



PGx Translation Service

From genotypes to phenotypes

Translation algorithm and web application

Evgenia Kartsaki¹, Kleanthi Lakiotaki¹, Alexandros Kanterakis¹,
George P. Patrinos² and George Potamias¹

¹Institute of Computer Science, Foundation for Research & Technology – Hellas, Heraklion,
Crete

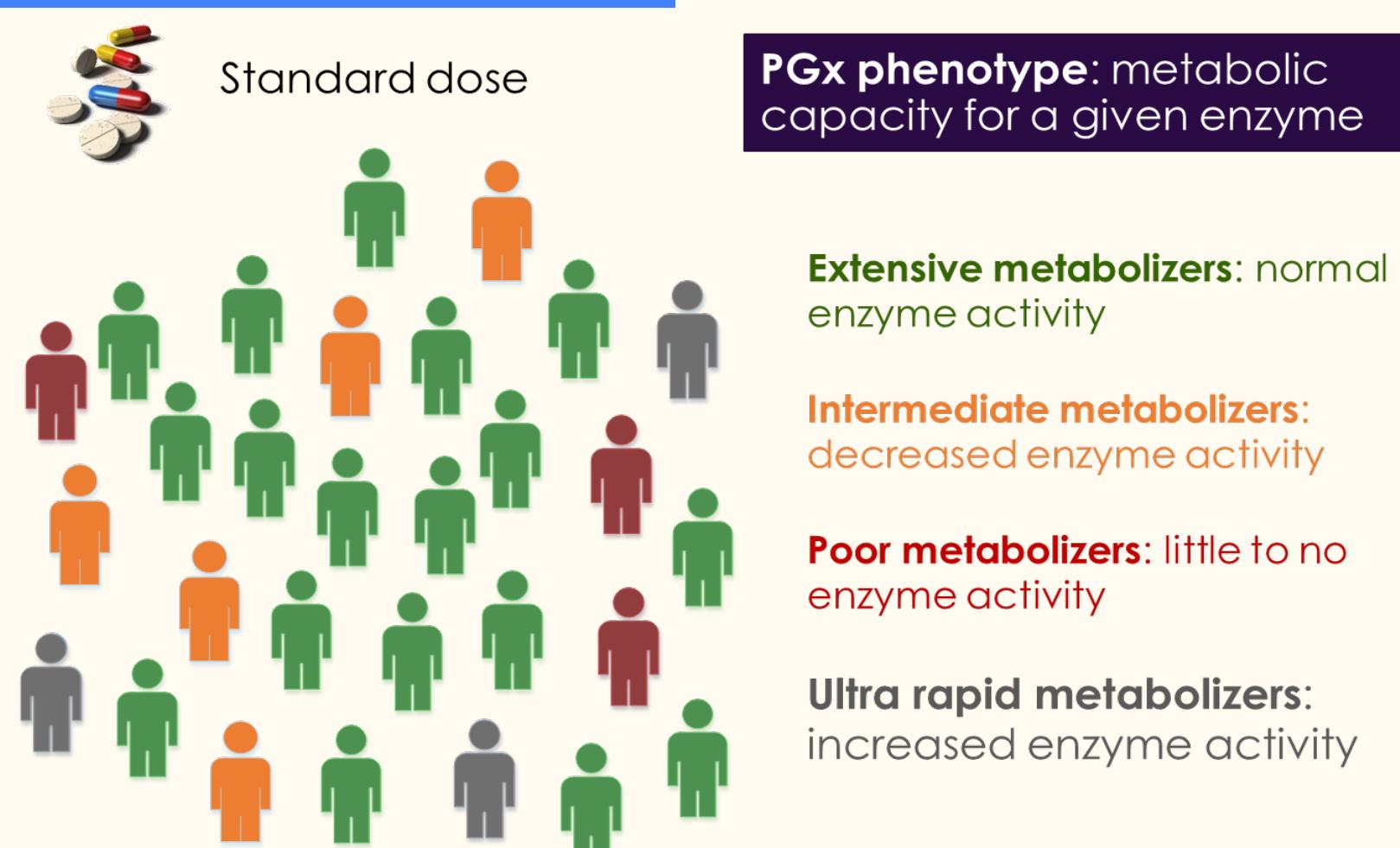
{ekartsak, kliolak, kantale, potamias}@ics.forth.gr

²Department of Pharmacy, University of Patras, Hellas, gpatrinos@upatras.gr





THE CLINICAL PROBLEM



OBJECTIVE. To develop an integrated electronic **Pharmacogenomics (PGx)** web service that provides **personalized genotype-to-phenotype** translation services, linked with **drug recommendations**.

MOTIVATIONS. Drug response varies among individuals, ranging from expected beneficial effects to adverse reactions, and sometimes to even fatal events. Various pharmacogenes relate and affect drug response. Different populations carry different profiles of rare and common gene variants.

→ In this work we study how 69 ADMET (core) genes and their variants (508 SNP biomarkers) affect the PGx **metabolizer status** of phase 3 1000 Genomes (**1kG**) samples across 26 populations following an elaborate PGx **translation** process.

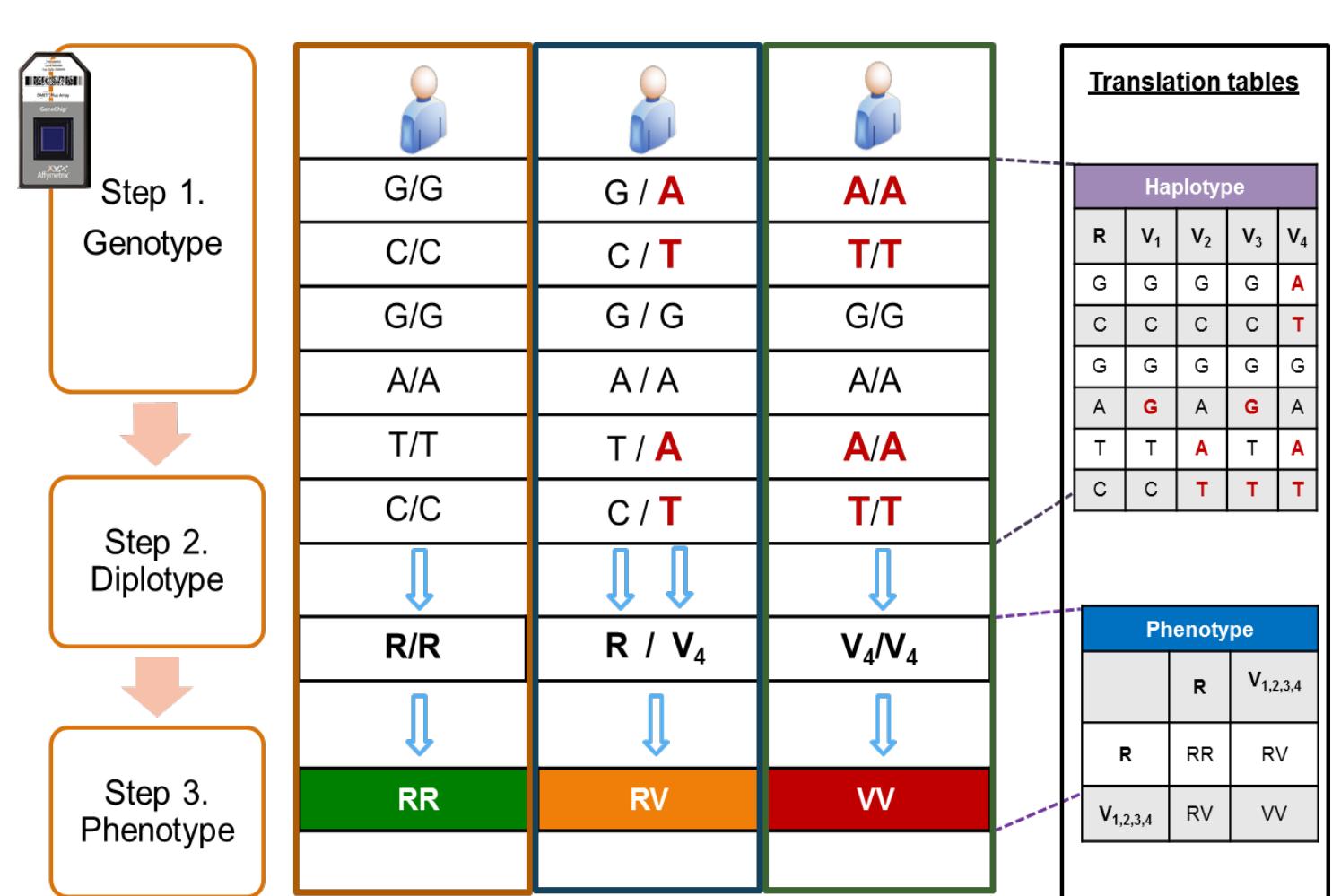
eMoDia: electronic Molecular Diagnostic Assistant

	Personalization	Quality of information	Cost	Updatable
Current scientific Knowledge bases	x	v	v	x
Commercial Direct To Consumer (DTC) companies	v	?	x	x
eMoDia: electronic Molecular Diagnostic Assistant	v	v	v	v

- To integrate heterogeneous PGx information from several valid PGx resources (PharmGKB, dbSNP, Ensembl ...)
- To offer automated personalized PGx translation (genotype-to-phenotype) services, linked to drug recommendations
- To provide a user friendly interface for submitting newly discovered PGx related gene-variants and alleles

PGx Translation Service

METHODOLOGY. We developed an automated **PGx translation algorithm**, which infers metabolic phenotypes from individual genetic (SNP) profiles. For each pharmacogene, and based on available (PharmGKB) haplotype/allele tables, an individual's genotype-profile is **matched** against the available gene-alleles. Next, each inferred diplotype is assigned to a PGx phenotype status, according to available "look up" tables [the algorithm was verified with the Affymetrix® DMET Plus respective translation results]



Glossary.

Pharmacogene - any evidence based gene related to drug Absorption, Distribution, Metabolism, Excretion and Toxicity (ADMET);

Pharmacogenomic marker - any Single Nucleotide Polymorphism (SNP) present in a pharmacogene locus.

Reference/Reference (R/R) - refers to the combination of two reference (wild-type) haplotypes and corresponds to a 'normal' metabolic status.

Reference/Variant (R/V) - refers to the combination of one reference and one variant haplotype (any non wild-type haplotype) and corresponds to an 'intermediate' metabolic status.

Variant/Variant (V/V) - refers to any combination of two variant haplotypes that may lead to either a poor or an ultra-rapid metabolic status; an 'abnormal' metabolic status.

IMPLEMENTATION. The core of the translation process is implemented in the open-source **R** environment (www.r-project.org) and uses R Studio's **Shiny** web-application framework (<http://shiny.rstudio.com/>) to build its web-based interface.

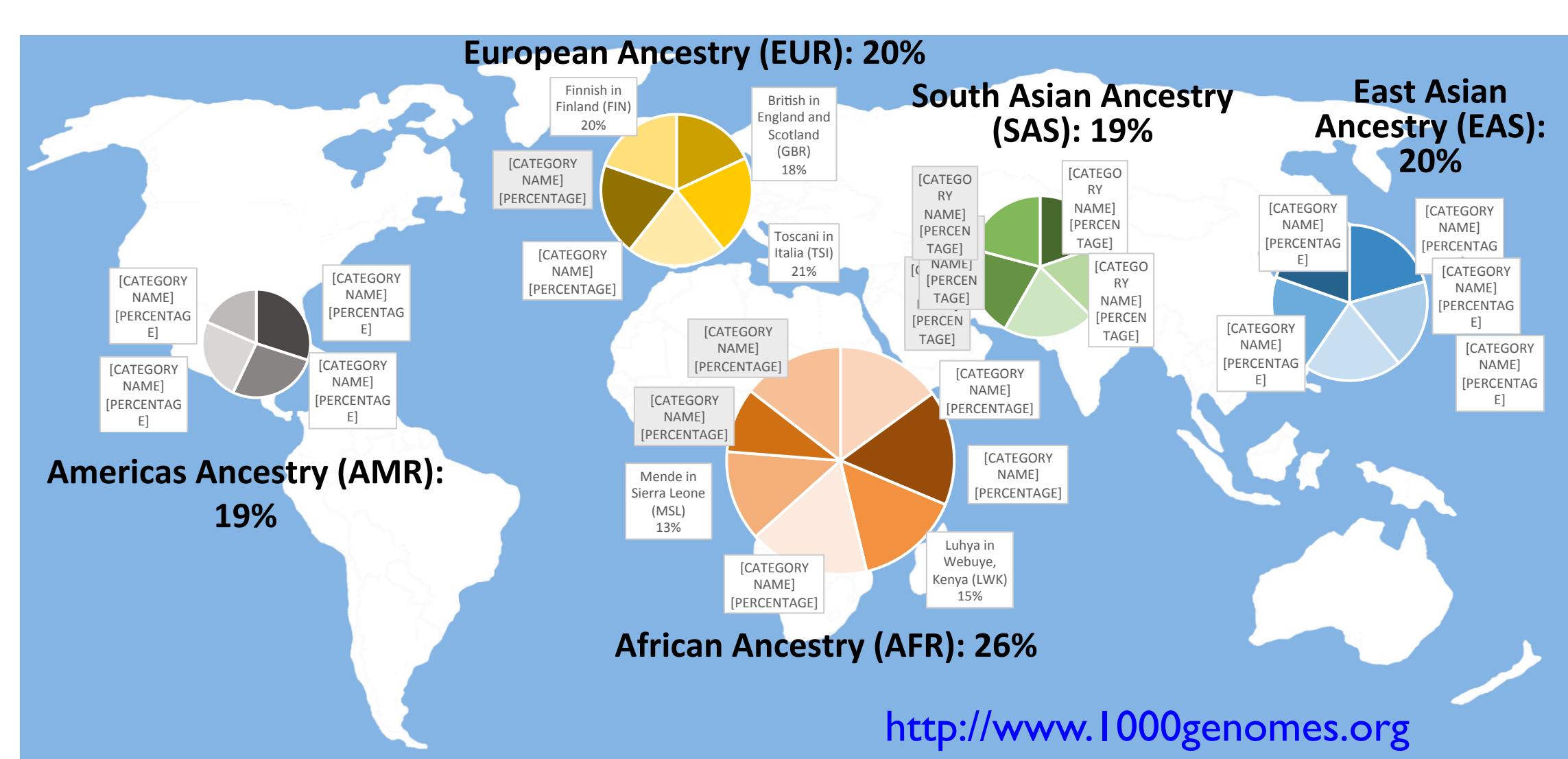
The screenshot shows the eMoDia Shiny application interface. It includes a header with a login field ('Please insert a username: 1KG'), a file upload section ('Please upload a genotype(s) file: Choose File ...'), and a custom translation section ('Please upload a file for Custom Translation: Choose File ...'). The main area displays a table of results with columns: Sample, Gene, Diplootype, Phenotype, and Recommendations. The table lists various individuals with their corresponding genotypes and phenotype status (e.g., RR, RV, VV).

METHODOLOGY. The **explore** functionality couples drugs, genes and diplotypes with personalized recommendations in a simple database schema. Currently eMoDia relies on the PharmGKB dosing guidelines. Health professionals or biomedical researchers browse this information through a lightweight web service. eMoDia shows recommendations that form an expandable tree with user friendly and printable content.

The screenshot shows the eMoDia Explore Service interface. It features a search bar with 'ePGA' and a results section titled 'ePGA electronic PharmacoGenomic Assistance services'. Below the title is a 3D graphic of a DNA helix and colored capsules. Buttons for 'Explore' and 'Translate' are visible.

The screenshot shows the 'Recommendations and Clinical Annotations' section of the eMoDia interface. It displays a hierarchical tree for CYP2D6 amitriptyline (CPIC). The tree includes sections for Source (CPIC), Summary (The CPIC Dosing Guideline for amitriptyline recommends...), Details (Guidelines regarding the use of pharmacogenomic te...), Alleles (Allele: *1/*1XN, Allele: *1/*1I), Metabolizer Status Processed (Poor), Metabolizer Status (Poor metabolizers (~5-10% of patients)), Recommendations (Avoid tricyclic use due to potential for side effe...), Implications (Greatly reduced metabolism of tricyclics to less a...), Phenotype (An individual carrying only non-functional alleles), Activity Score (0), and Clinical Annotations (Allele: *10/*11). A note at the bottom states: 'At the time of the development of this recommendation, there is a lack of strong evidence available on the possible role of CYP2D6 in amitriptyline response in pediatric patient populations; however, there is no reason to suspect that CYP2D6 variant alleles would affect amitriptyline metabolism differently in children compared to adults. Please see below for full details of these guidelines, with supporting evidence and disclaimers.'

POPULATION PHARMACOGENOMICS ANALYSIS ILLUSTRATES THE IMPORTANCE OF STRATIFIED MEDICINE FOR THE SAFER USE OF EXISTING



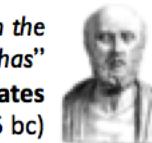
Translation PGx Service : Scope & Targets

Results ★ Welcome ePGA Translation Custom Translation Gene Summary Sample Summary Genotypes

Pharmacogenomics (PGx) Translation Service

As an integral part of Genomic-based Personalized Medicine, Pharmacogenomics (PGx) targets the delineation of the relationship between genetic variation and drug efficacy or toxicity. To date, there are several genes, referred to as pharmacogenes, which play a role in the absorption, distribution, metabolism, excretion and toxicity of several drugs (ADMET)

"It's far more important to know what person the disease has than what disease the person has"



Hippocrates
(c. 460 bc - c. 375 bc)

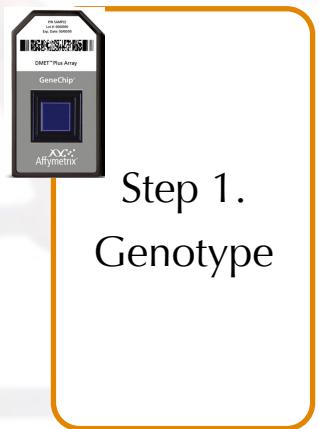
- ❖ The Translation service of the electronic PGx Assistant (ePGA) is meant **SOLELY FOR RESEARCH PURPOSES** [please refer to the respective [Disclaimer](#)].
- ❖ The ePGA/Translation adheres to the emerging trend of *pre-emptive high-throughput genotyping as a vital clinical (diagnostic and prognostic) decision making component* aiming to offer respective genotype-to-phenotype inference services as a mean to translate PGx knowledge from bench-to-bedside.
- ❖ The ePGA/Translation service is based on 'matching' individual genotype (SNP) profiles with PGx gene haplotypes, and the subsequent inference of the corresponding metabolizer phenotypes. Currently ePGA/Translation component employs harmonized haplotypes-tables as registered and curated by [PharmGKB](#) (the most advanced PGx knowledge base).



Adverse drug reactions (ADRs) continue to be a major public health problem. According to the [European Medicines Agency's EudraVigilance](#) database held a total of ~24,000,000 ADRs by 31 December 2013. Moreover, the US/[Centers for Disease Control and Prevention](#), reports that each year in the US 700,000 emergency department visits and 120,000 hospitalizations are due to Adverse Drug Events (ADEs)



Translation: Outline



G/G	G / A	A/A
C/C	C / T	T/T
G/G	G / G	G/G
A/A	A / A	A/A
T/T	T / A	A/A
C/C	C / T	T/T
R/R	R / V ₄	V ₄ /V ₄
RR	RV	VV

Translation tables

Haplotype					
R	V ₁	V ₂	V ₃	V ₄	
G	G	G	G	A	
C	C	C	C	T	
G	G	G	G	G	
A	G	A	G	A	
T	T	A	T	A	
C	C	T	T	T	

Phenotype		
	R	V _{1,2,3,4}
R	RR	RV
V _{1,2,3,4}	RV	VV

Translation Service in general ...



PGx translation service infers phenotypes from individual genotype samples

- Use of two types of PGx translation tables :
 - i. Haplotype tables
 - ii. Phenotype table
- Determine gene-level diplotypes (pair of haplotypes)
- Assign a phenotype to each diplotype
- Form a PGx profile on available pharmacogenes for each genotype sample

Haplotype Tables



- **PharmGKB** haplotype tables on **69** Pharmacogenes
- **718** pharmacovariants with rsID – SNPs or IN/DEls

Gene X	rs12513	rs123512	rs214126	rs10478	130G>C
*1	G	T	C	A	G
*2	G	A	T	A	G
*3	G	A	C	T	G
*4	C	T	C	T	C

- Each row represents a haplotype
- Each column represents a structural variant (SNP or IN/DEL)
- *1 is considered to represent the **reference** (wild-type) haplotype
- Black color → major allele
- Red color → minor allele

Haplotype Tables: Modifications

Identify variants with no rsID and haplotypes that include a minor allele in them

Delete any variants (columns) with no rsID

Delete any haplotype that includes a minor allele in the formerly deleted SNPs

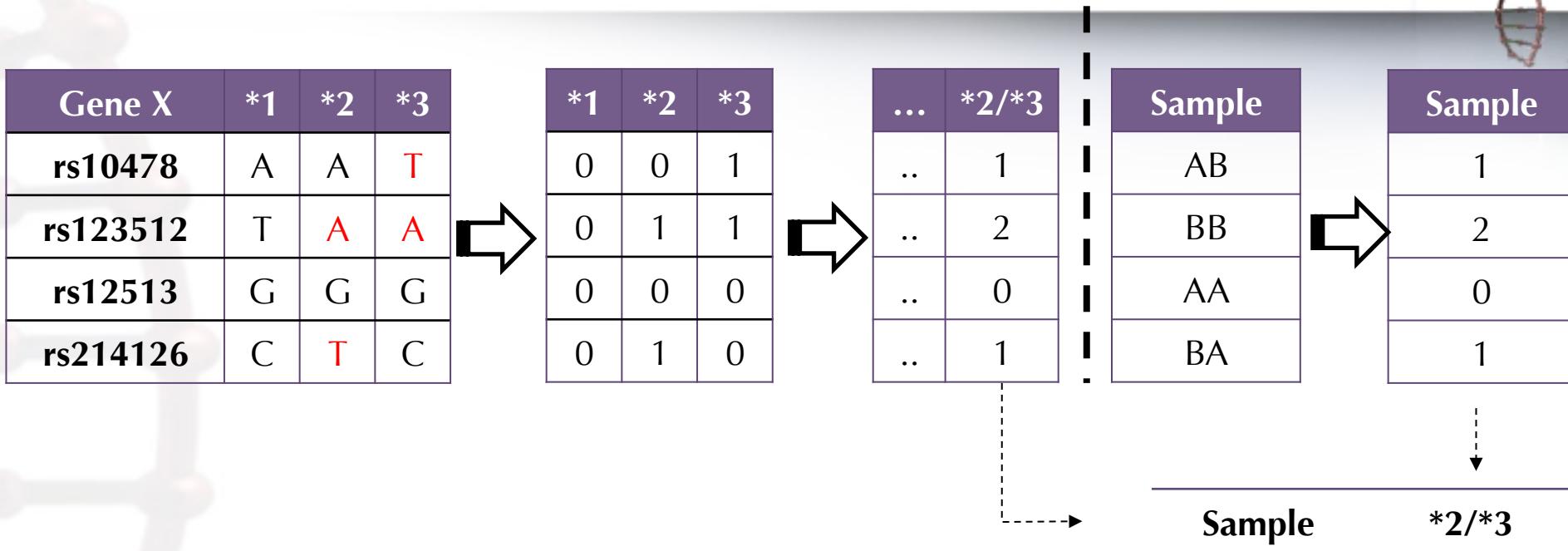
Transpose table and sort rows by rsID

Gene X	*1	*2	*3
rs10478	A	A	T
rs123512	T	A	A
rs12513	G	G	G
rs214126	C	T	C





Binary representation of haplotype tables



Haplotype table's Annotation



Platform's Annotation



Translation of Drug Metabolic Enzyme and Transporter (DMET) Genetic Variants into Star Allele Notation using SAS

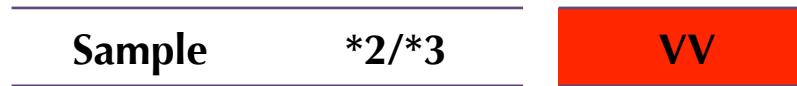
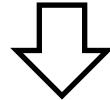
Mark Farmen et al.



Phenotype Inference



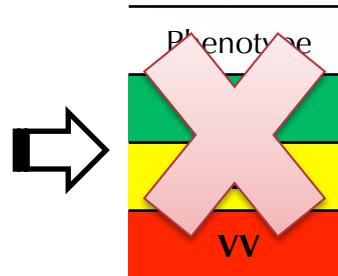
Phenotype Table	
Two reference haplotypes	RR
One reference & One variant haplotype	RV
Two variant haplotypes	VV



Genotype → Phenotype → Recommendations



Gene	Diplotype
CYP2C9	*1/*1
CYP2C19	*1/*2
TPMT	*1S/*1S



Metabolizer Status
Extensive
Intermediate
Poor or Ultra

Find genotype-based dosing recommendation

Pick CYP2C19 alleles: *1 ▾ *2 ▾

Phenotype (Genotype)

An individual carrying one functional allele and one loss-of-function allele

Metabolizer Status

Intermediate metabolizer (~18-45%)

Implications

Reduced metabolism of amitriptyline when compared to extensive metabolizers

Recommendations (Strength: Strong)

Initiate therapy with recommended starting dose.

At the time of the development of this recommendation, there is a lack of strong evidence available on the possible role of CYP2C19 in amitriptyline response in pediatric patient populations; however, there is no reason to suspect that CYP2C19 variant alleles would affect amitriptyline metabolism differently in children compared to adults. Please see below for full details of these guidelines, with supporting evidence and disclaimers.

PharmGKB



Translation PGx Service : Input

- Currently only public genotype sample files (e.g., from 1000 Genomes Project)



Input

Genotype file

1000 Genomes Phase-I (xxxx samples)

... in [VCF](#) format; you may download an [example VCF file](#)

[Get help](#)

File for Customized translation (optional)

[Browse...](#) No file selected.

... list of RS ids, you may download and save an [example file](#) which then you may store and upload

[Get help](#)



Please refer to the [disclaimer](#)



Translation Service : Genotype Profiles

~/R/Translation - Shiny
http://127.0.0.1:6538 | Open in Browser

eMoDia Personalized PGx Translation Service

Please insert your file :

Choose File ...G-EnsemblGenotypes.txt
Upload complete

Please insert a csv file, containing on the first column the rsIDs of your SNPs and on the following columns the genotype(s) of your sample(s). You can download a [sample](#) and then try uploading it.

Summary

Translated Samples : 10
Genes (matched): 72
Haplotypes(matched): 164

PGx Profiles Individuals Genes Genotypes

50 records per page Search:

rs_Id	HG00096	HG00097	HG00099	HG00100	HG00101	HG00102	HG00103	HG00104	HG00106
rs10012	G/G	G/G	G/C	G/G	G/G	C/G	G/G	G/C	G/G
rs10038095	T/A	T/A	T/A	T/T	A/A	T/A	A/T	A/A	T/T
rs1004749	A/C	A/A	A/C	A/A	A/C	A/A	A/A	A/A	A/A
rs10187694	G/G								
rs10264272	C/C								
rs10276036	T/T	C/T	T/T	T/T	C/C	C/C	T/T	T/C	C/C
rs1041983	C/C	C/C	C/T	T/T	T/T	C/C	C/T	T/T	C/C
rs1042597	C/C	C/C	G/C	G/G	G/G	C/G	C/C	C/C	C/C
rs1042605	A/A	A/G	A/A	A/A	A/A	A/A	A/A	A/A	A/G
rs1042640	C/G	C/C	G/C	G/C	C/C	C/C	C/C	C/C	C/C
rs1042713	A/A	A/G	G/G	A/A	G/G	G/A	G/A	G/A	A/A
rs1042714	C/C	C/G	C/G	C/C	C/C	G/C	G/C	C/C	C/C
rs10445704	G/A	A/G	G/A	G/G	G/G	A/G	A/A	A/A	A/G
rs1045642	A/G	A/G	G/G	G/G	A/A	A/A	G/G	A/G	A/A
rs1048943	T/T	T/T	C/T	T/T	T/T	T/T	T/T	T/T	T/T
rs1048977	C/C	C/T	C/T	C/C	T/C	T/C	T/C	T/C	C/T
rs1050828	C/	C/C	C/C	C/C	C/	C/C	C/	C/C	C/C
rs1050829	T/	T/T	T/T	T/T	T/	T/T	T/	T/T	T/T
rs10509681	T/T								



Translation Service : Reporting



~ /R/Translation - Shiny
http://127.0.0.1:6867 | Open in Browser

eMoDiA

Personalized PGx Translation Service

Please insert your file :

Choose File ...G-EnsemblGenotypes.txt
Upload complete

Please insert a csv file, containing on the first column the rsIDs of your SNPs and on the following columns the genotype(s) of your sample(s). You can download a sample and then try uploading it.

Summary

Translated Samples : 10
Genes (matched): 72
Haplotypes (matched): 164

Download

Combination of two reference haplotypes results in R/R phenotype

Sample	Gene	Translation	Phenotype
HG00096	ABCB1	NO_MATCH	
HG00096	ABCC10	NO_MATCH	
HG00096	ABCC2	H1/H15	
HG00096	ABCC2	H2/H12	
HG00096	ABCC2	H9/H14	
HG00096	ABCC2	H10/H13	
HG00096	ADRB1	H2/H2	
HG00096	ADRB2	2/2	
HG00096	APOE	E3/E3	
HG00096	BRCA1	NO_MATCH	
HG00096	CDA	*1A/*1A	
HG00096	CFTR	Ref/Ref	
HG00096	CHRNAS	NO_MATCH	
HG00096	COMT	Haplotype high activity/Haplotype intermediate activity	
HG00096	CYP1A1	*1/*1	
HG00096	CYP1A2	NO_MATCH	
HG00096	CYP1B1	NO_MATCH	
HG00096	CYP2A13	*1/*1	
HG00096	CYP2A6	*18/*18	
HG00096	CYP2A6	*18/*18A	
HG00096	CYP2A6	*18A/*18A	



Translation Service : Sample Summary Statistics

~/R/Translation - Shiny
http://127.0.0.1:6538 | Open in Browser

eMoDiA Personalized PGx Translation Service

Please insert your file :

Choose File ...G-EnsemblGenotypes.txt
Upload complete

Please insert a csv file, containing on the first column the rsids of your SNPs and on the following columns the genotype(s) of your sample(s). You can download a sample and then try uploading it.

Summary

Translated Samples : 10
Genes (matched) : 72
Haplotypes(matched) : 164

Download

PGx Profiles Individuals Genes Genotypes

10 records per page Search:

Samples	EM	IM	PM_or_UM
HG00096	38.59649	28.07018	33.33333
HG00097	41.17647	23.52941	35.29412
HG00099	28.78788	25.75758	45.45455
HG00100	47.36842	15.78947	36.84211
HG00101	44.82759	15.51724	39.65517
HG00102	36.06557	31.14754	32.78689
HG00103	38.33333	10.00000	51.66667
HG00104	50.00000	13.46154	36.53846
HG00106	44.64286	16.07143	39.28571
HG00108	46.26866	16.41791	37.31343

Showing 1 to 10 of 10 entries

← Previous 1 Next →



Translation Service : Gene Summary Statistics

MoDia

Personalized PGx Translation Service

Please insert your file :

Choose File ...G-EnsemblGenotypes.txt
Upload complete

Please insert a csv file, containing on the first column the rsids of your SNPs and on the following columns the genotype(s) of your sample(s). You can download a sample and then try uploading it.

Summary

Translated Samples : 10
Genes (matched) : 72
Haplotypes (matched) : 164

Download

PGx Profiles Individuals Genes Genotypes

25 records per page Search:

Genes	40.000000	20.000000	0	30.000000	10.000000
ABCB1	40.000000	20.000000	0	30.000000	10.000000
ABCC10	0.000000	0.000000	0	100.000000	0.000000
ABCC2	10.526316	36.842111	0	0.000000	52.631579
ADRB1	20.000000	10.000000	0	30.000000	40.000000
ADRB2	0.000000	30.76923	0	0.000000	69.230769
APOE	70.000000	30.000000	0	0.000000	0.000000
BRCA1	30.000000	0.000000	0	70.000000	0.000000
CDA	6.666667	40.000000	0	0.000000	53.333333
CFTR	100.000000	0.000000	0	0.000000	0.000000
CHRNA5	10.000000	10.000000	0	40.000000	40.000000
COMT	0.000000	60.000000	0	20.000000	20.000000
CYP1A1	72.727273	18.18182	0	0.000000	9.090909
CYP1A2	20.000000	0.000000	0	70.000000	10.000000
CYP1B1	18.181818	18.18182	0	27.27273	36.363636
CYP2A13	90.000000	0.000000	0	10.000000	0.000000
CYP2A6	0.000000	0.000000	0	12.500000	87.500000
CYP2B6	0.000000	0.000000	0	75.000000	25.000000
CYP2C19	0.000000	0.000000	0	100.000000	0.000000
CYP2C8	90.000000	10.000000	0	0.000000	0.000000
CYP2C9	90.000000	0.000000	0	10.000000	0.000000

Summary



Translation service :

- Provides inference of PGx phenotypes based on state-of-the-art accumulated PGx knowledge
- Provides PGx clinical-annotations and recommendations
- provides summary statistics on samples and genes

Work in progress...

- Updating mechanisms (new genes, new variants, new Haplotypes/Haplotype tables)
- Use of both genotype and phased data