *et al*. Transport of topoisomerase I inhibitors by the breast cancer resistance protein. Potential clinical implications. *Ann N Y Acad Sci* **922**:188-94 (2000).
37. van der Kolk D. M., *et al*. The role of drug efflux pumps in acute myeloid leukemia. *Leuk Lymphoma* **43**(4):685-701 (2002).
38. Rudin C. M., *et al*. Pharmacogenomic and pharmacokinetic determinants of erlotinib toxicity. *J Clin Oncol* **26**(7):1119-27 (2008).
39. Zaher H., *et al*. Breast cancer resistance protein (Bcrp/abcg2) is a major determinant of sulfasalazine absorption and elimination in the mouse. *Mol Pharm* **3**(1):55-61 (2006).
40. Cusatis G., *et al*. Pharmacogenetics of ABCG2 and adverse reactions to gefitinib. *J Natl Cancer Inst* **98**(23):1739-42 (2006).
41. Hu L. L., *et al*. BCRP gene polymorphisms are associated with susceptibility and survival of diffuse large B-cell lymphoma. *Carcinogenesis* **28**(8):1740-4 (2007).
42. Aklillu E., *et al*. VKORC1 Asp36Tyr warfarin resistance marker is common in Ethiopian individuals. *Blood* **111**(7):3903-4 (2008).
43. Momary K. M., *et al*. Factors influencing warfarin dose requirements in African-Americans. *Pharmacogenomics* **8**(11):1535-44 (2007).
44. Nakai K., *et al*. Ethnic differences in the VKORC1 gene polymorphism and an association with warfarin dosage requirements in cardiovascular surgery patients. *Pharmacogenomics* **8**(7):713-9 (2007).
45. Schoedel K. A., *et al*. Ethnic variation in CYP2A6 and association of genetically slow nicotine metabolism and smoking in adult Caucasians. *Pharmacogenetics* **14**(9):615-26 (2004).
46. Ambrosone C. B., *et al*. Polymorphisms in glutathione S-transferases (GSTM1 and GSTT1) and survival after treatment for breast cancer. *Cancer Res* **61**(19):7130-5 (2001).
47. Babu K. A., *et al*. GSTM1, GSTT1 and CYP1A1 detoxification gene polymorphisms and their relationship with advanced stages of endometriosis in South Indian women. *Pharmacogenet Genomics* **15**(3):167-72 (2005).
48. Pesch B., *et al*. Polymorphic metabolic susceptibility genes and longevity: a study in octogonarians. *Toxicol Lett* **151**(1):283-90 (2004).
49. Nakajima M., *et al*. Genetic polymorphisms of CYP2C8 in Japanese population. *Drug Metab Dispos* **31**(6):687-90 (2003).
50. Al-Dayel F., *et al*. Polymorphisms of drug-metabolizing enzymes CYP1A1, GSTT and GSTP contribute to the development of diffuse large B-cell lymphoma risk in the Saudi Arabian population. *Leuk Lymphoma* **49**(1):122-9 (2008).
51. Sivadorai P., *et al*. Genetic heterogeneity and minor CYP1B1 involvement in the molecular basis of primary congenital glaucoma in Gypsies. *Clin Genet* **74**(1):82-7 (2008).

52. Yoshimura K., *et al*. Allele frequencies of single nucleotide polymorphisms (SNPs) in 40 candidate genes for gene-environment studies on cancer: data from population-based Japanese random samples. *J Hum Genet* **48**(12):654-8 (2003).
53. Abbas, A. *et al*. GSTM1, GSTT1, GSTP1 and CYP1A1 genetic polymorphisms and susceptibility to esophageal cancer in a French population: different pattern of squamous cell carcinoma and adenocarcinoma. *World J Gastroenterol* **10**(23):3389-93 (2004).
54. Bowen D. T., *et al*. CYP1A1*2B (Val) allele is overrepresented in a subgroup of acute myeloid leukemia patients with poor-risk karyotype associated with NRAS mutation, but not associated with FLT3 internal tandem duplication. *Blood* **101**(7):2770-4 (2003).
55. Zhou S. F. Polymorphism of human cytochrome P450 2D6 and its clinical significance: Part I. *Clin Pharmacokinet*. **48**(11):689-723 (2009)
56. Shen H, *et al*. Comparative metabolic capabilities and inhibitory profiles of CYP2D6.1, CYP2D6.10, and CYP2D6.17. *Drug Metab Dispos*. **35**(8):1292-300 (2007).
57. Rebsamen M. C., *et al*. The AmpliChip CYP450 test: cytochrome P450 2D6 genotype assessment and phenotype prediction. *Pharmacogenomics J.* **9**(1):34-41 (2009).