Accurate detection of copy number variants in ADME genes using PharmacoScan Solution

R. Varma, D. Black, J. Gollub, C. Bruckner, J. Law, M. Mittmann, T. Webster, A. Roter, M. Shapero, J. Schmidt

Affymetrix, Inc., Santa Clara, CA 95051

affymetrix

part of Thermo Fisher Scientific

ABSTRACT 965T

Copy number variations (CNVs) in genes relevant to drug absorption, distribution, metabolism, and excretion (ADME) have been characterized using several technologies. CNV detection using a standard single-nucleotide polymorphism (SNP) genotyping assay is efficient and advantageous because the genotypes of these genes are of utmost importance. We report here on the development of a CNV detection method for five ADME genes using PharmacoScan™ Solution, a cost-effective, high-throughput pharmacogenomics analysis solution. Copy number calls are compared to TaqMan® copy number assays, published data, or both, when available.

We focus on CNV calling for five genes: CYP2A6, CYP2D6, GSTM1, GSTT1, and UGT2B17. CYP2A6 and CYP2D6 are particularly difficult to characterize because of the presence of closely related pseudogenes, which may form hybrids with their respective genes. Copy number states called include gains and homozygous and hemizygous losses. Both SNP and non-polymorphic probes are used for copy number.

Several hundred samples from the HapMap and 1000 Genome Projects were analyzed using PharmacoScan Solution, and CNVs were called in nine regions: GSTM1, GSTT1, UGT2B17, three CYP2D6 sub-regions, and three CYP2A6 sub-regions. The predicted copy number states were compared to TaqMan Assay results, and to sequencing results reported by the 1000 Genomes Project. Results show high concordance in copy number calls for all nine regions between PharmacoScan Solution and other technologies. For CYP2D6 and CYP2A6, many samples were identified with differing copy number calls in the sub-regions, demonstrating the ability to distinguish copy number changes that are likely due to complex recombination events. In order to obtain equivalent results using TaqMan Assays, multiple assays probing different locations within the gene are required. In contrast, the method presented here has the advantage of using a single assay to call CNVs in multiple sub-regions, while also providing comprehensive genotypic information in genes of interest.

Finally, copy number calls are used to improve genotyping accuracy and translation to star-allele nomenclature on these important ADME genes.

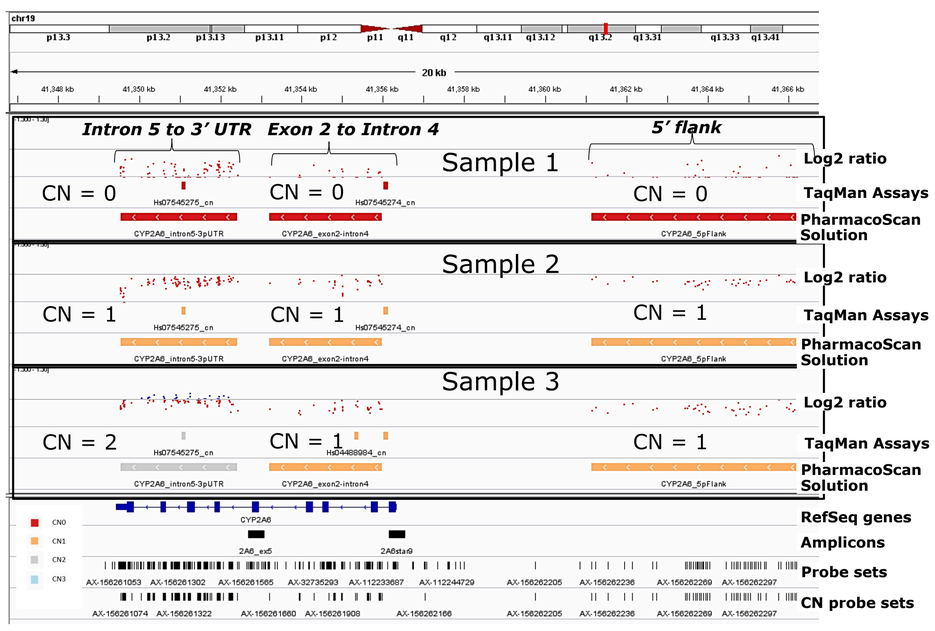
Copy number of ADME genes

Gene	Substrates (partial list)
CYP2A6	Nicotine, cotinine, coumarin, halothane, losigamone
CYP2D6	Various antidepressants, neuroleptics, beta-blockers, antiarrhythmics
GSTM1	Intermediates of polycyclic aromatic hydrocarbons
GSTT1	Cytostatic drugs, hydrocarbons
UGT2B17	Eugenol, steroid hormones

ADME genes with common copy number changes: Deletion or amplification of CYP2A6 and CYP2D6 impacts the rate of drug metabolism, with significant implications for treatment. The very common deletions GSTM1 and GSTT1 are partially compensated by other GST enzymes of the same classes, but nevertheless affect drug metabolism and cancer risk, in particular from smoking. Individuals with a double deletion of UGT2B17 have been shown to have significantly higher levels of testosterone and higher risk of prostate cancer.

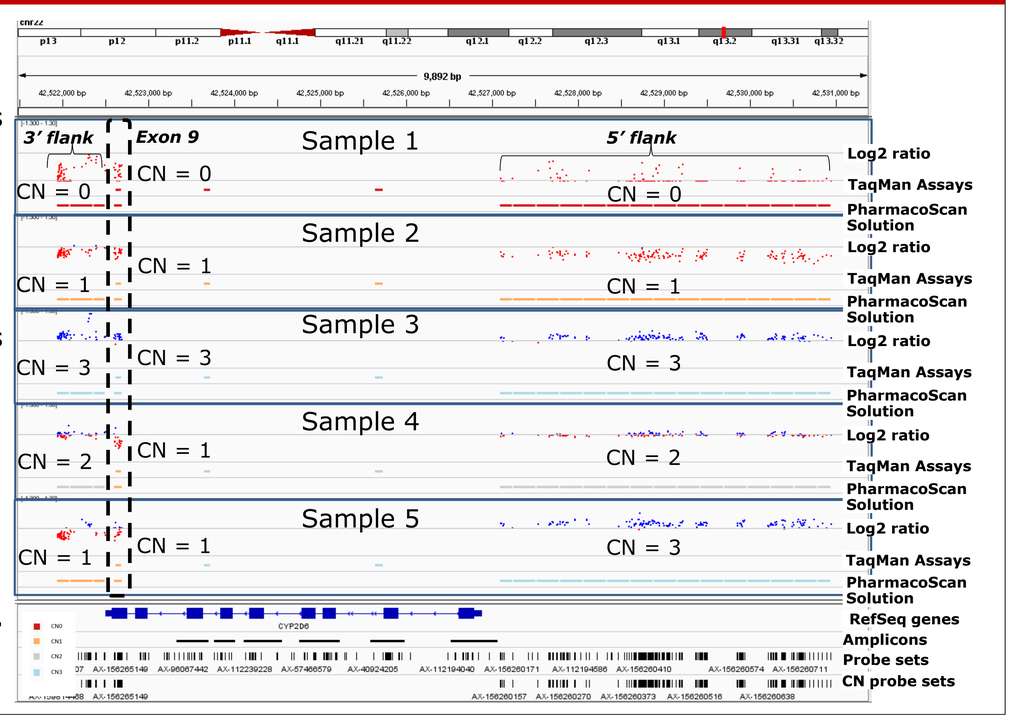
CYP2A6

- Copy number calls are made in three sub-regions of the gene: intron 5 to 3′ UTR, exon 2 to intron 4, and 5′ flank.
- Samples with differing copy numbers in the three sub-regions are identified (Sample 3).
- Calls show high concordance with TaqMan Assays in sub-regions overlapping TaqMan locations.

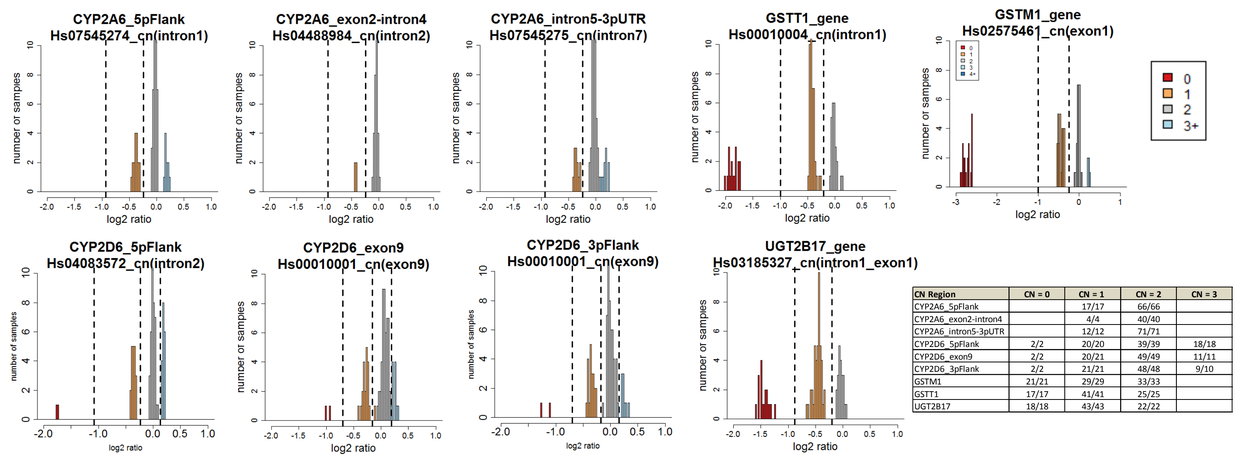


CYP2D6

- Copy number calls are made in three sub-regions of the gene: 3′ flank, exon 9, and 5′ flank.
- Samples with differing copy numbers in the three sub-regions are identified (Samples 4 and 5).
- Calls show high concordance with TaqMan Assays in sub-regions overlapping TaqMan locations.



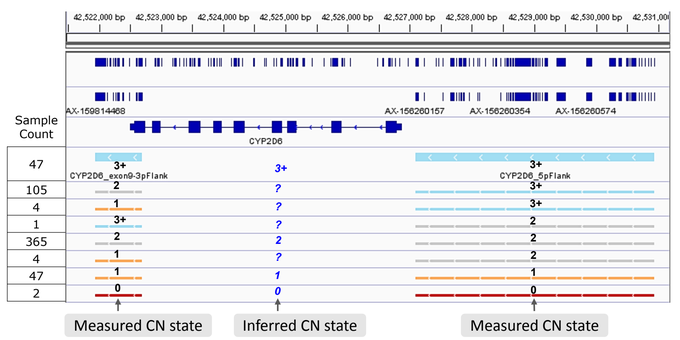
Validation



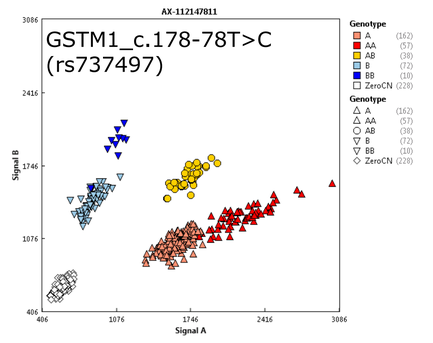
CN Region	CN = 0	CN = 1	CN = 2	CN = 3
CYP2A6_5pFlank		17/17	66/66	
CYP2A6_exon2-intron4		4/4	40/40	
CYP2A6_intron5-3pUTR		12/12	71/71	
CYP2D6_5pFlank	2/2	20/20	39/39	18/18
CYP2D6_exon9	2/2	20/21	49/49	11/11
CYP2D6_3pFlank	2/2	21/21	48/48	9/10
GSTM1	21/21	29/29	33/33	
GSTT1	17/17	41/41	25/25	
UGT2B17	18/18	43/43	22/22	

Seventy-seven samples with TaqMan copy number data were assayed with PharmacoScan Solution, six in duplicate. Figures show histograms of median log2 ratios of all samples colored by TaqMan calls. The table shows the number of concordant calls and total number of samples in each cell.

Copy number aware genotyping



SNP copy number is assessed as above. SNPs between copy number measurement regions are assigned the flanking copy number, or 2 (diploid) if the two regions disagree. PharmacoScan Solution CYP2D6 SNPs mostly use amplicons to avoid CYP2D7 interference. Amplicon regions are not informative for CN measurement, requiring CN state to be inferred for most of CYP2D6.



CN = 1 samples are genotyped using a haploid model. CN ≥2 samples are genotyped using a diploid model. CN = 0 samples are assigned a zero copy state and not genotyped.

Translation to star allele nomenclature

- CYPD2D6*5 is defined as deletion of an entire gene copy.
- CN information enables reporting of single and double copy deletions.
- CN gains are captured by the Interpretation Code field.

CN = 1 haploid SNP calls

```
#Interpretation Code Description:
#UNIQ: Unique haplotype pair
#MULT: Multiple haplotype pairs possible due to phase ambiguity
#UNIQ+UNK: Unique annotated haplotype pair, with other haplotype pairs requiring unannotated haplotypes also possible
#MULT+UNK: Multiple annotated haplotype calls possible, with other haplotype pairs requiring unannotated haplotypes also possible
#UNDH: Only haplotype pairs with undefined haplotypes possible
#NC/PRA/NotAvailable: NoCall, PossibleRareAllele, or NotAvailable call for one or more markers resulting in multiple haplotype pairs
#NoHAP: No haplotypes defined at this gene
#CN_HybridLoss: Partial gene deletion detected, so haplotype pair calling not available
#CN_HybridGain: Partial gene duplication detected
#CN_Gain: Gene duplication detected, but cannot determine which haplotype(s) are duplicated
#CN_NoCall: Copy Number state not reported, so less confidence in reported haplotype pairs
#CN_Error: Genotypes from multiple Copy Number states, so Haplotype pair calling not available
```

Index	Filename	Gene	Known Call	Interpretation Code	Summary Flag	Common Name	Probe Set ID	Basecall	Reference Base	Variant Base	Call
0002-0029-28	HG00260_JS_CYP2D6		*2/*5	UNIQ	S486T	CYP2D6_4180G>C(S486T)	AX-11155878	G	C	G	*5/Var
0002-0029-17	HG00260_JS_CYP2D6		*2/*5	UNIQ	R296C	CYP2D6_2850C>T(R296C)	AX-112192888	A	G	A	*5/Var
0010-0029-28	NA11829_T01	CYP2D6	*4/*5	UNIQ	S486T	CYP2D6_4180G>C(S486T)	AX-11155878	G	C	G	*5/Var
0010-0029-09	NA11829_T01	CYP2D6	*4/*5	UNIQ	*4	CYP2D6*4_1846G>A(SpliceDefect)	AX-112193595	T	C	T	*5/Var
0010-0029-01	NA11829_T01	CYP2D6	*4/*5	UNIQ	P34S	CYP2D6_100C>T(P34S)	AX-112194311	A	G	A	*5/Var
0015-0029-28	NA18973_CS	CYP2D6	*1*21	UNIQ+UNK	S486T	CYP2D6_4180G>C(S486T)	AX-11155878	C/G	C	G	Ref/Var
0015-0029-17	NA18973_CS	CYP2D6	*1*21	UNIQ+UNK	R296C	CYP2D6_2850C>T(R296C)	AX-112192888	G/A	G	A	Ref/Var
0015-0029-14	NA18973_CS	CYP2D6	*1*21	UNIQ+UNK	*21	CYP2D6*21_2573insC	AX-112192915	-/G	-	G	Ref/Var
0019-0029-28	NA19315_T01	CYP2D6	*1*29	UNIQ+UNK	S486T	CYP2D6_4180G>C(S486T)	AX-11155878	C/G	C	G	Ref/Var
0019-0029-17	NA19315_T01	CYP2D6	*1*29	UNIQ+UNK	R296C	CYP2D6_2850C>T(R296C)	AX-112192888	G/A	G	A	Ref/Var
0019-0029-20	NA19315_T01	CYP2D6	*1*29	UNIQ+UNK	*29	CYP2D6*29_3183G>A(V338M)	AX-122945650	C/T	C	T	Ref/Var
0019-0029-05	NA19315_T01	CYP2D6	*1*29	UNIQ+UNK	*29	CYP2D6*29_1659G>A(V136I)	AX-165876996	C/T	C	T	Ref/Var
0023-0029-01	NA19317_CS	CYP2D6	*5/*5	UNIQ	Deletion alleles have been detected for this gene, CN=0						
0025-0029-28	NA19318_CS	CYP2D6	*4/*40	UNIQ+UNK	S486T	CYP2D6_4180G>C(S486T)	AX-11155878	G/G	C	G	Var/Var
0025-0029-17	NA19318_CS	CYP2D6	*4/*40	UNIQ+UNK	R296C	CYP2D6_2850C>T(R296C)	AX-112192888	G/A	G	A	Ref/Var
0025-0029-09	NA19318_CS	CYP2D6	*4/*40	UNIQ+UNK	*4	CYP2D6*4_1846G>A(SpliceDefect)	AX-112193595	C/T	C	T	Ref/Var
0025-0029-04	NA19318_CS	CYP2D6	*4/*40	UNIQ+UNK	T107I	CYP2D6_1023C>T(T107I)	AX-112193805	G/A	G	A	Ref/Var
0025-0029-01	NA19318_CS	CYP2D6	*4/*40	UNIQ+UNK	P34S	CYP2D6_100C>T(P34S)	AX-112194311	G/A	G	A	Ref/Var
0025-0029-10	NA19318_CS	CYP2D6	*4/*40	UNIQ+UNK	*40	CYP2D6*40_1863ins(TTTCGCCCC)2	AX-112283720	-/GGGGCGAAAGGC	-	GGGGCGAAAGGC	Ref/Var
0031-0029-28	NA19782_JS	CYP2D6	*1/*4	UNIQ+UNK	S486T	CYP2D6_4180G>C(S486T)	AX-11155878	C/G	C	G	Ref/Var
0031-0029-09	NA19782_JS	CYP2D6	*1/*4	UNIQ+UNK	*4	CYP2D6*4_1846G>A(SpliceDefect)	AX-112193595	C/T	C	T	Ref/Var
0031-0029-01	NA19782_JS	CYP2D6	*1/*4	UNIQ+UNK	P34S	CYP2D6_100C>T(P34S)	AX-112194311	G/A	G	A	Ref/Var
0032-0029-28	NA19785_CS	CYP2D6	*1/*2	CN_HybridGain	S486T	CYP2D6_4180G>C(S486T)	AX-11155878	C/G	C	G	Ref/Var
0032-0029-17	NA19785_CS	CYP2D6	*1/*2	CN_HybridGain	R296C	CYP2D6_2850C>T(R296C)	AX-112192888	G/A	G	A	Ref/Var

For Research Use Only. Not for use in diagnostic procedures.