AlleleTyper™ Software: a Flexible Application for Mapping SNP Genotype and CNV Data Patterns to Pharmacogenomic Allele Nomenclature

Toinette Hartshorne, Necip Mehmet, Elliot Shelton, Harrison Leong. Life Technologies, South San Francisco, CA 94080

ABSTRACT

Pharmacogenomic (PGx) studies require genetic testing of individuals for multiple variants in drug metabolism enzyme and transporter genes. For phenotype interpretation purposes, genotyping results must be translated to star (*) allele nomenclature. Star alleles are haplotype patterns that have been defined at the gene level and, in many cases, associated with protein activity levels. Genetic variants within a haplotype can include single nucleotide polymorphisms (SNPs), Insertion/Deletions (InDels), and copy number variants (CNVs). Knowing the combination of variants within a given haplotype, and the diploid content in an individual, is of key importance for studying drug metabolism, drug response and adverse drug reactions. To facilitate the translation of results for individuals genotyped in studies using TaqMan® SNP and Drug Metabolism Genotyping Assays and TaqMan® Copy Number Assays, we developed a web-based flexible software tool called AlleleTyper™. This software uses genetic pattern information in user-defined translation tables to map sample genotyping data to star allele or other nomenclature.

INTRODUCTION

AlleleTyper™ Software is an automated data analysis application that translates genetic pattern information from TaqMan® SNP and DME Genotyping Assays and/or TaqMan® Copy Number Assay results to user-defined genotype or diplotype nomenclature. SNP assay and copy number assay experiment data generated on an Life Technologies real-time PCR system must be analyzed by TaqMan® Genotyper™ Software and CopyCaller® Software, respectively, as results files exported from these software are input files for AlleleTyper™ Software. User-defined monoallelic translation tables containing haplotype genetic information for the targeted gene variants in a study are automatically converted by the software to biallelic translators containing diploid genetic patterns. AlleleTyper™ matches the sample genotypes in results files from TaqMan® Genotyper™ Software and/or CopyCaller® Software to the patterns in the biallelic translator and reports the star allele genotypes determined for each individual.

ALLELE NOMENCLATURE

Standardized star (*) allele nomenclature (e.g. CYP2D6*4) is used to describe genetic variants in CYP P450 and other drug metabolism genes and their phenotypic outcome. Star alleles are gene-level haplotypes (a set of DNA polymorphisms that tend to be inherited together on the same chromosome), which in many cases have been associated with DME phenotypes (e.g. functional, decreased function, or nonfunctional variants). Genetic variants within an allele can include SNPs, InDels and CNVs. This nomenclature can be complicated as many alleles contain more than one polymorphism (i.e. they are haplotypes) and conversely, many polymorphisms are associated with several alleles. Translation analysis is required to associate sample genetic patterns with star allele haplotypes.

SUMMARY

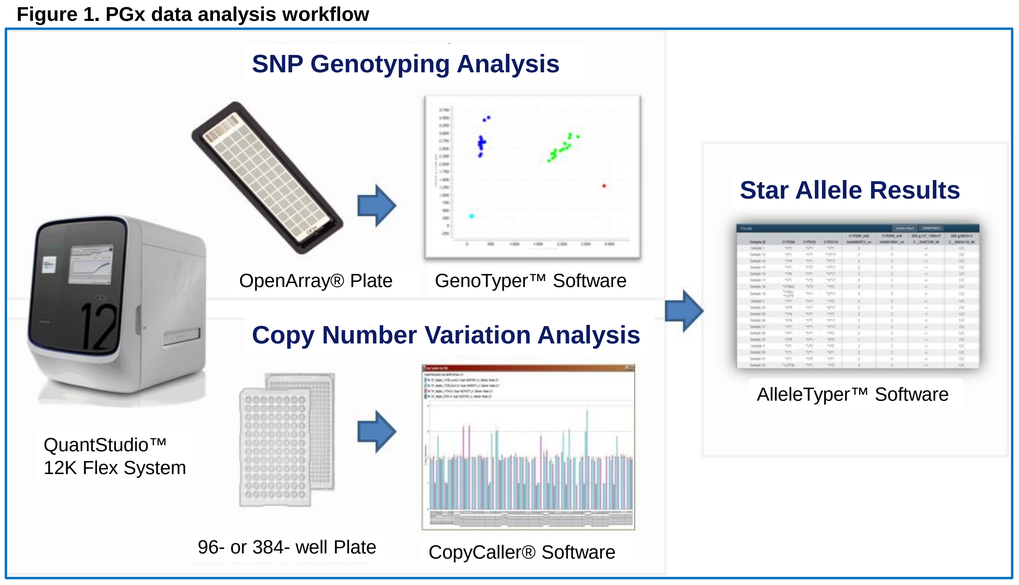
AlleleTyper™ Software greatly facilitates PGx study data analysis, particularly for high throughput studies. This software is also flexible enough to be used for other genotyping applications requiring translation of data from multiple TaqMan® assays, including triallelic SNP data analysis and blood genotyping. As well, it can be used to translate results for special cases such as triallelic SNP interrogation by two SNP assays, or simply to provide a name for a particular genetic outcome for a given assay.

REFERENCES

1. Cytochrome P450 (CYP) Allele Nomenclature website: www.cypalleles.ki.se
2. PharmGKB website: www.pharmgkb.org

METHODS AND RESULTS

Figure 1. PGx data analysis workflow



Shown is the basic PGx experiment and data analysis workflow that was used for the sample data shown on this poster. Other Applied Biosystems™ real time instruments can also be used to generate data for analysis by AlleleTyper™.

Genotyping experiments were performed using custom TaqMan® OpenArray® containing DME assays to major CYP gene variants. Reactions were run on the QuantStudio™ 12K Flex OpenArray® Genotyping System and data was examined using the instrument software. The run data .eds files were imported into TaqMan® Genotyper™ Software and data were analyzed using the Real Time Experiment type and Autocalling method settings. Allele Discrimination plots were reviewed and edited, if needed; then genotype call results were exported in a .txt file.

Copy number experiments were performed by running duplex PCRs containing TaqMan® Copy Number Assays and the TaqMan® Copy Number Reference Assay (Rnase P) on 384-well plates on the QuantStudio 12K Flex Real-Time PCR System. Data were analyzed by the instrument software using manual Ct threshold level of 0.2 and autobaseline settings, to determine Ct values for each assay. Exported results .txt files were then imported into CopyCaller® for copy number analysis by the ΔΔCt method. Copy number results were exported in .txt files.

Genotyper™ and CopyCaller® results files were imported into AlleleTyper™ along with a translation table for translation of sample genotypes to star allele nomenclature calls.

Figure 2. Creating a translation table

Translator Specifications — Bi-Allelic | Mono-Allelic

CYP2D6	CYP2D6_ex9 Hs00010001_cn	2850C>T C__27102425_10	4180G>C C__27102414_10	100C>T C__11484460_40	137_138insT C__32407245_40	883G>C C__30634118_A0	1023C>T C___2222771_40	1707delT C__32407243_20
*1	1	G	C	G	-	C	G	A
*1x2	2	G	C	G	-	C	G	A
*2	1	A	G	G	-	C	G	A
*2x2	2	A	G	G	-	C	G	A
*3	1	G	C	G	-	C	G	A
*4	1	G	G	A	-	C	G	A
*4x2	2	G	G	A	-	C	G	A
*5	0	noamp	noamp	noamp	noamp	noamp	noamp	noamp
*6	1	G		G	-	C	G	-
*7	1	G	C	G	-	C	G	A
*8	1	A	G	G	-	C	G	A
*9	1	G	C	G	-	C	G	A
*10	1	G	G	A	-	C	G	A
*10x2	2	G	G	A	-	C	G	A

Translator Specifications — Bi-Allelic | Mono-Allelic

CYP2D6	CYP2D6_ex9 Hs00010001_cn	2850C>T C__27102425_10	4180G>C C__27102414_10	100C>T C__11484460_40	137_138insT C__32407245_40	883G>C C__30634118_A0	1023C>T C___2222771_40	1707delT C__32407243
*1/*1	2	G/G	C/C	G/G	-/-	C/C	G/G	A/A
*1/*1x2	3	G/G	C/C	G/G	-/-	C/C	G/G	A/A
*1/*2	2	G/A	C/G	G/G	-/-	C/C	G/G	A/A
*1/*2x2	3	G/A	C/G	G/G	-/-	C/C	G/G	A/A
*1/*3	2	G/G	C/C	G/G	-/-	C/C	G/G	A/A
*1/*4	2	G/G	C/G	G/A	-/-	C/C	G/G	A/A
*1/*4x2	3	G/G	C/G	G/A	-/-	C/C	G/G	A/A
*1/*5	1	G/G	C/C	G/G	-/-	C/C	G/G	A/A
*1/*6	2	G/G	C/C;C/G	G/G	-/-	C/C	G/G	A/-
*1/*7	2	G/G	C/C	G/G	-/-	C/C	G/G	A/A
*1/*8	2	G/A	C/G	G/G	-/-	C/C	G/G	A/A
*1/*9	2	G/G	C/C	G/G	-/-	C/C	G/G	A/A
*1/*10	2	G/G	C/G	G/A	-/-	C/C	G/G	A/A
*1/*10x2	3	G/G	C/G	G/A	-/-	C/C	G/G	A/A

Shown is the automatic conversion of a user-defined mono-allelic translation table to a bi-allelic translation table by AlleleTyper™ Software. Portions of the translation tables for CYP2D6 alleles are shown.

Typically, allele haplotype information in public resources such as the Cytochrome P450 Allele Nomenclature or PharmGKB databases is used to create a mono-allelic translation table for the variants of interest in a study; the translation table may contain one or more genes. A monoallelic translation table is first prepared as a .csv file, which contains the star allele haplotype pattern information for each TaqMan® assay in a study.

This file is imported into AlleleTyper™, which automatically converts it to a biallelic translator containing all possible diplotype combinations. The biallelic translator can be exported for review and editing, if needed.

The Mono-Allelic and Bi-Allelic translation specifications are displayed in separate tables.

Note: CYP2D6 is a highly polymorphic gene with >100 described star allele groups including reduced function and non-functional alleles. Individuals may carry null alleles (*5) or extra copies of CYP2D6 (*1, *2, *4, *9,*10, *17, *35). Some CYP2D6 alleles contain sequences derived from the upstream CYP2D7 pseudogene; e.g. CYP2D6*36 has a gene conversion to 2D7 sequences in exon 9 and is a nonfunctional allele.
CYP2D6 copy number assays were used to detect both CNV and hybrid alleles: Hs00010001_cn targets 2D6 exon 9 and will not amplify 2D6/2D7 hybrid alleles having 2D7 exon 9 sequences (e.g. *36), whereas Hs04083572_cn targets CYP2D6 intron 2 and will amplify *36 hybrid alleles.

Figure 3. Results: Summary Report

Results — Summary Report | Detailed Report

Sample ID	CYP2D6	CYP2C9	CYP2C19	CYP2D6_int2 Hs04083572_cn	CYP2D6_ex9 Hs00010001_cn	2D6 -1584C>G C__32407252_30	2D6 100C>T C__11484460_40	2D6 137_138insT C__32407245_40
NA14476	*2/*41	*1/*5	*1/*1	2	2	G/G	G/G	-/-
NA17104	*2/*2x2	*1/*1	*1/*17	3	3	C/C	G/G	-/-
NA17105		*1/*1	*1/*2	3	3	G/G	A/G	-/-
NA17109	*10x2/*36	*1/*1	*2/*17	3	2	G/G	A/A	-/-
NA17112	*1/*1	*1/*1	*1/*1	2	2	G/G	G/G	-/-
NA17114	*1/*5	*1/*1	*1/*1	1	1	G/G	G/G	-/-
NA17117	*4/*10x2, *4x2/*10	*1/*1	UND	3	3	G/G	A/A	-/-
NA17118	*1/*2	*1/*1	*1/*1	2	2	G/G	G/G	-/-
NA17120	*1/*10	*1/*1	*1/*17	2	2	G/G	A/G	-/-
NA17125	*1/*1	*1/*1	*1/*1	2	2	G/G	G/G	-/-
NA17126	*1/*29	*1/*1	*1/*2	2	2	G/G	G/G	-/-
NA17128	*1/*15	*1/*1	*1/*1	2	2	G/G	G/G	-/A
NA17130	*1/*2	*1/*3	*1/*17	2	2	G/C	G/G	-/-
NA17132	*10/*17	*1/*1	*1/*1	2	2	G/G	A/G	-/-
NA17137	*2/*29	*1/*1	*1/*1	2	2	G/G	G/G	-/-
NA17139	*1/*29	*1/*1	*1/*2	2	2	G/G	G/G	-/-
NA17144	*1/*17	*1/*5	*1/*2	2	2	G/G	G/G	-/-
NA17148	*1/*1	*1/*1	*1/*17	2	2	G/G	G/G	-/-

The AlleleTyper™ Software Summary Results table shows Coriell DNA sample star allele genotypes for CYP2D6, CYP2C9 and CYP2C19 gene variants tested with DME assays on custom PGx OpenArray panels and CYP2D6 copy number assays.

Subsequent to creating a biallelic translator and genotyping and/or copy number assay data, in AlleleTyper™:

1. Project - A Project is created and named
2. Setup – the Translation Specifications (the biallelic translator) is imported
3. Data - results export files from TaqMan® Genotyper™ Software and/or CopyCaller® Software are imported
4. Results – Summary and Detailed Results tables are viewed and can be exported

AlleleTyper™ matches the genetic information in the data files to the patterns in the biallelic translator and reports the star allele genotypes determined for each individual.

Figure 4. Results: Detailed Report

Results — Summary Report | Detailed Report

Sample ID	CYP2D6	CYP2C9	CYP2C19	Notes
NA17105	no translation available	*1/*1	*1/*2	1. manual calls involved. Checked genetic pattern of sample: contains *4 & *70 haplotypes but *70 information was not in the translator
NA17109	*10x2/*36	*1/*1	*2/*17	
NA17112	*1/*1	*1/*1	*1/*1	
NA17114	*1/*5	*1/*1	*1/*1	When data is missing for an assay, possible translations are provided
NA17117	*4/*10x2, *4x2/*10	*1/*1	{*1/*1, *1/*5, *5/*5}	1. there are undetermined values for these assays: C__27861810_10.

Sample ID	CYP2D6	Notes
NA17105	*4x2/*70	1. manual calls involved. Added missing *70 haplotype to translator

A portion of the Detailed Results table shows examples of error reporting by the software.

AlleleTyper™ reports errors for samples that are missing data, or have genotype patterns for which there is no translation, etc. In the examples shown:

• NA17105 did not get a CYP2D6 gene translation. A review of the sample genotype data and CYP2D6 star alleles on the CYP Nomenclature web site indicated that *4 and *70 alleles were present but *70 had not been included in the translator. The sample got a *4x2/*70 call when *70 allele was included in a translator.

• NA17117 call was undetermined (UND) for the CYP2C19 gene in the Summary Report. The Detailed Report notes that this is because the call was UND for the 2C19*5 assay (C__27861810_10). AlleleTyper™ used the available genetic information to provide possible translations (in curly brackets) in the Detailed Report).