IDENTIFICATION OF 'PHARMA-SNPS' FOR PREDICTING RESPONSE TO DRUG THERAPIES

MAULANA BACHTIAR

NATIONAL UNIVERSITY OF SINGAPORE

2014

IDENTIFICATION OF 'PHARMA-SNPS' FOR PREDICTING RESPONSE TO DRUG THERAPIES

MAULANA BACHTIAR

B.Sc.(Hons.), NUS

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOCHEMISTRY

YONG LOO LIN SCHOOL OF MEDICINE

NATIONAL UNIVERSITY OF SINGAPORE

2014

DECLARATION

I hereby declare that this thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

_____

Maulana Bachtiar

10 August 2014

ACKNOWLEDGEMENTS

This PhD project had arisen from two personal motives. The first is impact. I am motivated to embark on something that can have an impact to society, particularly in human life and health. Secondly, having grown up in diverse societies, I have always been intrigued by human diversity. This thesis is a good combination of these two interests. It is also possible because of the kind supports from many individuals, most of whom are not addressed here due to space limitation. Despite this, I am very grateful and obliged to acknowledge the following individuals.

Assoc. Prof. Caroline Lee, my highly supportive and humble PhD supervisor. At our first meeting, upon realizing about my non-computational background, one of her initial advises (or even first words) to me was, "One requirement for this project is to have courage; including to not be afraid of computer." Now, she is proven right. It is this never-be-afraid-of-computer attitude that allowed me to continue, strive and complete this project. Thank you Caroline, for the continuous support, guidance, and supervision during my time in your lovely research group.

Assoc. Prof. Teo Yik Ying and **Assoc. Prof. Heng Chew Kiat**, my Thesis Advisory Committee; thanks for your strong support and believe in this project. **Prof. Greg Tucker-Kellogg;** thanks for the coffee and your kindness in teaching me how to script a good 'R' code. **Assoc. Prof. Thilo Hagen**, thanks for supporting and leading the research seminars, outings and other student/department activities without which, our PhD lifes will not be significantly memorable. **Dee, Amanda, Fatin** and the Biochemistry Department team, thanks for your never-ending assistance in school matters. **NUS YLL School of Medicine**, thanks for the academic and scholarship supports. **National Cancer Centre Singapore (NCCS)**, thanks for accommodating most of my research activities.

Past and present LCFG lab members including the following kind individuals (not in any order): Jingbo, Champ, Wei Bing, Steven Wolf, Thomas, Priya, Grace Pang, Jinyu, Steven Theng, Cheryl, Soo Ting, Marcus, Jingli, Yinyee, Gaoyun, Caoyi, Jianwei, and Tuan Tzen. Many thanks to Wang Jingbo and Priya for assistance in computer matters. Also to Champ, Wei Bing, Thomas, and Wolf for adding a lot of colors to my PhD life.

GIV startup team especially **Harry, Tony, Arvel, and Stella**, Thanks for supporting me in completing this study and many other unimaginable works or projects. My friends at NUS, Singapore and beyond; thank you. And to my Family; **Ayah, Ibu, Aam**, *terima kasih banyak atas sandang, papan, pangan, dukungan, bimbingan, love, dan lain-lainnya yang selalu ada kapan saja*.

Allow me to end with a quote that I (accidently) found and remember during my early encounter with computer sometime in the 90's. It was displayed on a desktop's screensaver in my mom's workplace. It says:

"The more I learn, the less I now feel knowledgeable about this place, but the more I now realize of Your greatness, Creator."

PUBLICATIONS AND ABSTRACTS

Published Manuscripts

1. **Maulana Bachtiar** and Caroline G.L Lee; *Genetics of Population Differences in Drug Response*. Current Genetic Medicine Reports 1, 167-170 (September 2013).

2. Steven J. Wolf #, **Maulana Bachtiar#**, Jingbo Wang, Tiow Suan Sim, Samuel S. Chong, Caroline G.L Lee (#both authors contributed equally to this work); *An update on ABCB1 pharmacogenetics – Insights from a 3D model into the location and evolutionary conservation of residues corresponding to SNPs associated with drug pharmacokinetics*. The Pharmacogenomics Journal 11, 315-325 (October 2011). This article was highlighted as the "most accessed article in October 2011" by The Pharmacogenomics Journal.

Manuscript in Preparation for Submission

Maulana Bachtiar and Caroline G.L Lee; *Elucidating Pattern of Population Differentiation and its Potential Functional Effect in the Human Genomes*.

Oral Presentation

Maulana Bachtiar, Steven J. Wolf, Jingbo Wang, Tiow Suan Sim, Samue S. Chong, Caroline G.L Lee, *Evolutionary and structural insights of coding SNPs from 3D model of the ABCB1 multidrug resistance protein*; Oral presentation at School of Medicine Graduate Congress, January 2013, Singapore.

Conference Abstracts

1. **Maulana Bachtiar**, Jingbo Wang, Cynthia Sung and Caroline G.L Lee; *The Pharmacogenomics behind Population Differences of Enalapril Response in Singapore*; Poster presentation at Pacific Symposium on Biocomputing (PSB) 2013, January 2013, Hawaii USA (Awarded Travel Fellowship from NIH/NLM).

2. **Maulana Bachtiar**, Jingbo Wang and Caroline G.L Lee; *Architecture of Single Nucleotide Polymorphisms in Drug Response Pathways*; Poster presentation at Bio-IT World Asia Conference, June 2012, Singapore (Awarded Conference Student Fellowship).

3. **Maulana Bachtiar**, Jingbo Wang and Caroline G.L Lee; *The Global SNP Architecture and population differentiation pattern of drug-response pathways*; Poster presentation at FAOBMB Conference, October 2011, Singapore.

4. Caroline G.L Lee, **Maulana Bachtiar**, Jingbo Wang, Samuel S. Chong; *Architecture and patterns of population differentiation of SNPs in drug response pathways*. European Human Genetics Conference 2011, 28 - 31 May 2011, Amsterdam, Netherlands.

5. **Maulana Bachtiar**, Jingbo Wang, Caroline G.L Lee. *The SNP architecture & population differences in drug-response pathways*. School of Medicine Graduate Congress, January 2011, Singapore (Received best poster award).

Awards Received

1. NUS Research Scholarship, Yong Loo Lin School of Medicine, 2010-2014, NUS.

2. 1st runner up for Oral Presenter at the 6th Biochemistry Student Symposium 2014, YLL School of Medicine, NUS.

3. Selected as 'Leader of Tomorrow' at the GapSummit 2014, University of Cambridge, UK.

4. Third Best Research in Progress Presenter 2012/2013, Department of Biochemistry, YLL School of Medicine, NUS.

5. NIH/NLM Travel Fellowship to Pacific Symposium on Biocomputing (PSB) 2013, USA.

6. Biochemistry Student Travel Fellowship, Yong Loo Lin School of Medicine, 2012, NUS.

7. Bio-IT World Asia Conference Student Fellowship 2012, Singapore.

8. Best Poster Award at Yong Loo Lin School of Medicine Inaugural Graduate Scientific Congress, 2011, NUS.

Table of Contents

SUMMARY

In medicine, one size does not fit all. Population differences in response, including adverse drug reactions (ADRs) are associated with commonly prescribed drugs, ranging from anticancers to hormone therapies. Genetic diversity in the form of Single Nucleotide Polymorphisms (SNPs) has been reported to play a role in this phenomenon. This thesis focuses on identifying SNPs that are extremely population differentiated and seeks to utilize the genes that carry these SNPs for profiling population differentiation of drug response.

Initially, I evaluated the role of coding SNPs in a multidrug resistance protein, the ABCB1. The potential effect of coding SNPs to the homolog 3D structure of ABCB1 was accessed based on residue location and conservation status. Nonetheless, as ABCB1 is not the only protein that plays a role in drug response, I then expanded the study to 750 drug-response genes that are linked to 41 conventional drug pharmacokinetic or pharmacodynamic pathways. The architecture of SNPs in these genes was elucidated and it was discovered that there is an abundant presence of both coding and non-coding genic SNPs. The latter are relatively less studied and can potentially affect gene expression, RNA structure or stability. Moreover, compared to coding SNPs, more of these regulatory SNPs are extremely population-differentiated. Subsequently, the next focus was on SNPs that are extremely population differentiated. However, the HapMap tag-SNPs data, which totaled 1.4 million SNPs, may not be the best representation of all SNPs in the human genome.

Therefore, using the genome-wide SNP allele frequencies from the 1000 Genomes project, I calculated pairwise F_{ST} score of 23 million SNPs. Here, the main focus was to identify top chromosome differentiated SNPs (tcdSNP) in addition to the tcdGenes containing these SNPs. Many pathways that are found to be enriched by tcdGenes include those that are connected with highly variable phenotypes, such as the olfactory transduction, antigen presentation, and immune system-related pathways.

This genome-wide population genetic differentiation data was then integrated with an expanded collection of drug-response genes that are originated from four major pharmacogenomics databases. Gene sets information was available for a total of 1,151 drugs that are approved by the US Food and Drug Administration. Subsequently, tcdGene enrichment analysis with this vast collection of drug-response genes was conducted. With this approach, I identified drug clusters that are associated with strong population genetic differentiation profile between African and other populations, in addition to those that are differentiated between the East Asian and European populations. Whilst the former is dominated by drugs that are associated with the musculo-skeletal system anatomical group, the latter cluster is mostly occupied by nervous system drugs, including psycholeptics.

Finally, the vast knowledge that is generated in this thesis is significant as it can be utilized for the development of drug population-genetic profiling, which can be useful for early prevention of ADR. The novel approach employed in this thesis can be replicated in drug clinical trial. This information is stored in an SQL database and publicly made available through the PharmaSNP web resource that is accessible at http://bit.ly/pharma-snp.

List of Tables

List of Figures

List of Abbreviations

Population Acronyms

ASW	African ancestries from Southwest United States
AF	African population
CEU	Northern and Western European ancestries in Utah, United States
CHB	Han Chinese in Beijing, China
CHD	Chinese in Denver
CHS	Han individuals in Southern China
CLM	Columbian in Medellin, Columbia
EA	East Asian population
EU	European population
FIN	Finish in Finland
GBR	British in England and Scotland, Great Britain
GIH	Gujararti Indian in Houston
IBS	Iberian in Spain
INS	Indian in Singapore
JPT	Japanese individuals in Tokyo, Japan
LA	Latin American population
LWK	Luhya individuals in Kenya
MAS	Malay in Singapore
MKK	Maasai in Kenya
MXL/MEX	Mexicans in Los Angeles, United States
PUR	Puerto Rican in Puerto Rico
TSI	Toscani in Italy
YRI	Yoruba individuals in Nigeria

Amino Acid Single Letter Codes

G	Glycine (Gly)
P	Proline (Pro)
A	Alanine (Ala)
V	Valine (Val)
L	Leucine (Leu)
I	Isoleucine (Ile)
M	Methionine (Met)
C	Cysteine (Cys)
F	Phenylalanine (Phe)
Y	Tyrosine (Tyr)
W	Tryptophan (Trp)
H	Histidine (His)
K	Lysine (Lys)
R	Arginine (Arg)

Q	Glutamine (Gln)
N	Asparagine (Asn)
E	Glutamic Acid (Glu)
D	Aspartic Acid (Asp)
S	Serine (Ser)
T	Threonine (Thr)

Nucleotide Single Letter Codes

A	Adenine
C	Cytosine
G	Guanine
T	Thymine

Other Abbreviations

#ns1	Non-synonymous SNP #1
#s1	Synonymous SNP #1
2D	Two dimensional
3D	Three dimensional
5-FU	5-Fluorouracil
Å	Amstrong unit
ABC	ATP-binding cassette
*ABCB*1	ATP-binding cassette sub-family B member 1
*Abcb*1a	ATP-binding cassette, sub-family B, member 1A
ADME	Absorption, distribution, metabolism, excretion
ADP	Adenosine diphosphate
ADR	Adverse drug reaction
AIDS	Acquired immune deficiency syndrome
ATC	Anatomical Therapeutic Classification
ATP	Adenosine 5'-triphosphate
C. elegans	*Caenorhabditis elegans*
C. l. Familiaris	*Canis lupus familiaris*
CAMs	Cell adhesion molecules
ChEMBL	Large-scale bioactivity database by European Molecular Biology Laboratory
CMS	Content management system
cSNP	Coding SNP
CTD	Comparative Toxicogenomics Databases
C-terminal	Carboxyl-terminal
CYP	Cytochrome P450
D. melanogaster	*Drosophila melanogaster*
dbSNP	The Single Nucleotide Polymorphism Database

DNA	Deoxyribonucleic acid
DPD	Dihydropyrimidine dehydrogenase
DPPR	Differentiated Population Pairs Ratio
DPYD	Dihydropyrimidine dehydrogenase
DRP	Drug-response Pathways
E1	Exon 1
E1/1000A>T	Substitution of A to T at nucleotide position 1000 of Exon 1
EGFR	Epidermal growth factor receptor
eQTL	Expression quantitative loci
ESE	Exonic splicing enhancer
ESE/S	Exonic splicing enhancer/silencer
ESS	Exonic splicing silencer
FDA	Food and Drug Administration
FDR	False discovery rate
Fig.	Figure
F_{ST}	Fixation index
G. gallus	*Gallus gallus*
G1000V	Substitution of Glycine to Valine at residue 1000
G6PD	Glucose-6-phosphate dehydrogenase
GEO	Gene Expression Omnibus
GO	Gene Ontology
HapMap	Haplotype Map
ID	Identification
IPA	Ingenuity Pathway Analysis
ISRE	Intronic Splicing Regulatory Element
KEGG	Kyoto Encyclopedia of Genes and Genomes
LCFG	Liver Cancer Functional Genomics
LCL	Lymphoblastoid cell line
MDR	Multidrug resistance
MDR1	Multidrug resistance protein 1
Mdr50	Multi drug resistance 50
MHC	Major Histocompatibility Complex
miRBS	miRNA binding site
miRNA	Mirco RNA
mRNA	Messenger RNA
RNA	Ribonucleic Acid
NBD	Nucleotide-binding domain
NCBI	National Center for Biotechnology Information
NMD	Nonsense-mediated decay
nsSNP	Non-synonymous SNP
NUS	National University of Singapore
P. falciparum	Plasmodium falciparum
Pc value	Percentile value

PD	Pharmacodynamics
PDB	Protein Data Bank
*PFMDR*1	*Plasmodium falciparum* multidrug resistance gene
pfSNP	Potentially Functional SNP
Pgp	P-glycoprotein
pgp-1	P-glycoprotein-related protein 1
PGP18	ABC transporter B family member 18
PharmGKB	Pharmacogenomics Knowledge Base
PHP	Hypertext Preprocessor
PK	Pharmacokinetics
pmd1	Leptomycin efflux transporter Pmd1
PTM	Post-translational modification
PXR	Pregnane-X-receptor
Py score	Pathway prioritization score
RNA	Ribonucleic acid
rs	Reference SNP
rSNP	Regulatory SNP
S. aureus	*Staphylococcus aureus*
S. pombe	*Schizosaccharomyces pombe*
Sav1866	*Staphylococcus aureus* Sav1866
SGVP	Singapore Genome Variation Project
*SLCO1B*1	Solute carrier organic anion transporter family, member 1B1
SNP	Single Nucleotide Polymorphism
SQL	Structured Query Language
srSNP	Structural-RNA SNP
sSNP	Synonymous SNP
tcdGene	Top chromosome differentiated gene
tcdSNP	Top chromosome differentiated SNP
TF	Transcription factor
TFBS	Transcription factor binding site
TMD	Transmembrane domain
TSI	Toscani in Italy
US	United States
UTR	Untranslated region
*VKORC*1	Vitamin K epoxide reductase complex, subunit 1
WHO	World Health Organization

Chapter 1. General Introduction

Part of this chapter is adapted from

Maulana Bachtiar and Caroline G. L. Lee, ***Current Genetic Medicine Report****, 2013, 1(3):162-170*

This thesis focuses on the study of human genetics diversity that could potentially affect population differences in drug response. Drug response in this case, is the outcome, whether positive or negative, that an individual exhibits after consuming medicine prescribed by a doctor.

1.1 One size does not fit all

Human diversity, particularly one that arises as a result of genetic differences, is manifested at various levels. This includes differential response to xenobiotic or external compounds such as drugs, which are not inherently produced and found in the human body. The long term vision for developing a personalized approach to medicine is to tailor drug prescription according to individual's need so as to optimize outcome and prevent potential adverse drug reaction (ADR) [1].

ADRs cases are not uncommon and do not benefit the patient as not only do they worsen the clinical outcome, frequent ADR occurrence can translate to higher hospitalization cost and is a public health concern [2-5]. In the United States, hospitalization cost as a result of ADR-related cases had reached $136 billion in 2012 [3].

Moreover, variation in the susceptibility of ADR in different populations can also affect drug development. At often times, clinical trial is performed in a country that has a different population profile to the final destination of the drug. These trials may employ subjects who have significant differences in genetic background and cultural behavior. Due to such approach in drug testing, it is therefore unsurprising to observe unexpected ADR cases of drug that was observed to be highly effective during trials. Ideally, a drug should always be initially tested in individuals who have similar background to the intended population where it is going to be marketed. This pushes for a greater involvement of a more diverse clinical trial subjects. However, most of the clinical trials are still conducted in developed economies, despite a growing number of clinical trials that are now organized in developing countries [6, 7].

One reason for this uneven global distribution in drug trial is the extraordinary cost that is associated with organizing clinical trials in multiple geographic regions. As complexity arises, so does the cost in developing the drug, which will have a direct impact to the end consumers, the patients. Therefore, diversity in population profiles can be one of the biggest hurdles in the production of effective medicine [8]. However, if a drug trial is conducted only with one population few could foresee the potential occurrence of side effects in another population. At the same time, there is also a concern that when an ADR is observed in one or more populations within a complex trial, more potentially effective drug could be terminated in the early phase of development.

One potential solution is to identify the population that is more susceptible to

ADR early in the drug development phase. Therefore, ADR can be prevented even during clinical trials. In this regard, drug developers would also be saved from potential problem that would have surfaced in the early phase of a drug development pipeline such as those that are associated with "bad drug" candidates. At the same time, they can also expect to expedite approval of potential block bluster drugs that will be beneficial for the correct audience.

Hence in drug development, the big question is: what if we can prioritize trials in subject population that are identified to be less susceptible to drug toxicity? And what if we can identify drugs that are more population-specific, hence when it induces toxicity in one population; we do not have to recall the drug for the other population? What is the best way to identify the factor that induces population differences in the first place?

Motivated by these questions, my PhD thesis focuses on addressing population differences of drug response and attempted to attack the two above challenges using novel pharmacogenomic approaches, for a better medicine. The method that is developed throughout the course of this study is aimed to serve as a reliable proxy in determining the important factors that account for population differences in treatment response.

1.2 Population Differentiation and Drug Response

Medicine is the epicenter of healthcare. However, population differences in the way patients' response to medication could complicate treatment and elevate the

cost associated with healthcare. This problem becomes more significant in especially two scenarios. The first is in countries with a strong degree of demographic background variation, such as in places that are populated by multiracial demographics carrying significant genetic and lifestyle differences. In this case, no same treatment can be applied to all individuals and tailoring of dug dosage may become a necessity. The other scenario involved drugs that had passed clinical trials in country(s) that has a different population profile to the intended market.

Part of the reason for treatment complication is due to unforeseen ADR occurrence, as a result of variation in a drug response profile in different individuals. These inter-individual differences in drug response are often manifested at the population level, where it has been frequently reported that the outcome of treatment could be manifested differently in individuals belonging to different race or population background [9-12]. It is a common phenomenon and can significantly affect the outcome of treatment including those reported in anticancers, anticoagulation and beta-blockers [13-16]. Population differences in ADR is especially significant if such cases can lead to fatal consequences [17], hence it is important to be able to identify the group of patients who are more susceptible to this negative side effect of a drug.

Furthermore, during literature preview, it can also be observed that there are great number of studies reporting population differences between the European and other populations, particularly Asians or Africans (Fig. 1.1). It is believed that there are more studies that were conducted in these population groups, owing to

the possibly bigger budget and accessibility of clinical trials information in these populations. In the course of this thesis, I had the opportunity to review some of the most highly-studied drugs that are associated with population differences, which is summarized in a graphic format (Fig 1.1).

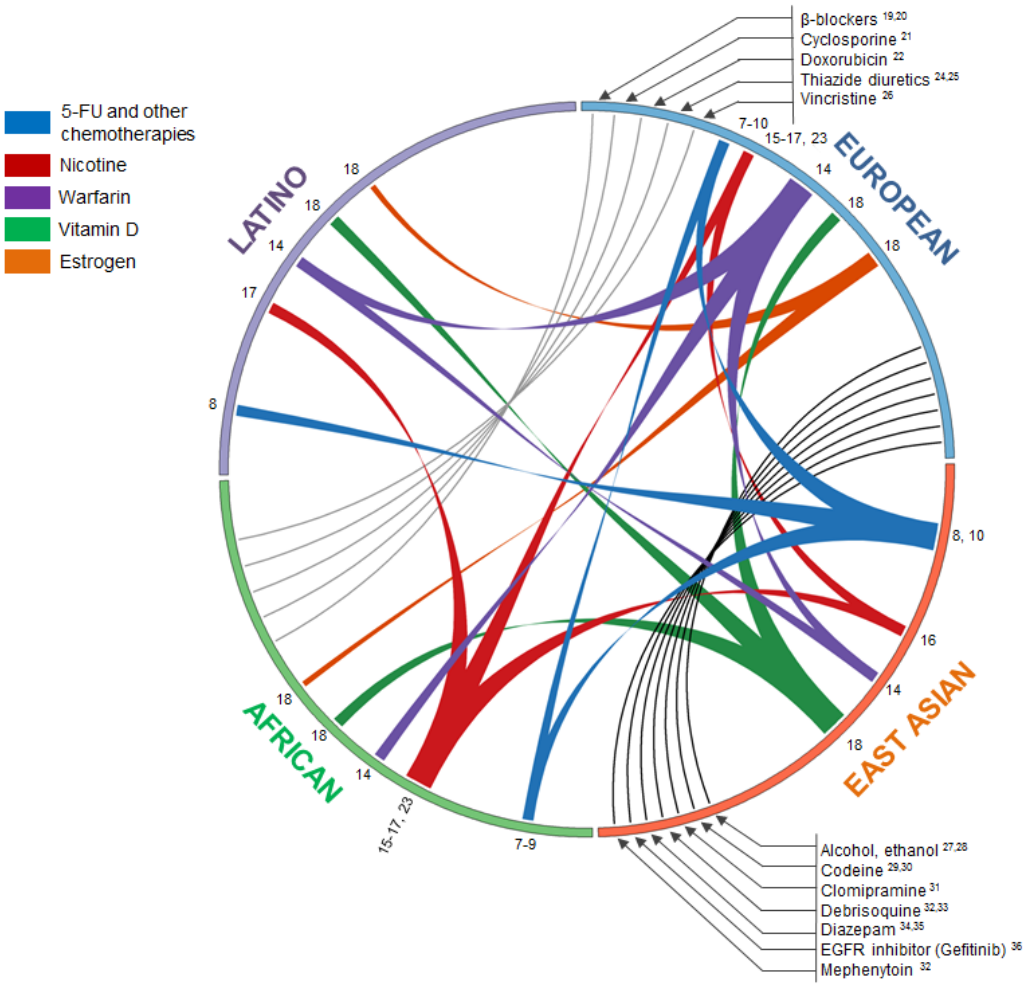


Figure 1.1 Circular plot showing population differences in response to commonly used drugs. The drugs shown are reported to have a different response profile in two or more populations. A drug response difference is shown by ribbon joining at least two populations. The number of population differences that are involved in a drug response profile can be judged by the thickness of the root of the ribbon. Drugs that are reported to have only two population differences are shown by thin black/grey lines. References are marked by the reference number as published in *Maulana Bachtiar and Caroline G. L. Lee, Current Genetic Medicine Report, 2013, 1(3):162-170.*

Frequently cited to have differences in drug response in different populations, 5-Fluorouracil (5-FU) is a widely-prescribed chemotherapy in the treatment of various cancers. 5-FU, which is a fluoropyrimidine-based drug that is known to be associated with hematologic toxicities such as leukopenia and anemia, has been

reported to have outcome differences between Europeans and either African Americans or East Asians [18-21]. The East Asians were also reported to have a different response compared to the Latino or African descents [19]. Individuals with a deficient dihydropyrimidine dehydrogenase (*DPD/DPYD*) were reported to be more susceptible to 5-FU-induced hematologic toxicities [22].

Another popular drug that exerts population differences in response is warfarin, an anticoagulant used to prevent thrombosis and embolism. Patients who are more susceptible to ADR are at higher risk of bleeding. Warfarin treatment is best done through proper optimization of dosage so that efficacy and patient's safety can be ensured. Furthermore, variation in warfarin response is observed in patients of different ethnicities and has been reported albeit not much appreciated [11]. East Asians and either European or Latino populations are known to have significant difference in optimal warfarin dose, in addition to what is observed between the Europeans and either Latinos or Africans [13] (Fig. 1.1, purple ribbons).

Besides 5-FU and Warfarin, another 'drug' that exhibits population difference in response is nicotine. Perez-Stable et al argue that ethnic differences in nicotine response could have accounted for the observation of ethnic differences in tobacco-related disease [23]. The body eliminates nicotine by initially metabolizing it into cotinine. Genes that are involved in this metabolism pathway is known to carry genetic variations, which is a possible factor of addiction to nicotine and tobacco-associated diseases [24]. It has been reported that East Asians and Europeans or Africans, in addition to those between the African and Latino populations, do have significant variation in their serum cotinine levels

[24, 25] (Fig. 1.1, red ribbons).

Reports on ethnic or population differences in the response of other drugs such as codeine, vincristine and b-blockers were also observed, particularly between European and the Asian or African populations. Nonetheless, these reports that are briefly reviewed in this thesis probably over-represent those cases that are seen in developed countries while not highlighting the cases in developing economies. Lastly, Appendix 6 provides a non-exhaustive list of studies describing differential response to various drug treaments.

1.3 Genetic Basis behind Population Differences in Drug Response

Based on the assumption that any two individuals from the same population has a higher chance to be more similar than those from different population backgrounds, 'population differences' can been used as a proxy of 'interindividual differences'. As introduced above, population differentiation is a phenomenon that shall not be neglected in medicine as it is a factor that can significantly affect outcome. Ideally, no two patients are to be given equal treatment as they may respond differently. However economically, individual tailoring of treatments is extremely expensive and requires extensive investigation on patient's background hence it is not a standard practice today.

Drug or xenobiotic response is highly influenced by a number of factors which can be broadly categorized into inherent (which includes but is not restricted to

genetic predisposition, disease status, age, and weight) and external (such as environment condition, socioeconomic status, and education level) factors. Out of all these variables that may affect a person's response to medication, the genetic factor is the most prevalent variable. Hypothetically, with the exception of somatic mutations, most if not all of our genetic material will stay the same from birth to adulthood. For this reason, studying the genetic components that are affecting drug response is the most feasible and is the primary driving force of pharmacogenomics, the field of study involving genes and gene variants that are important in drug-response pathways.

1.3.1 Single Nucleotide Polymorphisms (SNPs)

Single Nucleotide Polymorphisms (SNPs) are the most abundant form of genetic variants in humans, with up to 38 million validated reference SNPs that are recorded in the dbSNP (build 137) database [26]. To be classified as SNP, a genetic mutation has to be observed in a relatively high frequency (usually more than one percent in a particular population). The frequency elevation of these mutations is most probably due to natural pressures that had positively selected these mutations due to them producing survival or reproductive advantages to the organism.

SNPs residing within genes could potentially affect gene function, depending on the location of the sequence variation. Coding region SNPs, particularly those that are associated with substitution in a protein amino acid sequence, could potentially pose a functional effect to a protein structure or post-translational

modification activity [27, 28] (Fig. 1.2). Moreover, when residing in transcription binding site, miRNA binding site, or exon/intron splicing regulatory site within gene regulatory regions, SNPs could pose functional changes to gene expression [29, 30].

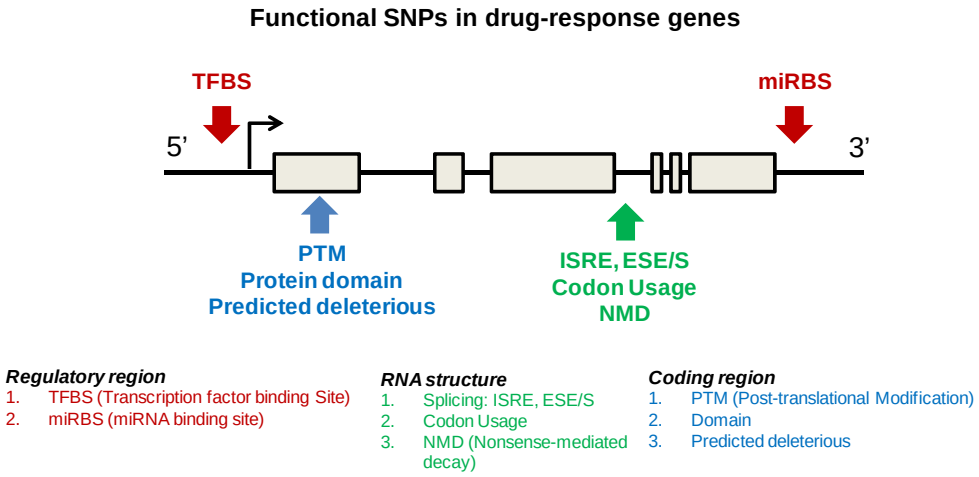


Figure 1.2 SNPs have potential implication for gene function. Those residing in regulatory region such as transcription factor binding site (TFBS), miRNA binding site (miRBS), intronic splicing regulatory elements (ISRE), exonic splicing enhancer/silencer (ESE/S), may affect gene expression. On the other hand, coding SNPs may affect protein structure or post-translational modification (PTM), which may be important to protein function.

Due to its abundance, SNPs play a significant role in the manifestation of population diversity in the human genome. Furthermore, the different environmental factors across various geographic regions would have acted as selection force throughout human migration and evolution, which could negatively or positively select SNPs that are associated with disadvantegous or advantageous traits, respectively [39]. Theoretically, these selection forces leave a "genomic signature" that can be measured by studying SNPs differentiation

pattern in the genome [39]. One way to estimate genomic differentiation between populations is by measuring genetic diversity. This thesis utilizes F_{ST} statistics in estimating population differentiation, which is a measurement that takes into account the reduction of heterozygosity in two or more sub-populations, compared to the total population heterozygosity level [40].

Traditionally, SNP frequencies were obtained by genotyping a targeted DNA region that has been previously sequenced. If one is to conduct this over the entire genome, not only that it will be a very slow process, the approach could only give us a small amount of data that does not represent the overall big picture of variants in the genome. Moreover, this candidate-based approach required one to develop a prior hypothesis in localizing the region where the SNP resides, hence limiting the possibility of finding SNPs that reside in genomic regions that were thought to be of less significance. The advent of high thoroughput genotyping SNP technology had also enabled scientists to genotype many individuals over a relatively shorter period. It is for this reason that the HapMap project was initiated, in which a selection of representative 'tag' SNPs were genotyped to create a haplotype map of the genome [31, 32]. This created a new foundation for pharmacogenomics, which can particularly serve as a referral point in association studies that involve diverse population backgrounds.

1.3.2 Pharmacogenomics

Pharmacogenomics is an extension of pharmacogenetics. Whilst the lalter usually refers to a study involving a select few SNPs or genes, pharmacogenomics

employed a more genome-wide perspective in identifying genetic markers that are important in drug response. In addition, since drug response is highly affected by group genes in the pharmacokinetic (PK) and pharmacodynamic (PD) pathways (Fig. 1.3), identifying the genes in these pathways would hypothetically give a good direction in pin pointing to the SNPs that can be associated with drug response differences.

Following completion of the human genome sequencing, more genes were identified to be a component of drug-response PK and PD pathways. These genes range from membrane transporters, metabolizers, and drug-target genes that are part of a drug PK and PD process. Furthermore, next-generation sequencing technology had propelled genomic studies into a higher level. Genomic variants including SNPs can be identified by comparison of individuals' genomic sequence, even without the initial mapping of these variants into gene regions [33]. Despite its relatively higher initial cost, this technology allows one to start with a hypothesis-free approach of identifying SNPs that could be population-differentiated. These are the SNPs that have significantly different distribution of allele frequency in different populations.

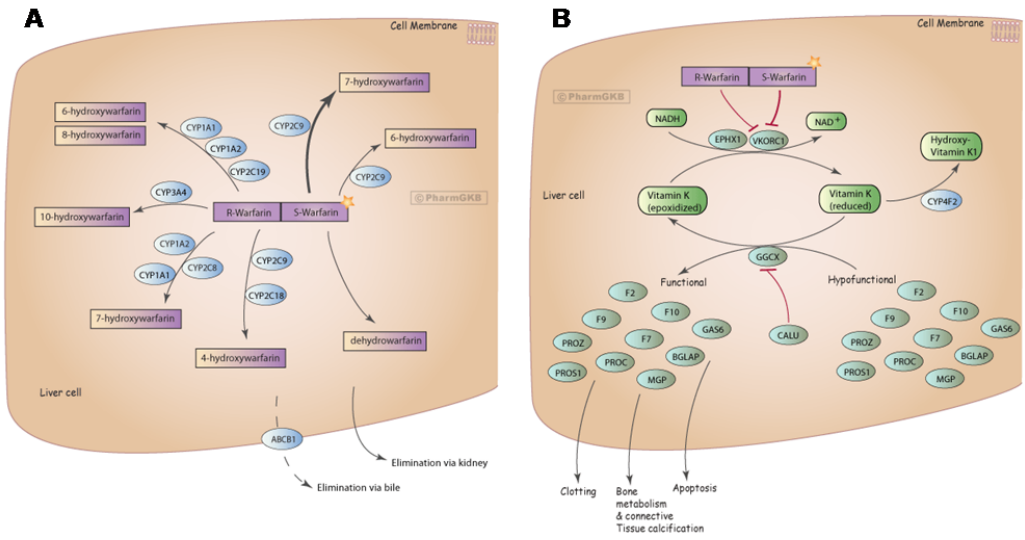


Figure 1.3 Genes in the Pharmacokinetic (A) and Pharmacodynamic (B) pathways that are associated with Warfarin. Genes are represented inside oval shape containers whilst drugs and drug metabolites are contained inside purple squares. Drug of significant interest is marked with a star. Biological intermediates are contained within green colour capsule boxes. In this diagram, a thick arrow indicates a primary route, whereas a thin arrow represents a more generic route and the dashed arrow indicates a secondary route. A red line indicates a repression phenomenon. This image is reproduced from PharmGKB [34].

In 2007, the US Food and Drugs Administration decided to update the labeling for the anticoagulant, warfarin. The new label would put dosing recomendation based on the genotype of SNPs residing in the *VKORC*1 and *CYP2C*9 genes (Fig. 1.3), which are associated with warfarin pharmacokinetics [13, 27]. Warfarin is one of the best examples of a successful implementation of pharmacogenomics in the clinics.

Besides the *VKORC*1 and *CYP2C*9, several other well-reported pharmacogenes that are potentially associated with clinical impact include the *ABCB*1, *CYP3A*4, *DPYD*, *G6PD*, and *SLCO1B*1 [34]. A more exhaustive list of well-known

pharmacogenomic associations is accessible at the PharmGKB website (www.pharmgkb.org/search/knownPairs.action).

Nonetheles, translational application of pharmacogenomics in most if not all drugs are still far away from realization due to the presence of multifactorial components. Currently, one challenge is to find the SNPs that are not only associated with population differences, but also the ones affect gene functions and drug response. Limitations exist particularly when it comes to applying newly discovered pharmacogenomics knowledge to clinics, which is particularly associated with the complexity of drug reponse phenotypes in general.

The vision is to be able to use these genomic variants as predictors to an individual's drug response hence helping us to identify patients who are more susceptible to ADRs. Going forward, pharmacogenomics studies would generally aim to translate more newly discovered knowledge into clinical application. In this thesis, for the purpose of consistency, I will adopt the term 'pharmacogenomics' when referring to any study that involves the genetic basis of drug response, regardless of its genomic scale.

1.4 The Era of Big Data

Today, not only has advancement in next-generation sequencing technology provided an abundant wealth of new knowledge, it also poses a new challenge on how to systematically catalog, mine and more importantly, study the massive amount of data. In the past years, there are efforts in having these data accessible

to the general public. Several early initiatives have embarked to provide population genetics data, ranging from one that genotyped SNPs in specialized candidate genes to one that utilized the tag-SNP approach but in a different set of populations.

The International HapMap Project initially provided genotype data of individuals from four representative populations: CEU (Caucasian from Utah, USA), CHB (Han Chinese from Beijing, China), JPT (Japanese from Tokyo, Japan), and YRI (Yoruban from Ibadan, Nigeria) [35]. It then expanded into genotyping seven additional populations, covering 1.4 million SNPs that were genotyped using the 'tag' or regional representative SNP coverage method [36]. As it is not feasible to genotype all individuals at one go, the HapMap approach of genotyping individual samples from representative of the diverse world populations was seen to be a more viable way in obtaining reference allele frequencies. Indeed, the HapMap is a good source of reference for allele frequency data. Moreover, it inspired the trend of studying global population genetics pattern and its relationship with human migration phenomenon.

The Environmental Genome Project made available the genotype data of SNPs residing in genes that 'interact' with the environment. These genes have role in the DNA repair, metabolism, oxidative stress, apoptosis and several additional pathways that are important in the immune system [37]. On the other hand, the Singapore Genome Variation Project provided the genotype data of one million SNPs from representative individuals who originated from three different racial groups that are defined in Singapore: the Chinese, Indian and Malay [38].

These initial efforts however, are faced by several limitations, including the unavailability of genome-wide data that can allow one to gauge the 'true strength' of population differentiation in a drug response gene, compared to other background genes. Moreover, the tag SNPs approach that are adopted by the HapMap project could have had various 'genomic blind spots' that are attributed for the lack of information that can represent many regions of the human genome.

Nonetheless, the advent of next-generation sequencing technology, which allowed high-throughput sequencing, greatly expanded the coverage to a whole-genome level, in addition to mining sequence information from more individuals. By capitalizing this high-throughput platform, the 1000 Genomes Project was empowered to produce the whole-genome sequence data of 1,092 individuals who originated from the same population groups as those of the HapMap project [33]. Phase 1 of the project has resulted in the identification of 38 million SNPs in addition to insertion-deletions and structural variants that could have not been previously identified using the tag-SNP approach [33]. These SNPs, which have been deposited in dbSNP, are extremely useful in pharmacogenomics, in which it is now possible to identify the whole-genome SNPs that are population-differentiated, including those that are important in drug response.

1.5 General Hypothesis and Aim

My work was conducted based on the notion that population-differentiated SNPs are one of the determinants of population differences in drug response. This knowledge can serve to empower tomorrow's medicine, particularly in advancing the technology that supports personalized therapy. The general hypothesis is that these SNPs could affect drug response variation in individuals originating from different background. Having said that however, this thesis does not come with an ideal promise that by the end of the last chapter, there will be a direct application of personalized medicine. Instead, I aimed to accumulate, integrate, elucidate, and package this novel genomic information into an accessible format for determining drugs population genetic differentiation profile.

1.6 The Thesis Structure

This thesis is divided into two broad categories. The initial part, which is covered by chapter 2 and 3, presented an attempt to understand the pattern of SNPs in drug-response genes. In chapter 2, I presented a detailed study of SNPs that affect the ABCB1 protein, an evolutionary conserved yet highly polymorphic membrane transporter that is associated with multidrug resistance phenomenon. Because drug response is affected by multiple genes, in chapter 3, I then studied the general architecture of SNPs in 715 genes important for drug response, which not only covered coding SNPs, but also SNPs residing in non-coding regions.

The raw data in chapter 3 was derived from the HapMap, which is based on microarray technology. As next-generation sequencing technology emerged, I then attempted to perform a whole genome calculation of population differentiation scores to identify SNPs that could possibly be significant in manifesting population differences in drug response. This is the focus of chapter 4, where by utilizing data from the 1000 Genomes project, SNPs that are population-differentiated were identified. With this result, in chapter 5, I then specifically utilized the population-differentiated SNPs to study population genetic differentiation pattern of more than a thousand drugs. This 'PharmaSNP' information is then made available in an online resource described in chapter 6.

1.7 Reference

1. Zhang, W., Y. Zheng, and L. Hou, *Pharmacogenomic Discovery Delineating the Genetic Basis of Drug Response.* Current genetic medicine reports, 2013. **1**(3): p. 143-149.
2. Classen, D.C., et al., *Adverse drug events in hospitalized patients. Excess length of stay, extra costs, and attributable mortality.* JAMA, 1997. **277**(4): p. 301-6.
3. Bond, C.A. and C.L. Raehl, *Adverse drug reactions in United States hospitals.* Pharmacotherapy, 2006. **26**(5): p. 601-8.
4. Bond, C.A. and C.L. Raehl, *Clinical pharmacy services, pharmacy staffing, and adverse drug reactions in United States hospitals.* Pharmacotherapy, 2006. **26**(6): p. 735-47.
5. Johnson, J.A. and J.L. Bootman, *Drug-related morbidity and mortality. A cost-of-illness model.* Arch Intern Med, 1995. **155**(18): p. 1949-56.
6. Thiers, F.A., A.J. Sinskey, and E.R. Berndt, *Trends in the globalization of clinical trials.* Nat Rev Drug Discov, 2008. **7**(1): p. 13-14.
7. Karlberg, J.P.E., *Globalization of sponsored clinical trials.* Nat Rev Drug Discov, 2008. **7**(5): p. 458-458.
8. Allison, M., *Reinventing clinical trials.* Nature biotechnology, 2012. **30**(1): p. 41-9.
9. Yue, Q.Y., et al., *Interindividual and interethnic differences in the demethylation and glucuronidation of codeine.* British journal of clinical pharmacology, 1989. **28**(6): p. 629-37.
10. O'Donnell, P.H., et al., *Population differences in platinum toxicity as a means to identify novel genetic susceptibility variants.* Pharmacogenet Genomics, 2010. **20**(5): p. 327-37.
11. Johnson, J.A., *Ethnic differences in cardiovascular drug response: potential contribution of pharmacogenetics.* Circulation, 2008. **118**(13): p. 1383-93.

12. Dohmen, K., et al., *Ethnic differences in gastric sigma-alcohol dehydrogenase activity and ethanol first-pass metabolism.* Alcoholism, clinical and experimental research, 1996. **20**(9): p. 1569-76.
13. Dang, M.T., J. Hambleton, and S.R. Kayser, *The influence of ethnicity on warfarin dosage requirement.* The Annals of pharmacotherapy, 2005. **39**(6): p. 1008-12.
14. O'Donnell, P.H. and M.E. Dolan, *Cancer pharmacoethnicity: ethnic differences in susceptibility to the effects of chemotherapy.* Clin Cancer Res, 2009. **15**(15): p. 4806-14.
15. Ma, B.B., E.P. Hui, and T.S. Mok, *Population-based differences in treatment outcome following anticancer drug therapies.* The lancet oncology, 2010. **11**(1): p. 75-84.
16. Materson, B.J., et al., *Single-drug therapy for hypertension in men. A comparison of six antihypertensive agents with placebo. The Department of Veterans Affairs Cooperative Study Group on Antihypertensive Agents.* The New England journal of medicine, 1993. **328**(13): p. 914-21.
17. Edwards, I.R. and J.K. Aronson, *Adverse drug reactions: definitions, diagnosis, and management.* Lancet, 2000. **356**(9237): p. 1255-9.
18. McCollum, A.D., et al., *Outcomes and toxicity in african-american and caucasian patients in a randomized adjuvant chemotherapy trial for colon cancer.* Journal of the National Cancer Institute, 2002. **94**(15): p. 1160-7.
19. Han, H.S., et al., *Racial differences in acute toxicities of neoadjuvant or adjuvant chemotherapy in patients with early-stage breast cancer.* European journal of cancer, 2011. **47**(17): p. 2537-45.
20. Sanoff, H.K., et al., *Racial differences in advanced colorectal cancer outcomes and pharmacogenetics: a subgroup analysis of a large randomized clinical trial.* Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2009. **27**(25): p. 4109-15.
21. Sekine, I., et al., *Common arm analysis: one approach to develop the basis for global standardization in clinical trials of non-small cell lung cancer.* Lung cancer, 2006. **53**(2): p. 157-64.

22. Diasio, R.B., T.L. Beavers, and J.T. Carpenter, *Familial deficiency of dihydropyrimidine dehydrogenase. Biochemical basis for familial pyrimidinemia and severe 5-fluorouracil-induced toxicity.* The Journal of clinical investigation, 1988. **81**(1): p. 47-51.
23. Perez-Stable, E.J., et al., *Nicotine metabolism and intake in black and white smokers.* JAMA : the journal of the American Medical Association, 1998. **280**(2): p. 152-6.
24. Nakajima, M., et al., *Comprehensive evaluation of variability in nicotine metabolism and CYP2A6 polymorphic alleles in four ethnic populations.* Clinical pharmacology and therapeutics, 2006. **80**(3): p. 282-97.
25. Caraballo, R.S., et al., *Racial and ethnic differences in serum cotinine levels of cigarette smokers: Third National Health and Nutrition Examination Survey, 1988-1991.* JAMA : the journal of the American Medical Association, 1998. **280**(2): p. 135-9.
26. Sherry, S.T., et al., *dbSNP: the NCBI database of genetic variation.* Nucleic Acids Res, 2001. **29**(1): p. 308-11.
27. Limdi, N.A., et al., *Warfarin pharmacogenetics: a single VKORC1 polymorphism is predictive of dose across 3 racial groups.* Blood, 2010. **115**(18): p. 3827-34.
28. Adzhubei, I., D.M. Jordan, and S.R. Sunyaev, *Predicting functional effect of human missense mutations using PolyPhen-2.* Current protocols in human genetics / editorial board, Jonathan L. Haines ... [et al.], 2013. **Chapter 7**: p. Unit7 20.
29. Fairbrother, W.G., et al., *Predictive identification of exonic splicing enhancers in human genes.* Science, 2002. **297**(5583): p. 1007-13.
30. Matys, V., et al., *TRANSFAC: transcriptional regulation, from patterns to profiles.* Nucleic acids research, 2003. **31**(1): p. 374-8.
31. Frazer, K.A., et al., *A second generation human haplotype map of over 3.1 million SNPs.* Nature, 2007. **449**(7164): p. 851-61.
32. Thorisson, G.A., et al., *The International HapMap Project Web site.* Genome research, 2005. **15**(11): p. 1592-3.

33. Abecasis, G.R., et al., *An integrated map of genetic variation from 1,092 human genomes.* Nature, 2012. **491**(7422): p. 56-65.
34. Hodge, A.E., R.B. Altman, and T.E. Klein, *The PharmGKB: integration, aggregation, and annotation of pharmacogenomic data and knowledge.* Clinical pharmacology and therapeutics, 2007. **81**(1): p. 21-4.
35. *The International HapMap Project.* Nature, 2003. **426**(6968): p. 789-96.
36. Johnson, G.C., et al., *Haplotype tagging for the identification of common disease genes.* Nature genetics, 2001. **29**(2): p. 233-7.
37. Livingston, R.J., et al., *Pattern of sequence variation across 213 environmental response genes.* Genome research, 2004. **14**(10A): p. 1821-31.
38. Teo, Y.Y., et al., *Singapore Genome Variation Project: a haplotype map of three Southeast Asian populations.* Genome Res, 2009. **19**(11): p. 2154-62.
39. Michael Bamshad and Stephen P. Wooding, *Signatures of natural selection in the human genome*. Nature Review Genetics, 2003. **4**(1): p. 99-111.
40. Holsinger, K.E. and B.S. Weir, *Genetics in geographically structured populations: defining, estimating and interpreting F(ST)*. Nature reviews. Genetics, 2009. **10**(9): p. 639-50.

Chapter 2. An update on *ABCB*1 pharmacogenetics: insights from a 3D model into the location and evolutionary conservation of residues corresponding to SNPs associated with drug pharmacokinetics

This chapter is published as

Steven J. Wolf, Maulana Bachtiar*, Jingbo Wang, Tiow Suan Sim, Samuel S Chong, and Caroline G.L. Lee,*

***Pharmacogenomics Journal**, 2011, 11: 315-325*

**These authors contributed equally to this work. Majority of analyses in this work were performed by the candidate.*

Within month of publication, this paper was marked as the most accessed paper on the Pharmacogenomics Journal website.

2.1 Introduction

Differential response in drug therapies is a common clinical phenomenon with human genetic predisposition contributing as a major factor [1]. Single Nucleotide Polymorphisms (SNPs) in genes responsible for drug pharmacokinetics (PK) or pharmacodynamics (PD) activity are associated with variation in drug metabolism, transport and efficacy [2, 3]. Moreover, it

has been shown that SNPs could potentially exert certain functional changes to drug-response genes [4-6]. The potential effect to which SNPs could implicate relies on where they are located in genes [7].

This chapter focuses on studying coding SNPs in the ABCB1 multidrug resistance protein, a membrane transporter that plays a key role in drug pharmacokinetics. Coding SNPs can be categorized into non-synonymous (amino acid substituting) and synonymous (non amino acid substituting). The potential functional effect of non-synonymous SNPs in *ABCB*1 depends on the relative positioning of the affected amino acids. In this chapter, based on a mouse homolog crystal structure, I constructed a 3D protein model that allows investigation of these coding SNPs in the ABCB1 protein based on two perspectives: amino acid residue location and evolutionary conservation. This work was conducted in collaboration with Dr. Steven J. Wolf.

2.2 ABCB1 is involved in multidrug resistance and altered drug pharmacokinetics

The ATP Binding Cassette superfamily member protein, ABCB1 (P-glycoprotein or P-gp / MDR1), has a primary role in the unidirectional transport of drugs from the cytoplasm to the extra cellular environment [8]. Due to its involvement in resistance to drugs used in treatment of cancer, heart disease, AIDS, and epilepsy, this trans-membrane protein is widely referred as a multidrug resistance (MDR) protein [9, 10]. It is a phenotype that is often

associated with variation in the level of drug efficacy, adverse drug reaction (ADR) or alteration in drug-to-drug interaction [11, 12].

In drug pharmacokinetic (PK) and drug resistance phenomena, the role of this MDR protein is complex as it is often affected by other variables that may affect the expression and localization of the ABCB1 protein. For instance, the presence of other protein such as the pregnane-X-receptor (PXR), in addition to an ABCB1 upregulation, may alter a drug PK [13]. Other studies also reported that the family of ABC proteins can be found across the membrane of organelles such as lysosomes or endosomes, where drug transport takes place for further drug processing within these organelles [14, 15].

Genetic polymorphisms in drug transporter can potentially affect the drug PK, with various studies showing evidence of the impact on drug concentration variation in the cerebrospinal fluid or intracellular compartment [16]. Furthermore, ABCB1 protein structure and function have been shown to be important in its role in multidrug resistance [8, 17]. This role is well established since 1976, when the protein was first discovered [18]. Now, following the release of protein crystal structure of the closest ABCB1 mammalian homolog, the mouse Abcb1a [19], the role of genetic polymorphisms in altering the protein function can be established further. There is great structural and chemical diversity of substrates that are subject to ABCB1 protein activity. They are usually small molecules that bind the ABCB1 substrate binding region, which are found on the internal surface of the protein. This region is a large open cavity that is surrounded by 12 transmembrane domains (TMDs) and is integrated to the lipid-based cell membrane [8, 19].

These TMDs are connected with the nucleotide binding domains, where hydrolysis of ATP takes place in the cytoplasmic portion of the protein [20, 21]. ATP binding and hydrolysis in the ABCB1 protein are done at the region containing conserved motifs such as the Walker A, Walker B and Signature C motifs. In addition, the TMD also accomodates a number of conserved loops, including the A-loop, D-loop, H-loop, and Q-loop [22-25] . Most of the ABC protein family and homologs are known to carry the same motifs, which is proposed to be important in binding ATP when the NBD domains dimerize [26]. This dimerization would eventually change ABCB1 conformation, resulting in substrate dissociation and efflux from the protein [26, 27].

Because ABCB1 is a very active transporter that handles various substrates [25], genetic polymorphisms can incur a potentially significant consequence. These polymorphisms can arise in the form of amino acid change or alteration in gene expression. This will be more prominent where the sequence changes take place at the sites where substrate bind or at regions that are highly conserved, including the Nucleotide Binding Domain (NBD), which can affect ABCB1 function including drug efflux or small molecules pharmacokinetics. A comprehensive literature review summary on the involvement of *ABCB*1 coding SNPs in drug responses is provided in Appendix 1.

2.3 *ABCB*1 SNPs as potential contributor to variation in individual drug response

There has been much interest surrounding the role of genetic variation in the *ABCB*1 gene in altering protein expression and function. Single Nucleotide Polymorphisms (SNPs) is the most common type of genetic variation. As introduced in the previous chapter, these SNPs can potentially affect gene function depending on the location where they reside. In this chapter, the focus is on *ABCB*1 coding region SNPs, which can alter the protein structure in the case of non-synonymous SNPs. On the other hand, non amino acid substituting SNP can affect an mRNA sequence, in addition to affecting the stability and structure of a RNA transcript. How SNPs evolved, particularly in a way that is relevant to ABC proteins, have previously been discussed in literature [28, 29].

SNPs influence can be apparent in the occurrence of adverse drug reaction, drug efficacy difference or multidrug resistance. Using the high resolution crystal structure of a mammalian homolog, the Abcb1a, these SNPs can now be mapped to the 3D structure, allowing better observation of the location of the residue, in addition to conserved regions information and the residues within close vicinities, which may also be coding SNPs. Some of these SNPs are also implicated with drug pharmacokinetics. The analysis in this chapter is also portrayed in a flash movie, that is accessible at http://pfs.nus.edu.sg/demo_src/abcb1.html.

2.4 Polymorphic ABCB1 is also well conserved protein

The ABCB1 protein is well conserved and is 87% identical to the protein sequence of the mouse Abcb1a homolog (Figure 2.1). In order to perform a multiple sequence alignment, 11 homologous ABCB1 sequences from multiple organisms were obtained from the NCBI Homologene database (Table 2.1). Using the ClustalW algorithm, despite its half-transporter status in unicellular organism (Sav1866 and MsbaA), it can be observed that the sequence of this protein is conserved from *E. coli* to humans (Fig. 2.2). Here, the prefix 'E' refers to the exon location of a SNP in the *ABCB*1 gene sequence. Based on the dbSNP database (Build 131), out of 66 coding SNPs in total, 24 SNPs are categorized as synonymous (sSNPs – 's' in the Figures and Tables), whereas 42 SNPs are considered to be nonsynonymous (nsSNPs – 'ns' in the Figures and Tables). These *ABCB*1 coding polymorphisms are summarized in Table 2.2 and Table 2.3. When comparing the human and mouse protein sequences, it can be observed that all ABCB1 sSNPs residues are conserved (Shaded blue in Table 2.2, Fig. 2.2). Furthermore in 11 organisms, five residues corresponding to sSNPs are found to be conserved. These are R442, L554, T558, S565 and I598; which correspond to #s8, #s11, #s12, #13 and #s14 respectively in Fig. 2.3C.

In addition, out of 42 *ABCB*1 nsSNPs, there are 37 that can be mapped to the residues in the mouse crystal structure, which is in accordance with the pair-wise sequence alignment that was performed with ClustalW (Table 2.2, Fig. 2.1) [30, 31]. Based on this alignment, it was observed that out of 37 nsSNPs, the major amino acid alleles of 31 nsSNPs or 83.7%, are similar between

mouse and human (Table 2.2 and Fig. 2.2). In 11 organisms that were examined, only two residues from this total of 37 nsSNPs are conserved, whilst 12 residues are observed to be similar in 7 or more organisms (Table 2.2).

Globally, most of the coding SNPs are perceived to be located in the ABCB1 NBDs, are located in close vicinity to conserved regions such as the Walker A, Walker B and signature motif C (Fig. 2.2, Fig. 2.3B and Fig. 2.3C). One nsSNP, E26/3222A>C or the C1074W (ns#27 in Fig. 2.3B) is located inside the Walker A motif at the C-terminal NBD (Table 2.2, Fig. 2.2). For every 100 residues, there are around 6.2 SNPs in the NDBs, whereas in the TMDs, the number is almost half, which is 3.7 SNPs. Furthermore, out of a total of 14 SNPs that are located outside of the NBDs, there are five that are found at the protein external surface, while three SNPs are found within the internal structure that corresponds to the substrate binding site (Fig. 2.3).

Table 2.1 ABCB1 homologous proteins.

#	Homologous Protein (Oganisms)	Length (Amino Acids)	Accession Number
1	ABCB1 (*Homo sapiens*)	1280	NP_000918.2
2	ABCB1 (*Canis lupus familiaris*)	1280	NP_001003215.1
3	Abcb1a (*Mus musculus*)	1276	NP_035206.2
4	ABCB1 (*Gallus gallus*)	1288	NP_990225.1
5	pgp-1 (*Caenorhabditis elegans*)	1321	NP_502413.1
6	Mdr50 (*Drosophila melanogaster*)	1313	NP_523740.3
7	PGP18 (*Arabidopsis thaliana*)	1225	NP_189480.1
8	pmd1 (*Schizosaccharomyces pombe*)	1362	NP_588265.1
9	Sav1866 (*Staphylococcus aureus*)	578	NP_372390
10	MsbA (*Escherichia coli*)	582	NP_415434
11	PFMDR1 (*Plasmodium falciparum*)	1419	XP_001351787.1

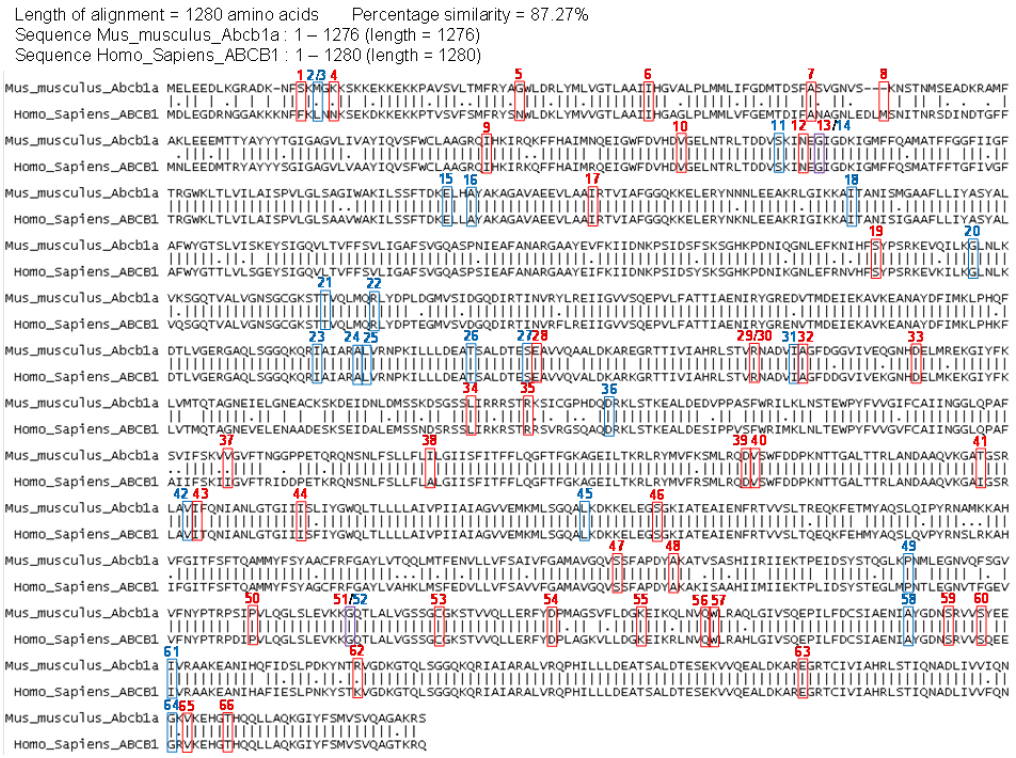


Figure 2.1 Pairwise sequence alignment of the ABCB1 human and mouse homologous protein sequences. The SNP numbering is identical to the ones used in Table 2.2 . Red and blue indicate position of nonsynonymous and synonymous SNPs, respectively. The purple boxes indicate residues concurrently housing both nonsynonymous and synonymous SNPs.

Figure 2.2 (next page): Homology map derived from the multiple sequence alignment of 11 ABCB1 homologs. Black bar height indicates the number of identical residues between the 11 homologous sequences at the respective amino acid site (between 11 and 0 identical residues may occur). sSNPs and nsSNPs whose corresponding residues can be mapped into the mouse protein are represented with blue and red dots, respectively. The C-terminal portion of the alignment where the bacterial half transporter sequences of Sav1866 and MsbA were repeated to produce a full sequence is marked with an arrow. This map also depicts the putative P-gp secondary structure shown in a linear chain format modified from the Protein Database (PDB). Conserved regions essential for protein function are highlighted both in the alignment and Abcb1a crystal structure.

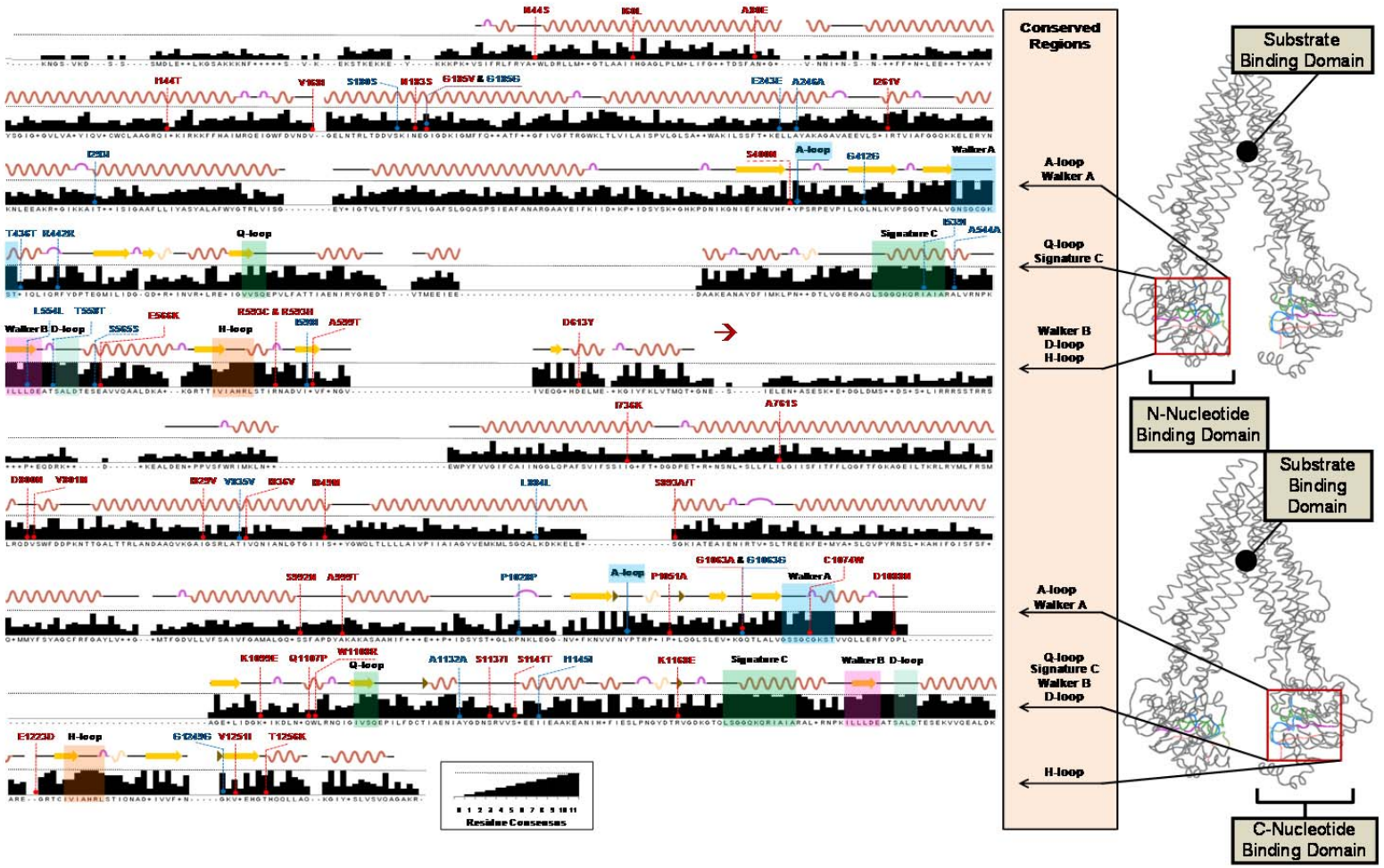

Conserved Regions
A-loop Walker A
Q-loop Signature C
Walker B D-loop H-loop
A-loop Walker A
Q-loop Signature C Walker B D-loop
H-loop
Substrate Binding Domain
N-Nucleotide Binding Domain
Substrate Binding Domain
C-Nucleotide Binding Domain
Residue Consensus

Table 2.2 Genetic conservation of amino acids corresponding to *ABCB*1 coding region SNPs.

#	rsNo	SNP (Amino Acid Substitution)		Mapped to Abcb1a domain (Internal/External Surface)	Amino Acid Residue Housing SNP											Conservation (%)^	
					H. Sapiens	*C. l. Familiaris*	*M. Musculus*	*G. gallus*	*C. elegans*	*D. melanogaster*	*A. thaliana*	*S. pombe*	*S. aureus*	*E. coli*	*P. falciparum*	Individual	Regional 3^{O} Structure
1	-	rs28381804	E3/49T>C	(F17L)	F17	W16	S16	Y17	Y35	F39	-	A42	-	-	G11	-	-
2	-	rs41304191	E3/55C>T	*(L19L)*	L19	M18	M18	I19	G37	P41	-	E44	-	-	L13	-	-
3	-	rs76199854	E3/57G>A	*(L19L)*	L19	M18	M18	I19	G37	P41	-	E44	-	-	L13	-	-
4	-	rs9282564	E3/61A>G	(N21D)	N21	K20	K20	N21	N39	K43	-	H46	-	-	I15	-	-
5	ns1	rs1202183	E5/131A>G	(N44S)	N44	N43	G43	S55	T70	T81	D16	D88	K10	A21	P50	18.2	38.7
6	ns2	rs41315618	E5/178A>C	(I60L)	I60	I59	I59	A71	I86	A97	G32	G104	K26	N37	L66	27.3	46.6
7	ns3	rs9282565	E5/239C>A	(A80E)	A80	A79	A79	V91	I106	I117	G52	Y124	N46	G57	N86	36.4	29.1
8	-	rs35810889	E5/266C>T	(M89T)	M89	F89	S85	T96	I112	-	-	-	-	-	-	-	-
9	ns4	rs61607171	E7/431T>C	(I144T)	I144	I145	I140	V152	N168	I176	A98	A172	L93	V97	L124	36.4	40.9
10	ns5	rs61122623	E7/502G>A	(V168I)	V168	V169	V164	A176	S192	S200	T122	A196	V117	T121	G148	54.5	46.2
11	s1	rs1128500	E8/540C>T	*(S180S)*	S180	S181	S176	S188	E204	S212	L136	N208	E129	E133	E160	27.3	42.0
12	ns6	rs60419673	E8/548A>G	(N183S)	N183	N184	N179	N191	K207	E215	Q139	Q211	K132	A136	S163	27.3	39.9
13	**ns7**	**rs1128501**	**E8/554G>T**	**(G185V)**	**G185**	**G186**	**G181**	**G193**	**G209**	**G217**	**F141**	**G213**	**F134**	**S138**	**G165**	**45.5**	**40.9**
14	s2	rs1128502	E8/555A>T	*(G185G)*	G185	G186	G181	G193	G209	G217	F141	G213	F134	S138	G165	45.5	40.9
15	s3	rs2235022	E9/729A>G	*(E243E)*	E243	E244	E239	E251	E267	E275	I199	Q271	R192	M196	S223	18.2	33.3
16	s4	rs28381867	E9/738G>A	*(A246A)*	A246	A247	A242	A254	R270	M278	E202	V274	A195	T199	Y226	9.1	34.5
17	ns8	rs36008564	E9/781A>G	(I261V)	I261	I262	I257	V267	I285	I293	V217	I289	I210	H214	I241	36.4	50.3
18	s5	rs80153317	E10/879T>C	*(I293I)*	I293	I294	I289	I301	L317	M325	L249	I321	R242	S246	F273	18.2	36.3
19	**ns9**	**rs2229109**	**E12/1199G>A**	**(S400N)**	**S400**	**S401**	**S396**	**N408**	**T424**	**Q439**	**T355**	**V428**	**Q348**	**T350**	**H386**	**18.2**	**60.7**
20	**s6**	**rs1128503**	**E13/1236C>T**	***(G412G)***	**G412**	**G413**	**G408**	**G420**	**G436**	**K451**	**D367**	**N440**	**D359**	**N361**	**D398**	**27.3**	**55.9**

Table 2.2. Genetic conservation of amino acids corresponding to *ABCB*1 coding region SNPs (con't).

#		rsNo	SNP (Amino Acid Substitution)		Mapped to Abcb1a domain (Internal/External Surface)	Amino Acid Residue Housing SNP											Conservation[A]	
						H. Sapiens	*C. l. Familiaris*	*M. Musculus*	*G. gallus*	*C. elegans*	*D. melanogaster*	*A. thaliana*	*S. pombe*	*S. aureus*	*E. coli*	*P. falciparum*	Individual	Regional 3[O] Structure
21	s7	rs35068177	E13/1308A>G	*(T436T)*	N-NBD (Internal)	T436	T437	T432	T44	I460	C475	V391	I464	L383	I385	I422	54.5	64.6
22	s8	rs41311775	E15/1326G>A	*(R442R)*		R442	R443	R438	R450	R466	R481	R397	R470	R389	R391	R428	100.0	54.5
23	s9	rs35633772	E15/1617C>T	*(I539I)*		I539	I540	I535	I547	I563	I578	I494	I576	L486	I489	I571	90.9	65.8
24	s10	rs60247941	E15/1632C>T	*(A544A)*		A544	A545	A540	A552	A568	A583	A499	A581	I491	A494	A576	63.6	65.1
25	s11	rs2235012	E15/1662G>C	*(L554L)*		L554	L555	L550	L562	L578	L593	L509	L591	L501	L504	L586	100.0	73.6
26	s12	rs56871767	E15/1674G>A	*(T558T)*		T558	T559	T554	T566	T582	T597	T513	T595	T505	T508	T590	100.0	78.7
27	s13	rs59697741	E15/1695C>T	*(S565S)*		S565	S566	S561	S573	S589	S604	S520	S602	S512	S515	S597	100.0	75.4
28	ns10	rs28381902	E15/1696G>A	(E566K)		E566	E567	E562	E574	E590	E605	E521	E603	E513	E516	E598	100.0	76.6
29	ns11	rs28381914	E16/1777C>T	(R593C)	N-NBD (External)	R593	R594	R589	R601	R617	R632	R548	R630	T540	E543	R627	54.5	67.3
30	ns12	rs56107566	E16/1778G>A	(R593H)		R593	R594	R589	R601	R617	R632	R548	R630	T540	E543	R627	54.5	67.3
31	s14	rs28381915	E16/1794C>T	*(I598I)*	N-NBD (Internal)	I598	I599	I594	I606	I622	I637	I553	I635	I545	I548	I632	100.0	65.5
32	ns13	rs2235036	E16/1795G>A	(A599T)		A599	A600	A595	A607	I623	V638	C554	V636	V546	V549	F633	54.5	63.6
33	ns14	rs57001392	E16/1837G>T	(D613Y)	N-NBD (External)	D613	D614	D609	S621	R637	Q652	E568	N650	R560	N563	D677	45.5	60.5
34	-	rs35657960	E17/1985T>C	(L662R)	-	L662	L663	L658	E671	M694	K698	E615	A708	-	-	E736	0.0	-
35	-	rs35023033	E17/2005C>T	(R669C)		R669	R670	R665	R678	I702	D705	S662	T715	-	-	N743	0.0	-
36	-	rs59340265	E17/2037C>T	*(D679D)*		D679	D680	D675	N688	D712	N715	S632	N725	-	-	E753	9.1	-
37	ns15	rs41316450	E18/2207T>A	(I736K)	TM7	I736	I737	V732	I745	M779	I771	V682	I820	I37	L48	V814	72.7	36.7
38	ns16	rs77144566	E19/2281A>C	(A761S)	TM8	A761	V763	I757	A769	V802	I796	G706	I844	I647	G655	L836	63.6	42.1
39	ns17	rs41305517	E21/2398G>A	(D800N)	C-NBD (Internal)	D800	D801	D796	D808	H841	D835	E745	D883	S108	P112	E875	9.1	45.4
40	ns18	rs2235039	E21/2401G>A	(V801M)	C-NBD (External)	V801	V802	V797	M809	I842	V836	V746	V884	A109	V113	M876	63.6	47.6

Table 2.2. Genetic conservation of amino acids corresponding to *ABCB*1 coding region SNPs (con't).

#		rsNo	SNP (Amino Acid Substitution)		Mapped to Abcb1a domain (Internal/External Surface)	Amino Acid Residue Housing SNP											Conservation	
						H. Sapiens	*C. l. Familiaris*	*M. Musculus*	*G. gallus*	*C. elegans*	*D. melanogaster*	*A. thaliana*	*S. pombe*	*S. aureus*	*E. coli*	*P. falciparum*	Individual	Regional 3^{O} Structure
41	ns19	rs2032581	E22/2485A>G	(I829V)	TM9	I829	I830	T825	T837	I870	T864	V744	S912	I135	S139	L904	45.5	30.2
42	s15	rs28381966	E22/2505A>G	*(V835V)*		V835	V836	V831	L843	T876	T870	L780	T918	N141	T145	F911	45.5	33.0
43	ns20	rs28381967	E22/2506A>G	(I836V)		I836	I837	I832	I844	V877	I871	L781	V919	I142	V146	F911	63.6	28.5
44	ns21	rs36105130	E22/2547A>G	(I849M)		I849	I850	I845	I857	I890	D884	I794	L932	I155	M159	M924	81.8	44.7
45	s16	rs9282563	E22/2650C>T	*(L884L)*	TM10	L884	L885	L880	K892	V925	M919	R829	E967	R190	K194	I959	27.3	31.2
46	**ns22**	**rs2032582**	**E22/2677G>T/A**	**(S893A/T)**		**S893**	**A894**	**S889**	**A901**	**S934**	**C928**	**S838**	**S976**	**V199**	**V203**	**P982**	**45.5**	**35.7**
47	ns23	rs56849127	E25/2975G>A	(S992N)	TM12	S992	S993	S988	S1000	T1035	L1027	G937	F1075	V298	T302	M1081	36.4	38.4
48	ns24	rs72552784	E25/2995G>A	(A999T)		A999	A1000	A995	A1007	A1042	Q1034	V944	T1082	T305	Q309	E1089	27.3	36.5
49	s17	rs2235044	E26/3084G>A	*(P1028P)*	C-NBD (External)	P1028	P1029	P1024	P1036	-	P1063	P973	V1111	P332	V334	I1117	0.0	33.0
50	ns25	rs28401798	E26/3151C>G	(P1051A)		P1051	P1052	P1047	K1059	E1093	Q1086	I996	K1135	P355	P357	P1142	18.2	57.6
51	ns26	rs2707944	E26/3188G>C	(G1063A)		G1063	G1064	G1059	G1071	G1105	G1098	G1008	G1147	G367	G369	K1154	45.5	53.8
52	s18	rs2707943	E26/3189C>G	*(G1063G)*		G1063	G1064	G1059	G1071	G1105	G1098	G1008	G1147	G367	G369	K1154	45.5	53.8
53	ns27	rs74755520	E26/3222A>C	(C1074W)	C-NBD (Internal)	C1074	C1075	C1070	C1082	C1116	C1109	S1019	C1158	G378	S380	S1165	63.6	67.8
54	ns28	rs57521326	E26/3262G>A	(D1088N)		D1088	D1089	D1084	D1096	D1130	D1123	D1033	D1172	D392	D394	D1179	100.0	53.9
55	ns29	rs41309225	E27/3295A>G	(K1099E)	C-NBD (External)	K1099	K1100	K1095	I1107	S1141	C1134	R1044	V1183	H403	H405	I1237	18.2	38.5
56	ns30	rs55852620	E27/3320A>C	(Q1107P)		Q1107	Q1108	Q1103	Q1115	E1149	T1142	R1052	N1191	G411	A413	R1245	27.3	43.7
57	ns31	rs35730308	E27/3322T>C	(W1108R)		W1108	Q1109	W1104	Q1116	H1150	N1143	S1053	D1192	S412	S414	D1246	27.3	43.5
58	s19	rs34748655	E27/3396C>T	*(A1132A)*	C-NBD (Internal)	A1132	A1133	A1128	A1140	I1174	S1167	M1077	V1216	L436	A438	K1270	27.3	56.8
59	ns32	rs41309228	E27/3410G>T	(S1137I)	C-NBD (External)	S1137	S1138	S1133	S1145	P1179	A1172	S1082	S1220	P440	E443	-	18.2	41.8
60	ns33	rs2229107	E27/3421T>A	(S1141T)		S1141	S1142	S1137	S1149	T1183	T1176	D1086	S1224	D444	R447	T1277	45.5	50.9
61	**s20**	**rs1045642**	**E27/3435C>T**	***(I1145I)***	**C-NBD (Internal)**	**I1145**	**I1146**	**I1141**	**I1153**	**V1187**	**I1180**	**I1090**	**M1228**	**V447**	**I450**	**V1281**	**72.7**	**57.6**
62	ns34	rs59241388	E28/3502A>G	(K1168E)	C-NBD (External)	K1168	R1169	R1164	R1176	R1210	R1203	C1113	L1251	E470	V473	N1304	27.3	61.9
63	ns35	rs41309231	E29/3669A>T	(E1223D)		E1223	E1224	E1219	E1231	E1265	E1258	V1168	Q1306	K525	K528	D1359	27.3	60.2
64	s21	rs2235051	E29/3747C>G	*(G1249G)*	C-NBD (Internal)	G1249	G1250	G1245	G1257	G1291	G1284	G1194	G1332	G551	G554	T1392	63.6	63.0
65	ns36	rs45456698	E29/3751G>A	(V1251I)	C-NBD (External)	V1251	V1252	V1247	V1259	I1293	V1286	V1196	I1334	I553	I556	V1394	81.8	59.2
66	ns37	rs35721439	E29/3767C>A	(T1256K)		T1256	T1257	T1252	T1264	T1298	D1291	N1201	T1339	T558	T561	T1399	72.7	59.4

Table 2.2 (previous page) Note. ^The conservation of residues corresponding to all coding regions SNPs was obtained following multiple sequence alignment of 11 confirmed ABCB1 homolog protein sequences. For each species the amino acid, as well as its position in the corresponding sequence, is indicated. Corresponding ends of each of the row containing *ABCB*1 SNPs of a particular pharmacogenomics interest are marked by the same colour that were used to highlight their names in Figure 2.3. Highlighted in faded blue are the rows containing synonymous SNPs. Shaded in green is the sequence alignment obtained after re-using the same bacterial half-transporter sequence that were used for the first half of the alignment. The SNP nomenclature (i.e. ns# (non-synonymous) or s# (synonymous)) is similar to what were used throughout this chapter.

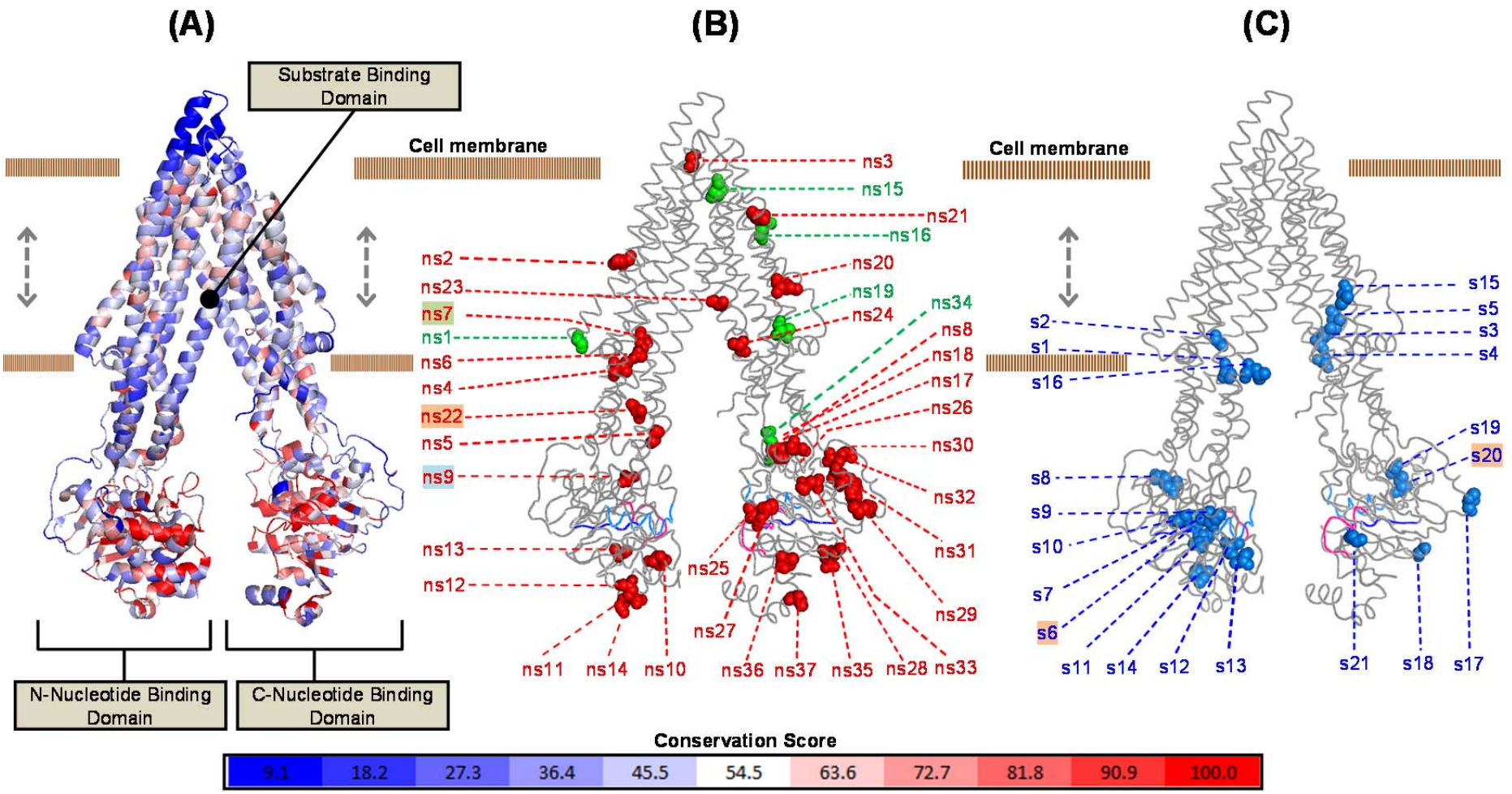

(A)
(B)
(C)
Substrate Binding Domain
Cell membrane
N-Nucleotide Binding Domain
C-Nucleotide Binding Domain
ns3
ns15
ns21
ns16
ns2
ns20
ns23
ns19
ns34
ns7
ns24
ns8
ns1
ns18
ns6
ns17
ns4
ns26
ns22
ns30
ns5
ns32
ns9
ns31
ns13
ns25
ns29
ns12
ns27
ns11
ns14
ns10
ns36
ns37
ns35
ns28
ns33
s15
s5
s2
s3
s1
s4
s16
s19
s20
s8
s9
s10
s7
s6
s11
s14
s12
s13
s21
s18
s17
Conservation Score
9.1
18.2
27.3
36.4
45.5
54.5
63.6
72.7
81.8
90.9
100.0

Figure 2.3 (previous page): (A) A global view of residue conservation following the multiple sequence alignment of 11 ABCB1 homologs using the ClustalW algorithm (see Table 2.1 and Table 2.2). A heat map with differing colors depending on the residue conservation score is shown (blue: low conservation, red: high conservation). (B) Global view of amino acid residues corresponding to non-synonymous SNPs (prefixed as 'ns') and (C) synonymous SNPs (prefixed as 's') within the human *ABCB1* gene mapped to the mouse Abcb1a crystal structure (PDB: 3G5U). Identical residues between human and mouse are represented as Blue (sSNPs) and Red (nsSNPs) balls/fonts. Those residues that are not homologous are represented as green balls/fonts. The conserved Walker A (light blue), Walker B (pink) and signature C motifs (dark blue) are also highlighted as coloured lines. Number labels correspond with the data/labels in Table 2.2. The residue housing SNPs of pharmacogenomics interest are consistently highlighted by the same colours in Tables 2.2 and 2.3.

2.5 SNPs associated with protein function or expression

Based on literature review, there are two sSNPs and 12 nsSNPs that are associated with the ABCB1 protein function or expression variation (Table 2.3). However, because they are located outside of the crystalized region of the Abcb1a protein, four of the twelve nsSNPs could not be located in the protein structure. These are the E3/61A>G, E5/266C>T, E17/1985T>C and E17/2005C>T.

Table 2.3 *ABCB*1 coding SNPs and their allele frequencies.

#		rsNo	SNP (Amino Acid Substitution)		Minor Allele Frequency (%)					Association Report(s) (Y/N)
					Asian		CAU	YRI	Ref.	
					CHB	JPT				
1	-	rs28381804	E3/49T>C	(F17L)	2.5		0	0	b	No
2	-	rs41304191	E3/55C>T	*(L19L)*			-			No
3	-	rs76199854	E3/57G>A	*(L19L)*			-			No
4	-	rs9282564	E3/61A>G	(N21D)	0	0	10	0	a	Yes
					0		19	0	b	
					-	-	6.5	2.1	c	
5	ns1	rs1202183	E5/131A>G	(N44S)	0	0	0	0	a	No
					-	-	0	0	c	
6	ns2	rs41315618	E5/178A>C	(I60L)			-			No
7	ns3	rs9282565	E5/239C>A	(A80E)	0	0	0	0	a	No
					-	-	0	2.1	c	
8	-	rs35810889	E5/266C>T	(M89T)			-			Yes
9	ns4	rs61607171	E7/431T>C	(I144T)			-			No
10	ns5	rs61122623	E7/502G>A	(V168I)			-			No
11	s1	rs1128500	E8/540C>T	*(S180S)*			-			No
12	ns6	rs60419673	E8/548A>G	(N183S)			-			No
13	**ns7**	**rs1128501**	**E8/554G>T**	**(G185V)**			**-**			**Yes**
14	s2	rs1128502	E8/555A>T	*(G185G)*			-			No
15	s3	rs2235022	E9/729A>G	*(E243E)*	0	0	0	0	a	No

16	s4	rs28381867	E9/738G>A	*(A246A)*	0		0	0	b	No
17	ns8	rs36008564	E9/781A>G	(I261V)			-			No
18	s5	rs80153317	E10/879T>C	*(I293I)*			-			No
19	**ns9**	**rs2229109**	**E12/1199G>A**	**(S400N)**	**0** **0** **-** **0**	**0** **-** **-**	**3.3** **2.3** **3.2** **2.1**	**0** **0** **2.1** **0**	**a** **b** **c** **d**	**Yes**
20	**s6**	**rs1128503**	**E13/1236C>T**	***(G412G)***	**68.9** **68.8** **-** **66.7**	**57.8** **-** **-**	**39.2** **38.6** **46.8** **47.7**	**12.3** **25** **18.7** **13.6**	**a** **b** **c** **d**	**Yes**
21	s7	rs35068177	E13/1308A>G	*(T436T)*			-			No
22	s8	rs41311775	E15/1326G>A	*(R442R)*			-			No
23	s9	rs35633772	E15/1617C>T	*(I539I)*			-			No
24	s10	rs60247941	E15/1632C>T	*(A544A)*			-			No
25	s11	rs2235012	E15/1662G>C	*(L554L)*	0 0	0	0 0	2.5 1.7 6.2	a c	No
26	s12	rs56871767	E15/1674G>A	*(T558T)*			-			No
27	s13	rs59697741	E15/1695C>T	*(S565S)*			-			No
28	ns10	rs28381902	E15/1696G>A	(E566K)	0.6 0	0.6	1.3 0	0.5 0	a b	No
29	ns11	rs28381914	E16/1777C>T	(R593C)	0		0	0	b	No
30	ns12	rs56107566	E16/1778G>A	(R593H)			-			No
31	s14	rs28381915	E16/1794C>T	*(I598I)*	-	-	0.9	-	a	No

						0	2.3	0	b	
32	ns13	rs2235036	E16/1795G>A	(A599T)	0	0	0	0	a	No
33	ns14	rs57001392	E16/1837G>T	(D613Y)			-			No
34	-	rs35657960	E17/1985T>C	(L662R)			-			Yes
35	-	rs35023033	E17/2005C>T	(R669C)			-			Yes
36	-	rs59340265	E17/2037C>T	*(D679D)*			-			No
37	ns15	rs41316450	E18/2207T>A	(I736K)			-			No
38	ns16	rs77144566	E19/2281A>C	(A761S)						No
39	ns17	rs41305517	E21/2398G>A	(D800N)			-			No
40	ns18	rs2235039	E21/2401G>A	(V801M)	0	0	0	0	a	No
41	ns19	rs2032581	E22/2485A>G	(I829V)	0	0	0.9	0	a	No
42	s15	rs28381966	E22/2505A>G	*(V835V)*	- 0	-	0.4 2.4	- 0	a b	No
43	ns20	rs28381967	E22/2506A>G	(I836V)		0	0	4.2	b	No
44	ns21	rs36105130	E22/2547A>G	(I849M)			-			No
45	s16	rs9282563	E22/2650C>T	*(L884L)*	- 0	-	1.6 2.3	0 0	c b	No
46	**ns22**	**rs2032582**	**E22/2677G>T/A**	**(S893A/T)**	**52.2/15.2 -**	**-**	**38.6/0 43.5/0**	**0/0 12.5/2.1**	**b c**	**Yes**
47	ns23	rs56849127	E25/2975G>A	(S992N)			-			No
48	ns24	rs72552784	E25/2995G>A	(A999T)			-			No
49	s17	rs2235044	E26/3084G>A	*(P1028P)*	0	0	0	0	a	No
50	ns25	rs28401798	E26/3151C>G	(P1051A)		0	0	0	b	Yes
51	ns26	rs2707944	E26/3188G>C	(G1063A)			-			No
52	s18	rs2707943	E26/3189C>G	*(G1063G)*			-			No

53	ns27	rs74755520	E26/3222A>C	(C1074W)							No
54	ns28	rs57521326	E26/3262G>A	(D1088N)			-				Yes
55	ns29	rs41309225	E27/3295A>G	(K1099E)			-				No
56	ns30	rs55852620	E27/3320A>C	(Q1107P)			-				No
57	ns31	rs35730308	E27/3322T>C	(W1108R)			-				Yes
58	s19	rs34748655	E27/3396C>T	*(A1132A)*	-	-	0	2.1	c		No
59	ns32	rs41309228	E27/3410G>T	(S1137I)			-				No
60	ns33	rs2229107	E27/3421T>A	(S1141T)	0	0	0	4.5	a		Yes
					0	-	0	13.6	d		
					-	-	0	10.4	c		
					0		0	4.2	b		
61	**s20**	**rs1045642**	**E27/3435C>T**	***(I1145I)***	**40**	**47.8**	**54.2**	**11.7**	**a**		**Yes**
					52.3		**54.5**	**20.8**	**b**		
					-	**-**	**53.2**	**20.8**	**c**		
					37.5	**-**	**62.5**	**15.2**	**d**		
62	ns34	rs59241388	E28/3502A>G	(K1168E)			-				No
63	ns35	rs41309231	E29/3669A>T	(E1223D)			-				No
64	s21	rs2235051	E29/3747C>G	*(G1249G)*	-	-	0	1.4	c		No
65	ns36	rs45456698	E29/3751G>A	(V1251I)	0		0	0	b		Yes
66	ns37	rs35721439	E29/3767C>A	(T1256K)			-				No

~The corresponding ends of each of the row containing *ABCB*1 SNPs of a particular pharmacogenomics interest are marked by the same colour that were used to highlight their names in Figure 2.3. Highlighted in faded blue are the rows containing synonymous SNPs. Minor allele frequency source: HapMap (a), EGP SNPs (b), SNP500Cancer (c), PERLEGEN (d).

2.6 In 3D structure E13/1236C>T, E22/2677G>T/A and E27/3435C>T are located in distant regions and have varied conservation

The E13/1236C>T, E22/2677G>T/A and E27/3435C>T are among the SNPs that are most extensively studied in *ABCB*1 pharmacogenetics [32-37]. Not only are they present at diverse frequencies across different populations (Table 2.3), it is also reported that these three SNPs are in high linkage disequilibrium [35]. Whilst the E22/2677G>T/A is categorized as nsSNPs (#ns22 in Fig. 2.3B), the other two SNPs are synonymous in nature (#s6 and #s20 in Fig. 2.3C). Interestingly, unlike the conventional biallelic SNPs, nsSNP #22, which is located at residue 893, exists with three different alleles conferring a Serine to Alanine or Threonine substitution.

By using the 3D structure, the two sSNPs are observed to be located in different protein regions (Fig. 2.3C and flash movie at http://pfs.nus.edu.sg/demo_src/abcb1.html)), which are located more than 50 Å apart. They also have a very dissimilar evolutionary conservation level. To calculate regional conservation score, amino acid residues within 10 Å of the SNP were fetched from the Abcb1a crystal structure. Subsequently, an average score of all the individual conservation scores of these residues, in addition to the scores of five residues adjacent to them were calculated and assigned to the residue of interest. This score was then referred as the regional conservation score. Table 2.2 presents the regional conservation scores of the *ABCB*1 coding SNPs. The reason for choosing five amino acids that are

located on either side of each residue was to obtain a finer spread of data, in demarcating the regions that have a big difference in conservation scores.

Synonymous SNP #s6 or the E13/1236C>T, which corresponds to a glycine, is located on the external surface within the NBD that is closer to the N-terminal (Fig. 2.2 and Fig. 2.3C). This residue is found inside a beta-sheet, between the Walker A and A-loop motif, which are separated by 15 and 10 amino acids, respectively. In this study, it is observed that the individual conservation score of this residue is relatively low (28%), which suggests that there is a considerable degree of tolerance for variation in this residue during the protein evolution (Table 2.2, Fig. 2.4A). Furthermore, this notion is supported by the presence of lysine, aspartic acid and asparagine in residues of the same position that is found in other homologous sequences (Table 2.2). They only differ from glycine in term of polarity and charge of side chain, in addition to being less hydrophobic (-3.5 to -3.9 in compared to -0.4). Because #s6 is likely to have an impact at mRNA level, this variation of tolerance is postulated to be important in nature.

The second highly studied sSNP was the E27/3435C>T or #s20, which corresponds to I1145I in the protein (Fig. 2.3C). The residue that corresponds to this SNP is observed to be highly conserved (73%) in all 11 ABCB1 homologous protein sequences (Table 2.1 and Table 2.2). Except in *C. elegans, S. aureus* and *P. falciparum*, which has a valine residue, all other ABCB1 homologs have isoleucine at the corresponding site to SNP #s20 (Fig. 2.2, Table 2.2). In the 3D structure, it can be observed that Ile 1145 is located within the internal structure of the NBD that is closer to the C-terminal (Fig. 2.3B). Because both isoleucine and valine have neutral side chains with non-

polar and hydrophobic residues, it suggested that these properties are important for this site of the protein throughout its evolution.

Theoretically, E27/3435C>T is similar to E13/1236C>T in term of variability as it is at the transcript level that they may exert influence. And it could be predicted that in *S. pombe*, methionine 1228 could have been the result of a mutation that took place at the isoleucine codon, from ATC to ATG. Residue 1145 is located right next to the signature C (Fig. 2.4C), a highly conserved motif and functionally important for ATP binding and hydrolysis [38, 39]. The regional conservation scores of E13/1236C>T and E27/3435C>T are highly similar, 56% and 58%, respectively, despite the variation in their conservation level and their opposite NBDs placement. Furthermore, these regions are more conserved than the average conservation of other SNPs, with regional conservation score of more than 50%.

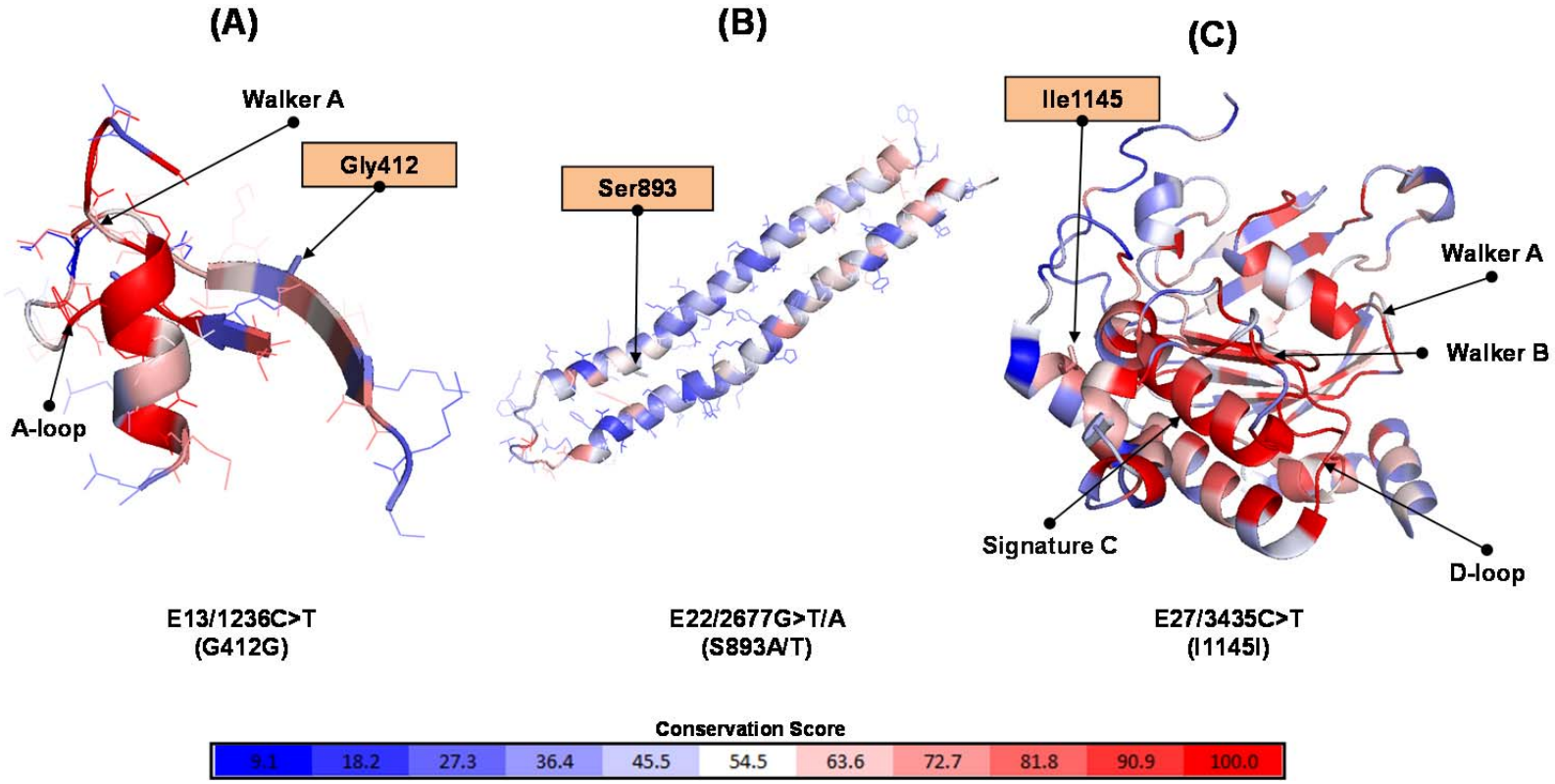


Figure 2.4 Location and conservation of (A) E13/1236C>T (G412G), (B) E22/2677G>T/A (S893A/T) and (C) E27/3435C>T (I1145I). Individual residue scores are mapped to the ribbon model with each colour representing a percentage of conservation where 100% is equal to 11 species expressing the same residue at the corresponding position. A heat map with differing colors depending on the conservation score is shown (blue: low conservation, red: high conservation).

The E27/3435C>T polymorphism (#s20 in Fig. 2.3C), which corresponds to the I1145I is a highly studied SNP of the ABCB1 protein, which is highly suggested to affect the protein 3D structure, stability or expression following the occurrence of a ribosomal stall during mRNA translation. This ribosomal stall is proposed to occur due to the presence of rare codon that changes the speed of mRNA translation and protein folding that is facilitated by chaperone [6, 40, 41]. This intriguing hypothesis however, is still supported by limited evidence. Moreover, rather than focusing on this SNP alone, most of the studies had largely been focusing on the SNPs haplotype consisting of #s6, #ns22, in addititon to the #s20 itself.

Similar to the #s20, the #ns22, which corresponds to E22/2677G>T/A has a higher conservation score than the residues in its vicinity (#ns22 in Fig. 2.3B). It is a non-synonymous SNP that is associated with substitution of amino acids at residue 893, from Serine to either Alanine or Threonine. Based on the regional conservation score of 36% and individual conservation score of 46%, the SNP probably lies in a region that was subjected to sequence variation throughout the protein evolution (Table 2.2, Fig. 2.4B). Moreover, several homologous residues are different to Serine in the polarity of side chains and hydrophaty index [42]. These are the Alanine in *C. l. Familiaris* and *G. gullus,* the cysteine in *D. melanogaster*, and the Valine in Sav1866 and MsbA.

There have been enormous efforts that focus in studying the role of these three *ABCB*1 SNPs in drug response, many of them often yielded conflicting conclusion (See Supplementary Table 3 in the published manuscript version). Where some studies concluded on the causative role of a single *ABCB*1 SNP, the majority of other studies have reported on the role of the SNP haplotypes. With many contradictory conclusion, the real evidence surrounding the significant role of the E13/1236C>T, E22/2677G>T/A and E27/3435C>T SNPs have therefore been largely controversial. Moreover, reviews on the association between *ABCB*1 polymorphisms and drug pharmacokinetics or protein functions have also carried mixed perspectives [10, 16, 43].

Because inferring a conclusion from many of these studies is not easy, especially in terms of *ABCB*1 pharmacogenetics, the availability of the 3D crystal structure provided an opportunity. With the mouse structure, we can now evaluate the potential effect of *ABCB*1 SNPs on the protein structure and function. This could complement the design of future functional studies, especially in the selection of candidate SNPs.

2.7 3D structure reveals that classic G185V polymorphism is in close proximity to two other non-synonymous SNPs in less evolutionary conserved region

Besides the E13/1236C>T, E22/2677G>T/A and E27/3435C>T SNPs, there are several nsSNPs that are associated with drug responses. The majority of these SNPs could be mapped to the mammalian Abcb1a crystal structure.

The E8/554G>T (#ns7), which is a G185V substitution (Fig. 2.3B) is one of the non-synonymous SNPs that have been reported to be associated with drug response. It is a "classic" polymorphism that induce an amino acid substitution from Glycine to Valine at residue 185, and is associated with vinblastine and colchicine drug specificity [44]. It is reported that glycine 185 is an important component in the alteration of ABCB1 conformation between the drug binding and catalytic sites [45]. Atomic detail homology modeling coupled with combining dynamics simulation predicted an improved drug efflux as a result of non-polar van der Waals force reduction that were supposed to bind colchicine near residue 185. Because Valine is bulkier than Glycine, this interaction would have been prevented, which increases efflux and dissociation [45, 46].

As can be observed in the crystal structure (Fig. 2.5 and 2.6), G185V is located within a close 3D vicinity to two other nsSNPs, the I144T (#ns4 in Fig. 2.3B) and N183S (#ns6 in Fig. 2.3B). With the #ns4, the G185V is located 10.9 Å apart, whilst with the N183S, it is 5.2 Å apart. However, because ABCB1 is known to have more than one structural confirmation [19], there is a concern that this close proximity may not be observed in a different

conformation. To address this, the same SNPs were mapped to the Sav1866 ADP-bound structure, which is a half transporter found in *Staphylococcus aureus*. Here, it can be observed that the 3D distance between these nsSNPs are roughly similar, with 11.3 and 5.3 Å, respectively (Fig. 2.5 and Fig. 2.6).

With an average score of around 40% for the I144, N183 and G185, the individual conservation scores are 36%, 27% and 45%, respectively. Within 10 Å region, there are only three residues that have conservation score of 80% or more. It can therefore be suggested that throughout the protein evolution, some degree of variation of residues are tolerated. One should evaluate the influence of I144T and N183S on ABCB1 conformation in addition to their influence on the role of G185V in protein functional change.

However, few studies focus on the role of these three SNPs in drug response. In fact, no genotype information is available (Table 2.3). Therefore, in order to determine whether drug specificity is influenced by these SNPs, the next focus should be in obtaining the genotype information, in addition to the haplotype information across different populations.

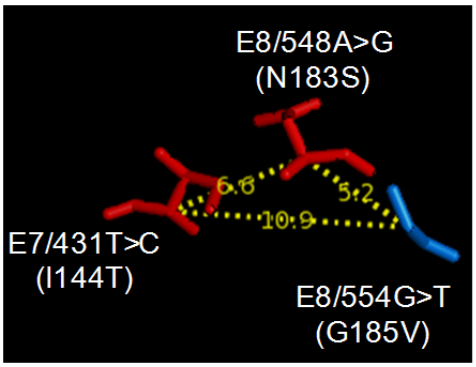


Figure 2.5 Three homologous residues housing I144T, N183S and G185V in mouse Abcb1a structure (ATP/ADP free form). The homologous SNP amino acid sites are identical between human and mouse. Distances are indicated with yellow dashed lines.

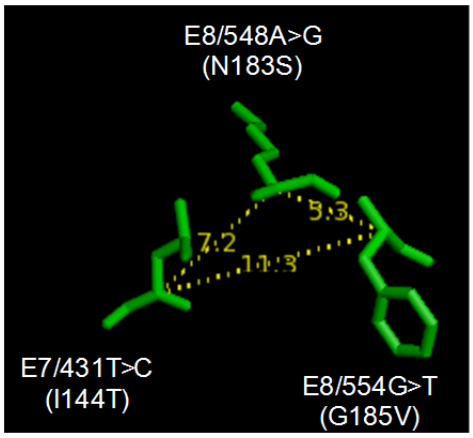
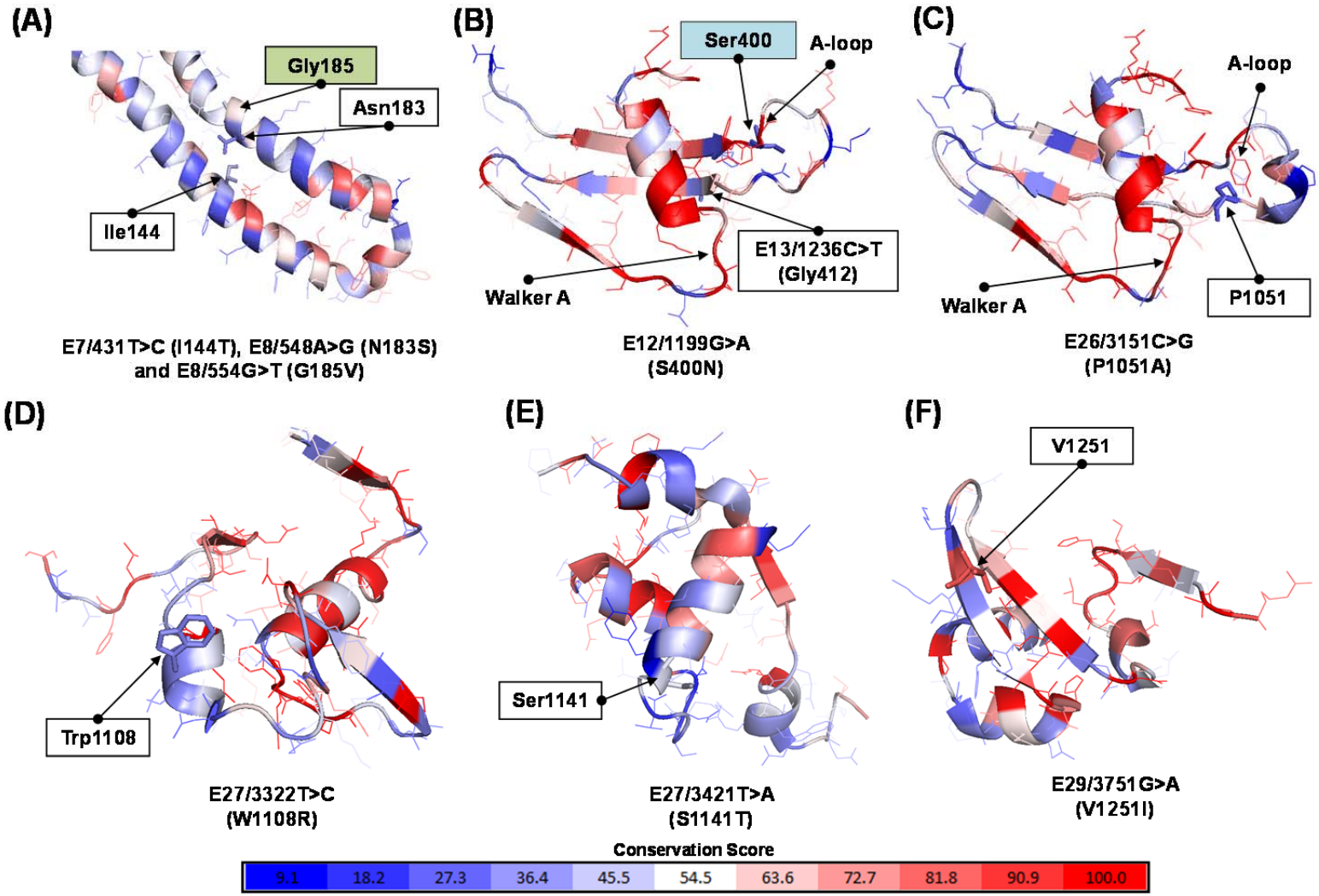


Figure 2.6 Three homologous residues housing I144T, N183S and G185V in Staphylococcus aureus Sav1866 structure (ADP-bound conformation). The homologous SNP amino acid sites are non-identical between human and *S. aureus*. Distances are indicated with yellow dashed lines.

Figure 2.7 (next page) Location and conservation of (A) E8/554G>T (G185V), (B) E12/1199G>A (S400N), (C) E26/3151C>G (P1051A), (D) E27/3322T>C (W1108R), (E) E27/3421T>A (S1141T), (F) E29/3751G>A (V1251I). Individual residue scores are mapped to the ribbon model with each colour representing a percentage of conservation where 100% is equal to 11 species expressing the same residue at the corresponding position. A heat map with differing colors depending on the conservation score is shown (blue: low conservation, red: high conservation).


(A)
Gly185
Asn183
Ile144
E7/431T>C (I144T), E8/548A>G (N183S)
and E8/554G>T (G185V)
(B)
Ser400
A-loop
E13/1236C>T
(Gly412)
Walker A
E12/1199G>A
(S400N)
(C)
A-loop
Walker A
P1051
E26/3151C>G
(P1051A)
(D)
Trp1108
E27/3322T>C
(W1108R)
(E)
Ser1141
E27/3421T>A
(S1141T)
(F)
V1251
E29/3751G>A
(V1251I)
Conservation Score
9.1
18.2
27.3
36.4
45.5
54.5
63.6
72.7
81.8
90.9
100.0

2.8 E12/1199G>A (S400N) is evolutionary non-conserved, but resides in an evolutionary conserved region

The E12/1199G>A (#ns9 in Fig. 2.3b) nsSNP, which induces an S400N residue substitution, has a frequency of less than 4% in Asians. This SNP has been previously reported to be influential in drug responses [47-53]. With the mouse protein structure, it can be observed that S400N is similarly located as Gly 412, which houses the synonymous SNP, E13/1236C>T (Fig. 2.7B and see flash movie at http://pfs.nus.edu.sg/demo_src/abcb1.html). It lies within the NBD, next to the A-loop (Fig. 2.2 and Fig. 2.8B) at the same turn region as Gly 412, in between two beta-sheets. Based on this analysis however, the S400 residue is observed to have a relatively low individual conservation score (18%), which indicates poor evolutionary conservation that may tolerate variation at this position (Table 2.2). At this position, serine is observed only in three species: human, canine and mouse. In other species, substitution takes place with amino acids having polar side chain, except in yeast where there is an occurrence of valine, which is hydrophobic in property. Furthermore, there is an interesting observation that this S400 residue is flanked by two residues that are completely conserved, with a score of 100% (Fig. 2.7B), in a region of relatively high conservation, with a regional conservation score of 60.72% (Table 2.2).

2.9 Four SNPs that are associated with drug response are mapped to the outer surface of C-terminal NBD

Using pair-wise protein sequence alignment, there are four out of eight SNPs that are reported to be associated with drug response that could be mapped to the crystal structure of the mouse Abcb1a protein (Fig. 2.1) [30, 31]. These SNPs are the P1051A, W1108R, S1141T, and V1251I, which corresponds to #ns25, #ns31, #ns33, and #ns36, respectively (Fig. 2.3B). These SNPs, which reside in the C-terminal Nucleotide Binding Domain, may affect protein function if they are reside in close vicinity to the protein surface that is important for ATP binding. The outer protein surface may also be important in NBDs dimerization, either during ATP binding and hydrolysis or opposing NBDs interaction.

Here, the SNP E26/3151C>G, which corresponds to a P1051A (#ns25) substitution is highlighted (Fig. 2.3B and Fig. 2.7C). It is a relatively well conserved amino acid substitution, affecting only protein hydropathy. In the 3D structure, #ns25 is mapped to the C-terminal NBD, at the surface forming the ATP binding pocket. The SNP is found to be located between two beta-sheets, within a turn region, in a fashion that is similar to the S400N polymorphism (Fig. 2.7B). SNP P1051A is located close to the A-loop and Walker A motif, which are both very conserved and important for ATP activity [22, 25]. Moreover, the individual conservation score of the P1051 is only 18.2%, albeit residing in a region with around 60% of conservation score (Table 2.2, Fig. 2.7C). This suggests that at this residue, variation was well tolerated during the protein evolution despite it not being equally tolerated at

the neighboring residues. In a yeast-based experiment, it was shown that the P1051 polymorphism affects valinomycin resistance, when it occurred in diplotype with E22/2677G>T/A [54].

Another SNP that has previously been reported to play a role in drug response is SNP E27/3421T>A (S1141T) (#ns33) and the E27/3322T>C (W1108R) (#ns31) [54, 55]. Based on the assessment using the 3D crystal structure, S1141T and W1108R are found at the C-terminal NBD. S1141 resides at the external surface, whilst the W1108R is found within the NBD interior (Fig. 2.3B, 2.7D, 2.7E, and flash movie http://pfs.nus.edu.sg/demo_src/abcb1.html).

Meanwhile, the E27/3421T>A (S1141T, #ns33) polymorphism, which has a relatively high minor allele frequency of more than 4% in Africans, is not associated with highly conserved residue and protein region. The individual conservation and regional conservation score for this residue is 45.5% and 50.9%, respectively (Table 2.2). This suggests that throughout the protein evolution, there was a good tolerance for a change of residue properties. Another non-synonymous polymorphism, which is less conserved than S1141T, is the E27/3322T>C (W1108R). This SNP, which is also referred as #ns31, corresponds to a less conserved residue (individual conservation score = 27.3%) compared to its surroundings (regional conservation score = 43.5%) (Table 2.2, Fig. 2.7E). Interestingly, both #ns31 and #ns33 are poorly evolutionary conserved and reside in the C-terminal NBD. These SNPs have previously been proposed to be associated with ABCB1 substrate discrimination [54].

Lastly, #ns36 is a non-synonymous polymorphisms that correspond to the V1251I substitution (E29/3751G>A) and has been reported to affect BODIPY-FL-paclitaxel pharmacokinetics [55]. Using the mouse crystal structure, it can be seen that the V1251I is located at the outer side of the C-terminal NBD (Fig. 2.3B and Fig. 2.7F). The individual conservation score of 81.8% signifies a highly conserved residue, which is also associated with conservative amino acid changes in 4 out of 11 species (Table 2.2). This SNP however, is located in a less conserved protein region, with regional conservation score of 59.2%, which suggests a relatively high tolerance level of change in the neighboring residues.

Generally, with the exception of E27/3322T>C (W1108R) (#ns31), most of these SNPs, which have been reported to have association with ABCB1 function alteration, are found at the external surface of the protein, specifically at the C-terminal NBD. These SNPs are #ns25, #ns33 and ns36, which correspond to E26/3151C>G (P1051A), E27/3421T>A and E29/3751G>A (V1251I), respectively (Table 2.2, Fig. 2.3B and flash movie at http://pfs.nus.edu.sg/demo_src/abcb1.html). Furthermore, all of these SNPs have relatively low average conservation scores (below 46%), except for #ns36 that has a relatively high individual conservation score of more than 80%. With the structure information, we can now deduce that all these SNPs are located in a region that have a moderate degree of evolutionary conservation, between 40 and 60%.

2.10 More inclination for nsSNPs to be located at less conserved residues

In this chapter, it can be observed that there is some correlation between the degree of conservation and SNPs. There is a higher percentage of non-synonymous SNPs (12%) versus synonymous SNPs (7%) that are associated with low individual conservation score (20-30%) (Fig. 2.8A). Based on this observation, it can be deduce that there is a tendency for nsSNPs to reside in amino acids that are less conserved. Nonetheless, because SNPs are found throughout both the non-conserved and conserved protein regions, no clear correlation can be deduced between regional conservation scores and the overall presence of SNPs in ABCB1 protein (Fig. 2.8B). Here, it is found that most of the SNPs (39%) that could be mapped to the protein structure correspond to residues having low regional and individual conservation scores (Table 2.2).

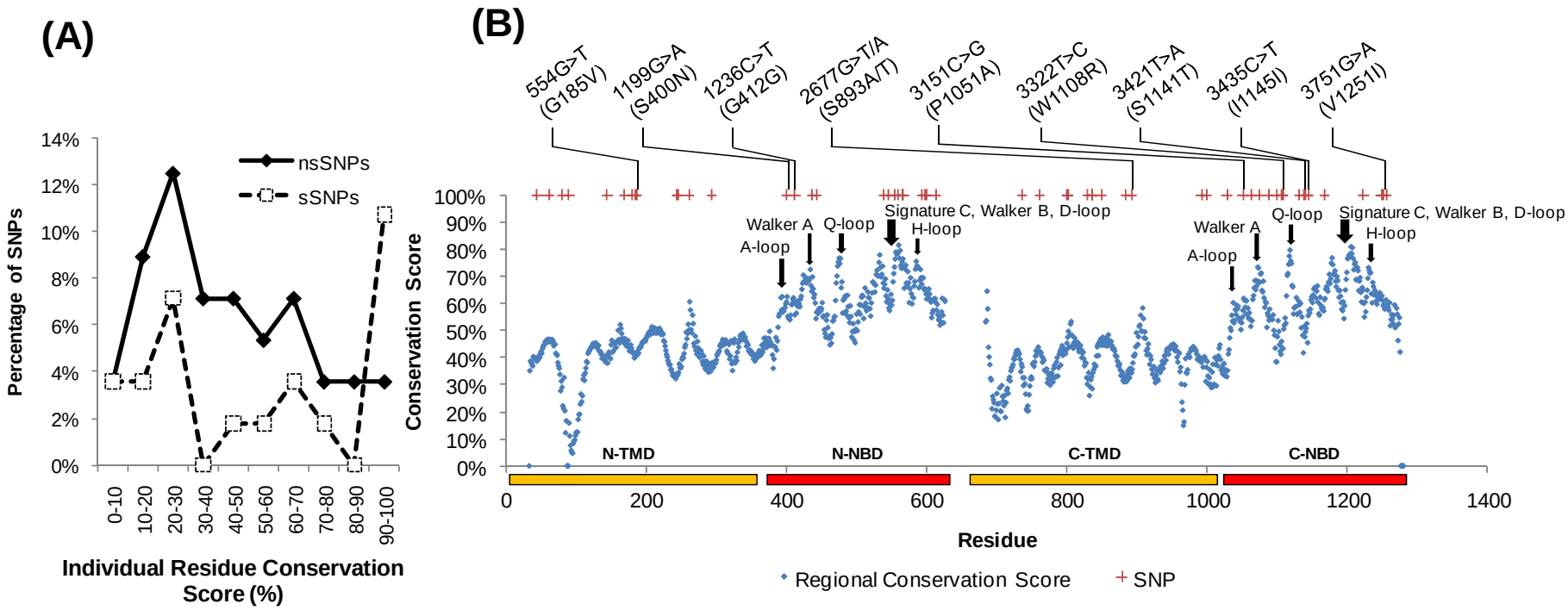


Figure 2.8 Distribution of conservation scores in ABCB1 protein. (A) The distribution of SNP individual conservation scores. (B) Presence of *ABCB*1 coding SNPs as well as the 3D regional conservation scores using the mouse structure. The SNPs that are highlighted in this thesis are presented.

2.11 Conclusion

In conclusion, out of a total of 66 coding SNPs in the ABCB1 protein, there are two sSNP and 12 nsSNPs that have been reported to be associated with protein function variation. There are ten SNPs that could be mapped into the mouse crystal structure and none of these SNPs are found at the region important for substrate binding as illustrated by Aller et al [19]. Here, it was observed that the only one of these SNPs that possibly resides within the lipid bilayer region is #ns7 or the E8/554G>T (G185V) polymorphism. In addition, although residing in a residue that is part of TM10, #ns22 or the E22/2677G>T/A (S893A/T) is located between the cytosolic N-terminal NBD and the membrane region (Fig. 2.3). The NBD can further be suggested to play a significant role in the functioning of ABCB1. This is based on the observation that most of the SNPs that are associated with function differences, are located at the external surface of the C-terminal NBD, not within the internal surface where ATP hydrolysis take place.

From these 14 SNPs, the genotype data of 8 SNPs indicated diverse frequency across various populations. In fact, #ns33 or the E27/3421T>A (S1141T) is only observed in the African American population (Table 2.3). Furthermore, whilst the conservation scores of these 14 SNPs (except for #ns36 and #ns28) are well below average (less than 46%), their regional conservation score only reached average level (40-60%). This indicates that variability is somewhat tolerated in these regions of ABCB1, which may facilitate the variation of substrate specificity of ABCB1 polymorphs.

In recent years, studies involving *ABCB*1 SNPs have grown in number. Many of them focus on studying the SNPs association with drug pharmacokinetics. However, there has been little clarification of the role of these SNPs over variation of drug response. The field is becoming more saturated as differences in methodologies and approaches of various group further confounded the issue. For instance, the difference in populations that are involved in the different studies would add an extra challenge in comparing the results of these studies. Hence, such variable must be addressed in future study design.

Moving forward, structural information of a drug response protein will be extremely useful, especially to enable more understanding of the potential effect of coding SNPs to the protein function. As illustrated in this chapter, 3D localization of the nsSNPs to the protein can aid the generation of a more knowledge-based hypothesis in functional studies. Moreover, using the evolutionary conservation methodology, one is also able to get more insight not only on the potential role of these non-synonymous SNPs, but also the synonymous SNPs that may exert impact in the mRNA level. Therefore, deducing the SNPs location and residue conservation in the 3D crystal structure can arguably provide a more accurate and realistic approach in visualizing the SNPs, in comparison to using the conventional 2D schematic diagram. The study elaborated in this chapter therefore highlighted the feasibility of using this approach in other proteins, given 3D crystal structure is available. It can also extend interpretation of results that are derived from association studies involving SNPs in drug response gene.

Nonetheless, this analysis was limited to only protein-coding SNPs in the *ABCB*1, which is only part of the general breath of polymorphisms that can affect drug response. Hence, in the next chapter, I present the general architecture of SNPs in drug-response pathways, which not only include the coding SNPs, but also SNPs residing outside of the exon region.

2.12 References

1. Weinshilboum, R., *Inheritance and drug response.* N Engl J Med, 2003. **348**(6): p. 529-37.
2. Evans, W.E. and M.V. Relling, *Pharmacogenomics: translating functional genomics into rational therapeutics.* Science, 1999. **286**(5439): p. 487-91.
3. Eichelbaum, M., M. Ingelman-Sundberg, and W.E. Evans, *Pharmacogenomics and individualized drug therapy.* Annu Rev Med, 2006. **57**: p. 119-37.
4. Wang, Z., et al., *A functional polymorphism within the MRP1 gene locus identified through its genomic signature of positive selection.* Hum Mol Genet, 2005. **14**(14): p. 2075-87.
5. Letourneau, I.J., R.G. Deeley, and S.P. Cole, *Functional characterization of non-synonymous single nucleotide polymorphisms in the gene encoding human multidrug resistance protein 1 (MRP1/ABCC1).* Pharmacogenet Genomics, 2005. **15**(9): p. 647-57.
6. Kimchi-Sarfaty, C., et al., *A "silent" polymorphism in the MDR1 gene changes substrate specificity.* Science, 2007. **315**(5811): p. 525-8.
7. Pang, G.S., et al., *Predicting potentially functional SNPs in drug-response genes.* Pharmacogenomics, 2009. **10**(4): p. 639-53.
8. Tusnady, G.E., et al., *Membrane topology of human ABC proteins.* FEBS Lett, 2006. **580**(4): p. 1017-22.
9. Tate, S.K. and S.M. Sisodiya, *Multidrug resistance in epilepsy: a pharmacogenomic update.* Expert Opin Pharmacother, 2007. **8**(10): p. 1441-9.
10. Leschziner, G.D., et al., *ABCB1 genotype and PGP expression, function and therapeutic drug response: a critical review and recommendations for future research.* Pharmacogenomics J, 2007. **7**(3): p. 154-79.
11. Pal, D. and A.K. Mitra, *MDR- and CYP3A4-mediated drug-drug interactions.* J Neuroimmune Pharmacol, 2006. **1**(3): p. 323-39.

12. Vourvahis, M. and A.D. Kashuba, *Mechanisms of pharmacokinetic and pharmacodynamic drug interactions associated with ritonavir-enhanced tipranavir.* Pharmacotherapy, 2007. **27**(6): p. 888-909.
13. Harmsen, S., et al., *The role of nuclear receptors in pharmacokinetic drug-drug interactions in oncology.* Cancer Treat Rev, 2007. **33**(4): p. 369-80.
14. Chapuy, B., et al., *Intracellular ABC transporter A3 confers multidrug resistance in leukemia cells by lysosomal drug sequestration.* Leukemia, 2008. **22**(8): p. 1576-86.
15. Rajagopal, A. and S.M. Simon, *Subcellular localization and activity of multidrug resistance proteins.* Mol Biol Cell, 2003. **14**(8): p. 3389-99.
16. Cascorbi, I., *Role of pharmacogenetics of ATP-binding cassette transporters in the pharmacokinetics of drugs.* Pharmacol Ther, 2006. **112**(2): p. 457-73.
17. Dean, M., A. Rzhetsky, and R. Allikmets, *The human ATP-binding cassette (ABC) transporter superfamily.* Genome Res, 2001. **11**(7): p. 1156-66.
18. Juliano, R.L. and V. Ling, *A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants.* Biochim Biophys Acta, 1976. **455**(1): p. 152-62.
19. Aller, S.G., et al., *Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding.* Science, 2009. **323**(5922): p. 1718-22.
20. Higgins, C.F. and K.J. Linton, *The ATP switch model for ABC transporters.* Nat Struct Mol Biol, 2004. **11**(10): p. 918-26.
21. Sauna, Z.E., et al., *The mechanism of action of multidrug-resistance-linked P-glycoprotein.* J Bioenerg Biomembr, 2001. **33**(6): p. 481-91.
22. Ambudkar, S.V., et al., *The A-loop, a novel conserved aromatic acid subdomain upstream of the Walker A motif in ABC transporters, is critical for ATP binding.* FEBS Lett, 2006. **580**(4): p. 1049-55.
23. Gottesman, M.M., I. Pastan, and S.V. Ambudkar, *P-glycoprotein and multidrug resistance.* Curr Opin Genet Dev, 1996. **6**(5): p. 610-7.

24. Kim, I.W., et al., *The conserved tyrosine residues 401 and 1044 in ATP sites of human P-glycoprotein are critical for ATP binding and hydrolysis: evidence for a conserved subdomain, the A-loop in the ATP-binding cassette.* Biochemistry, 2006. **45**(24): p. 7605-16.
25. Sharom, F.J., *ABC multidrug transporters: structure, function and role in chemoresistance.* Pharmacogenomics, 2008. **9**(1): p. 105-27.
26. Sauna, Z.E. and S.V. Ambudkar, *About a switch: how P-glycoprotein (ABCB1) harnesses the energy of ATP binding and hydrolysis to do mechanical work.* Mol Cancer Ther, 2007. **6**(1): p. 13-23.
27. Sheps, J.A., *Biochemistry. Through a mirror, differently.* Science, 2009. **323**(5922): p. 1679-80.
28. Wang, Z., et al., *Signatures of recent positive selection at the ATP-binding cassette drug transporter superfamily gene loci.* Hum Mol Genet, 2007. **16**(11): p. 1367-80.
29. Wang, Z., et al., *Mining Potential Functionally Significant Polymorphisms at the ATP-Binding- Cassette Transporter Genes.* Current Pharmacogenomics and Personalized Medicine, 2009. **7**(1): p. 40-58.
30. Larkin, M.A., et al., *Clustal W and Clustal X version 2.0.* Bioinformatics, 2007. **23**(21): p. 2947-8.
31. Thompson, J.D., T.J. Gibson, and D.G. Higgins, *Multiple sequence alignment using ClustalW and ClustalX.* Curr Protoc Bioinformatics, 2002. **Chapter 2**: p. Unit 2 3.
32. Leschziner, G., et al., *Exon sequencing and high resolution haplotype analysis of ABC transporter genes implicated in drug resistance.* Pharmacogenet Genomics, 2006. **16**(6): p. 439-50.
33. Kim, R.B., et al., *Identification of functionally variant MDR1 alleles among European Americans and African Americans.* Clin Pharmacol Ther, 2001. **70**(2): p. 189-99.
34. Tang, K., et al., *Genomic evidence for recent positive selection at the human MDR1 gene locus.* Hum Mol Genet, 2004. **13**(8): p. 783-97.

35. Tang, K., et al., *Distinct haplotype profiles and strong linkage disequilibrium at the MDR1 multidrug transporter gene locus in three ethnic Asian populations.* Pharmacogenetics, 2002. **12**(6): p. 437-50.
36. Sai, K., et al., *Haplotype analysis of ABCB1/MDR1 blocks in a Japanese population reveals genotype-dependent renal clearance of irinotecan.* Pharmacogenetics, 2003. **13**(12): p. 741-57.
37. Kroetz, D.L., et al., *Sequence diversity and haplotype structure in the human ABCB1 (MDR1, multidrug resistance transporter) gene.* Pharmacogenetics, 2003. **13**(8): p. 481-94.
38. Leslie, E.M., R.G. Deeley, and S.P. Cole, *Multidrug resistance proteins: role of P-glycoprotein, MRP1, MRP2, and BCRP (ABCG2) in tissue defense.* Toxicol Appl Pharmacol, 2005. **204**(3): p. 216-37.
39. Tombline, G., et al., *Synergy between conserved ABC signature Ser residues in P-glycoprotein catalysis.* J Biol Chem, 2004. **279**(7): p. 5363-73.
40. Fung, K.L. and M.M. Gottesman, *A synonymous polymorphism in a common MDR1 (ABCB1) haplotype shapes protein function.* Biochim Biophys Acta, 2009. **1794**(5): p. 860-71.
41. Tsai, C.J., et al., *Synonymous mutations and ribosome stalling can lead to altered folding pathways and distinct minima.* J Mol Biol, 2008. **383**(2): p. 281-91.
42. Kyte, J. and R.F. Doolittle, *A simple method for displaying the hydropathic character of a protein.* J Mol Biol, 1982. **157**(1): p. 105-32.
43. Marzolini, C., et al., *Polymorphisms in human MDR1 (P-glycoprotein): recent advances and clinical relevance.* Clin Pharmacol Ther, 2004. **75**(1): p. 13-33.
44. Choi, K.H., et al., *An altered pattern of cross-resistance in multidrug-resistant human cells results from spontaneous mutations in the mdr1 (P-glycoprotein) gene.* Cell, 1988. **53**(4): p. 519-29.

45. Omote, H., et al., *Improved energy coupling of human P-glycoprotein by the glycine 185 to valine mutation.* Biochemistry, 2004. **43**(13): p. 3917-28.

46. Omote, H. and M.K. Al-Shawi, *Interaction of transported drugs with the lipid bilayer and P-glycoprotein through a solvation exchange mechanism.* Biophys J, 2006. **90**(11): p. 4046-59.

47. Kimchi-Sarfaty, C., J.J. Gribar, and M.M. Gottesman, *Functional characterization of coding polymorphisms in the human MDR1 gene using a vaccinia virus expression system.* Mol Pharmacol, 2002. **62**(1): p. 1-6.

48. Sakurai, A., et al., *Quantitative structure--activity relationship analysis and molecular dynamics simulation to functionally validate nonsynonymous polymorphisms of human ABC transporter ABCB1 (P-glycoprotein/MDR1).* Biochemistry, 2007. **46**(26): p. 7678-93.

49. Woodahl, E.L., et al., *MDR1 G1199A polymorphism alters permeability of HIV protease inhibitors across P-glycoprotein-expressing epithelial cells.* AIDS, 2005. **19**(15): p. 1617-25.

50. Woodahl, E.L., et al., *MDR1 (ABCB1) G1199A (Ser400Asn) polymorphism alters transepithelial permeability and sensitivity to anticancer agents.* Cancer Chemother Pharmacol, 2009. **64**(1): p. 183-8.

51. Woodahl, E.L., et al., *Multidrug resistance gene G1199A polymorphism alters efflux transport activity of P-glycoprotein.* J Pharmacol Exp Ther, 2004. **310**(3): p. 1199-207.

52. Crouthamel, M.H., et al., *A novel MDR1 G1199T variant alters drug resistance and efflux transport activity of P-glycoprotein in recombinant Hek cells.* J Pharm Sci, 2006. **95**(12): p. 2767-77.

53. Green, H., et al., *ABCB1 G1199A polymorphism and ovarian cancer response to paclitaxel.* J Pharm Sci, 2008. **97**(6): p. 2045-8.

54. Jeong, H., et al., *Function-altering SNPs in the human multidrug transporter gene ABCB1 identified using a Saccharomyces-based assay.* PLoS Genet, 2007. **3**(3): p. e39.

55. Gow, J.M., et al., *Substrate-dependent effects of human ABCB1 coding polymorphisms.* J Pharmacol Exp Ther, 2008. **325**(2): p. 435-42.

Chapter 3. Architecture of SNPs in Drug Response Pathways

3.1 Introduction

Drug response variation is affected by individual's genetic background [1-3]. Elucidating the role of SNPs residing in drug-response genes is a central theme in many pharmacogenomics studies [4-7]. Moreover, as shown in the previous chapter, coding SNPs could affect protein 3D structure, which can potentially affect the protein function. There have been studies supporting the potential importance of protein-coding region SNPs (cSNPs) in drug response [5-6, 8-10]. However more recently, there has been an increasing awareness of the impending potential consequence of SNPs that reside on non-coding or regulatory regions despite their lack of effects on protein structure [11-13]. The potential functional effect of non-coding polymorphisms can be exerted in gene expression level (for regulatory SNPs - rSNPs) such as by affecting transcription factor (TF) binding or miRNA binding; or in RNA structural level (for structural RNA SNPs - srSNPs) by affecting splicing.

In this chapter, utilising the potentially functional SNPs (pfSNP^{TM} - http://pfs.nus.edu.sg/) resource [14], the architecture of rSNPs and srSNPs in addition to that of cSNPs in genes responsible for drug response were elucidated. In addition, the correlation pattern between SNPs genotype and changes of drug-response gene expression level was also investigated.

The relationship between drug response variation and population differentiation in SNP allele frequency has been discussed extensively [2, 15-16]. International consortiums that have publicized genotype data of major world populations include the HapMap, which genotyped SNPs in 11 populations and the Singapore Genome Variation Project (SGVP), which genotyped SNPs in three Asian populations in Singapore, a multiracial city-state [17-18]. Using this publicly available information, this chapter also reports the pattern of population genetics differentiation across conventional drug-response genes. This approach is arguably effective in studying the genetic background that is important in drug responses, especially when a clinical trial was done in populations that constitute a different genetic background.

Moreover, it is important to note that in many drugs, the PK/PD is a multigenic process [19-20]. However, many studies were conducted based on a candidate gene or candidate SNP approach, which could disregard other potentially important genetic factors affecting the drug response. Nonetheless, there has been an effort to systematically organize drug-response genes into biological pathways where their specific role on the drug PK/PD is well annotated [20-21]. These drug-response pathways (DRPs) represent diverse networks of PK/PD genes that are associated with various drug responses. Genes within a DRP may interact and regulate the overall therapeutic outcome through various roles such as in drug absorption, distribution, metabolism, and excretion.

In this chapter, I designed a systematic analysis of SNPs in the DRPs using a gene functional region-directed approach, one that could reliably detect pattern of

SNPs distribution in genes that belong to various DRPs. To my knowledge, at the time this study was initiated, a systemic investigation of the pattern of SNP architecture such as the gene regional distribution and predicted functional effect was not yet available.

3.2 Methods

3.2.1 Drug-response genes & pathways

The drug pathways in this study were mined from the PharmGKB database (http://www.pharmgkb.org) [21]. Sixty six pathways that contain genes information in the PharmGKB database were used to assemble the set of 715 drug-response genes that were used for analysis. Pathways that are associated with similar drugs were combined, which made up the 41 drug-response pathways (DRPs) used in this analyses (Appendix 2).

3.2.2 Mapping of SNPs to gene region

The SNPs data were obtained from the NCBI dbSNP (build 131) database [30]. A total of 10,512,313 SNPs from 22,333 genome genes were extracted for this purpose. Out of this total, 497,736 SNPs belong to the 715 drug-response genes. The work in this chapter employed the gene functional region-directed approach in mapping SNPs to the gene sets, in accordance to NCBI Genome build 37. SNPs

are categorised based on their location in genes. For SNPs in non-coding regions, the following classification was applied: *Promoter* for SNPs residing within 5.5Kb upstream of a gene transcription start site; *Intronic* for SNPs residing in introns; as well as *5' UTR* and *3' UTR* for SNPs residing in the 5' or 3' un-translated mRNA regions. In the coding region (i.e. exons), SNPs that cause amino acid substitution during mRNA translation is classified as non-synonymous SNPs (nsSNPs). On the other hand, silent or non-amino acid-substituting SNPs are referred as synonymous SNPs (sSNPs). SNP density was calculated on transcript level and its average was used to define the gene SNP density.

Using the gene functional region-directed approach the SNP density of each gene transcript was calculated according to the following formula:

$$\text{SNP Density (SNPs/Kbp)} = \frac{n \text{ number of SNPs}}{\text{region length (basepairs)}} \times 1{,}000$$

3.2.3 Potentially Functional SNPs

In studying the SNPs potential functional implication, the information in the Potentially Functional SNPs database (pfSNP^{TM} - http://pfs.nus.edu.sg/) was utilized [14]. The pfSNP reseource integrated SNPs functional prediction tools into a one-stop portal of SNPs predicted functional anottation. The predicted functional features in this resource was used to define a SNP into different functional categories: Transcription factor (TF) binding sites, miRNA binding sites, 3' UTR conserved regions, splicing regulatory sites, nonsense-mediated

decay (NMD), codon usage differentiation, protein deleterious, post-translational modification sites, and protein domains. These are summarized in Table 3.1. Using the gene functional region-directed approach of SNP mapping, the SNPs' potential functional effect could be derived and the proportion of potentially functional SNPs in each gene were calculated.

Table 3.1 Description of tool used for SNP functional categories.

Functional category	Tools Description
TF Binding Sites	TF binding site changes
miRNA Binding Sites	miRNA binding site changes
3' UTR Conserved	3' UTR conserved regions
Splicing Regulatory Sites	Exonic splicing enhancer/silencer (ESE/ESS), intronic splicing regulatory element (ISRE), abberrant splice sites
Nonsense-mediated Decay	Nonsense-mediated decay (NMD) sequence changes
Codon Usage Differentiation	Codon usage differentiation sequence changes
Protein Deleterious	Polyphen/SNP34/LS-SNP predicted deleterious amino acid substitution
Post-translational Modification Sites	Glycosylation and phosphorylation sites
Protein Domains	Transmembrane domain, Interpro Scan

3.2.4 eQTL analysis

Gene expression data was obtained from the Gene Expression Omnibus (GEO) database. The study in this chapter utilised lymphoblastoid cell line (LCL) expression data from 144 (43 CEU, 59 CHB-JPT, and 42 YRI) unrelated HapMap

individuals performed by Stranger et al (Series GSE6536) [31]. In order to capture as many SNPs that could be correlated with differential gene expression, a linear regression was performed involving the expression data and their matched-individuals genotype from three populations (CEU, CHB-JPT and YRI) of the 1000 Genomes Project pilot phase [32]. The analysis of eQTL was then performed with 2.9 million, 2.4 million and 3.6 million SNP-mRNA probe pairs in the CEU, CHB-JPT and YRI, respectively. A SNP is categorised as expression-associated or eQTL if its false discovery rate (FDR)-corrected *P*-value is less than 0.05 after performing linear regression. The eQTL variants (n = 37,756) were subsequently mapped according to the gene functional region-directed method employed in this study.

3.2.5 Population differentiation estimation

The allele frequency was calculated using genotype data from two sources. The first originated from the HapMap (Release 27), consisting of 1.4 million SNPs that were genotyped from more than a thousand individuals in 11 populations [17]. The second came from the Singapore Genome Variation Project (SGVP), which genotyped approximately 1.4 million SNPs in 292 individuals from three Asian populations [18].

The allele frequency data was used to calculate the estimate measure of SNP population differentiation using F_{ST} statistics [22]. Only common SNPs that have been genotyped in unrelated individuals from all populations were used in the calculation of F_{ST}, which for this chpater, measure population differentiation estimates across all populations. In this analysis, a total of 10 populations from the two genotyping sources were combined into four continental groups: Africans (consisting of LWK and YRI), Europeans (consisting of CEU and TSI), East Asians (consisting of CHB, CHD, CHS and JPT), and South Indians (consisting of GIH and INS). Pairwise F_{ST} scores within populations belonging to the same continental group are all 0.01 or less. An extremely differentiated SNP is defined as one having F_{ST} score within the top 5% of the whole-genome overall population F_{ST} distribution. The pathway population differentiation was analyzed by determining the proportion of genes that carry one or more extremely differentiated SNP in the DRP.

3.2.6 Random sampling simulation

The statistical random sampling simulation was performed in the 'R' environment (http://www.r-project.org/). For each DRP, the same number of genes was sampled and the sampling criteria were set such that only genes having approximately similar length (size of transcripts are within the range of what is observed in the pathway genes set) to the DRP genes were considered. Simulations were run independently for the analysis involving SNP density,

proportion of potentially functional SNPs, proportion of genes with eQTL, and proportion of genes with high-F_{ST} SNPs.

Each simulation required 10,000-time random sampling repeats before an empirical distribution was formed in the individual DRP. Using these results; the percentile (*Pc*) value of each DRP was calculated using the following equation:

$$Pc \text{ value} = \frac{h}{(h + l + e)}$$

Where:

h = number of sampling observations higher than expected

l = number of sampling observations lower than expected

e = number of sampling observations equal to what expected

A *Pc* value < 0.05 determined significantly non-random observation where 95 percent of the sampling results fall below the observed value of the DRP. Whereas a Pc value > 0.95 depicted the opposite.

3.2.7 Drug pathway priority score

For each DRP, the '*Py score*' was calculated using their SNP architecture signatory parameters that were gathered in this study (Table 3.2). The *Py score* was defined according to the following equation.

$$\text{Py Score} = \frac{\frac{a}{9} + \frac{b}{6} + \left(\frac{c}{6} * 1.5\right) + \left(\frac{d}{9} * 2\right) + \left(\frac{e}{9} * 2\right) + \left(\frac{f}{6} * 3\right) + \left(\frac{g}{9} * 4\right) + (\hat{e} * 2.5) + (\hat{f} * 3.5) + (\hat{g} * 4.5)}{25}$$

Table 3.2 Features used for calculating drug pathways prioritization (Py) scores.

Variable	Description	Weight
a	Potentially Functional SNPs - *No. of significant categories*	1
b	Expression SNPs - *No. of significant categories*	1
c	Population Differentiation - *No. of significant categories*	1.5
d	Expression-associated Potentially Functional SNPs - *No. of categories with SNPs*	2
e	Highly-differentiated Potentially Functional SNPs - *No. of categories with SNPs*	2
$\hat{e}$	Highest F_{ST}	2.5
f	Highly-differentiated Expression-associated SNPs - *No. of categories with SNPs*	3
$\hat{f}$	Highest F_{ST}	3.5
g	Highly-differentiated & Expression-associated Potentially Functional SNPs - *No. of categories with SNPs*	4
$\hat{g}$	Highest F_{ST}	4.5

The *Py score* was derived based on the following rationale. The first three variables: *a*, *b* and *c* were derived based on the question whether a pathway is associated with one or more SNP category enrichment that is above the statistical threshold (*Pc* value < 0.05). Variable *a* recognizes the presence of enrichment of potentially functional SNPs in up to nine pfSNP categories (Table 3.1), which also explained the use of 9 as denominator. The subsequent variable, *b*, takes into account the observed enrichment of expression-associated SNPs in up to 6 SNP categories (Promoter, 5' UTR, 3'UTR, Intron, Non-synonymous, and Synonymous). Variable *c* recognizes the presence of enrichment of extremely population-differentiated SNPs in up to 6 SNP categories. Whilst variables *a* and *b* carried an equal weightage of 1, variable *c*, due to its association with population differentiation, was given a heavier weightage of 1.5.

In addition to recognizing the presence of SNP enrichment in the above variables, the subsequent part of the equation accounted the presence of SNPs that can be categorized into two or more categories, hence were given twice the weightage of the above variables. These are the potentially functional SNPs that are associated with variation in gene expression (variable *d*) or high population differentiation F_{ST} scores (variable *e*). Variable *f*, which was given a greater weightage, recognizes the presence of potentially functional SNPs that are also associated with gene expression variation. Moreover, the rarity of potentially functional SNPs that are also associated with gene expression and extreme population differentiation was given an even higher weightage of 4 in variable *g*.

The last three variables, $\hat{e}$, $\hat{f}$ and $\hat{g}$ accounted the highest F_{ST} score for SNPs that are predicted to be potentially functional ($\hat{e}$) and associated with gene expression ($\hat{f}$), in addition to expression-associated SNPs that are predicted to be potentially functional ($\hat{g}$). In taking into account the relevance of their population differentiation status, greater weightages were given for these variables.

3.3 Results

3.3.1 SNP enrichment in drug-response pathways (DRPs)

To evaluate the general distribution of SNPs in 715 drug-response genes retrieved from the PharmGKB database, the SNP densities were calculated based on gene functional region-directed approach (see methods). Each SNP is mapped to their respective genes and classified according to the gene functional region where they are located. SNPs in the non-coding region are classified as Promoter (if it is located within 5.5 Kbp upstream of the transcription start site), 5' un-translated region (UTR), 3' UTR, or intronic. SNPs in coding regions are categorized as either non-synonymous (nsSNPs, amino acid-substituting) or synonymous (sSNPs, non-amino acid-substituting). Table 3.3 presents the SNP densities of drug-response genes most commonly found in the DRPs.

The whole-genome median SNP densities are 7.6, 3.1, 7.6, 6.9, 3.7, and 2.2 SNPs/Kbp in the Promoter, 5' UTR, Intronic, 3' UTR, nsSNPs and sSNPs categories, respectively. When compared to these values, we observed that the coding and non-coding regions of the drug-response genes are equally SNP-

enriched (Fig. 3.1A). In fact, SNPs enrichment in the 5' UTR and synonymous SNP categories are more than 1.5 fold (P-value < 0.001, Mann-Whitney U test). SNP enrichment is therefore a widespread occurrence that occurs not only in the protein encoding region but also in the gene expression regulatory region of drug-response genes.

Table 3.3 SNP density of the most common genes in drug-response pathways.

Gene	NCBI ID	Function	No. of Associated Pathways	SNP Density^					
				Promoter	5' UTR	Intronic	3' UTR	Non-synonymous	Synonymous
CYP3A4	1576	Metaboliser	25	12.55	28.85	10.72	18.23	22.49	2.65
CYP3A5	1577		17	8.91	68.97	8.01	27.03	11.93	1.99
ABCB1	5243	Transporter	15	10	19.14	7.94	29.46	12.23	5.98
CYP2C19	1557	Metaboliser	14	11.09	0	10.6	0	23.08	6.11
CYP2C9	1559		14	14.18	0	12.79	22.04	22.4	7.47
CYP2D6	1565		10	22.36	22.22	36.01	0	54.25	16.9
ABCG2	9429	Transporter	10	21.82	18.26	9.66	4.06	8.13	3.05
CYP1A2	1544	Metaboliser	9	9.09	0	16.41	11.9	19.34	7.09
ABCC2	1244	Transporter	8	10.36	14.39	13.25	7.3	8.41	4.74
UGT1A1	54658	Metaboliser	8	14.36	0	13.21	13.51	21.85	1.87
CYP2C8	1558		7	7.64	10.53	10.06	11.24	12.22	1.36
CYP2B6	1555		6	9.09	0	17.01	36.33	24.39	7.45
ABCC1	4363	Transporter	6	7.09	0	12.15	14.5	4.66	7.16
ABCC3	8714		6	5.09	0	7.45	6.81	6.18	3.89
MAPK1	108	Cellular Signalling	5	4.36	0	8.21	4.98	1.53	3.36
ADCY2	5594	cAMP formation	5	7.82	8.33	7.96	6.79	1.85	5.54
UGT2B7	7364	Metaboliser	5	11.45	0	10.12	35.86	6.92	10.06

^SNP density is in SNPs/Kbp. The whole-genome median SNP densities are 7.6, 3.1, 7.6, 6.9, 3.7, and 2.2 SNPs/Kbp in promoter, 5' UTR, intronic, 3' UTR, nonsynonymous and synonymous categories, respectively

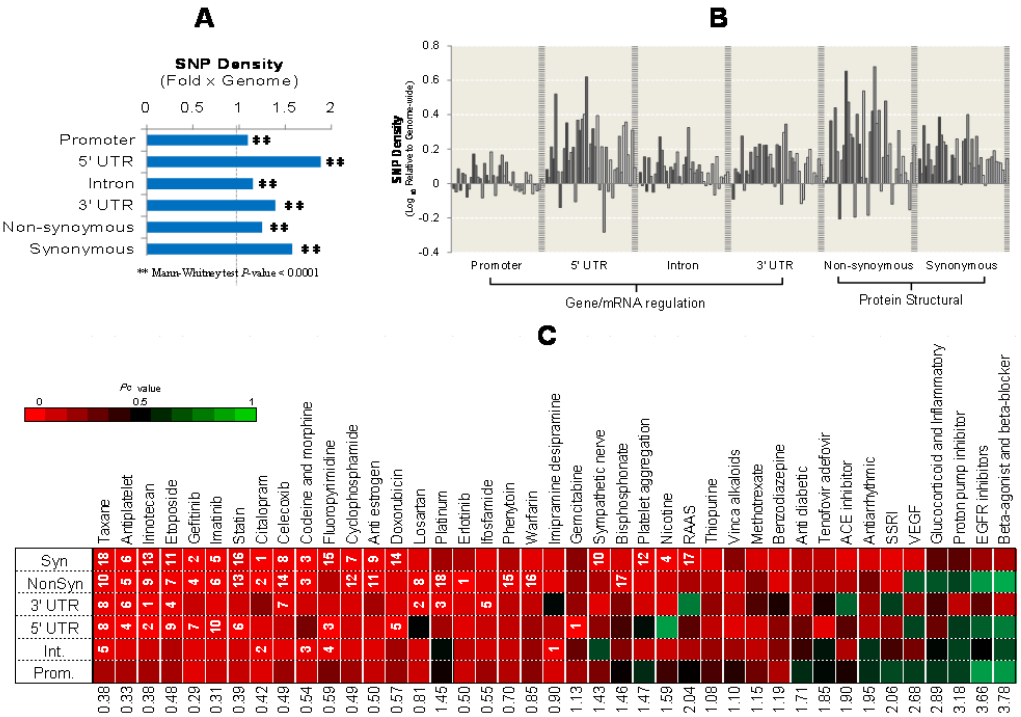


Figure 3.1 SNP enrichment in drug-response pathways (DRPs) is seen extensively across all gene functional regions. As shown by the median (fold versus genomic median) SNP density, genes taking part in drug-response (n = 715) are generally highly polymorphic from the promoter and intron to the coding and un-translated (UTR) regions (A). This enrichment of SNPs is also seen in the DRPs (B) where each bar represents DRP SNP enrichment in the gene functional region of interest. A heat map of DRP *Pc* values showed that many DRP *Pc* values are inclining towards zero, which signifies non-random enrichment of SNPs (C). In each DRP, the numbers in the blocks represent the enrichment rank associated with the specific SNP categories (ie. Syn, NonSyn, 3'UTR, 5'UTR, Intron, and Promoter).

As different drugs may target different tissues, metabolized and transported by diverse genes, we questioned whether the general SNPs enrichment observed in these genes could translate into significant projections of polymorphisms in the DRPs that are specific to certain drug types. In order to reduce background noise, the 66 PharmGKB PK/PD pathways information was summarized into 41 DRPs (Appendix 2). Out of 715 drug-response genes found in the DRPs collection 17 -

tioare associated with more than 5 pathways (Table 3.3). In fact the drug metabolizing enzyme, CYP3A4, is commonly associated with 25 pathways, whereas 15 pathways conceal the influential ABCB1 transporter. The individual DRP average SNP densities were calculated using the same gene functional region-directed approach (Appendix 3). The result indicates that in general, the DRPs are SNP-enriched in both the expression regulatory and protein coding SNP categories.

Because of the involvement of multiple genes and pathways, an early concern was that such observation could be affected by statistical randomness. Hence for each DRP, a 10,000-time statistical sampling simulation was performed using random genes. An evaluation if the same SNP enrichment was also observed in the DRP random sampling data set (see methods) was then carried out. The results showed that in the majority of DRPs, such SNP-enrichment was not observed in the random sampling set (Fig. 3.1C). This suggests that the SNP enrichment in the DRPs is not a random observation. Out of 41 DRPs, 27 pose a significant *Pc* value in one or more SNP categories representing SNP-enrichment. The DRP *Pc* values – the percentile score of which an observed pathway SNP density falls within its own empirical distribution – across all gene functional regions are well cumulated into the significant range (Fig. 3.2). Highly polymorphic pathways include those associated with taxane, antiplatelet, irinotecan and etoposide drugs, where significantly non-random SNP enrichments were observed in four or more SNP categories.

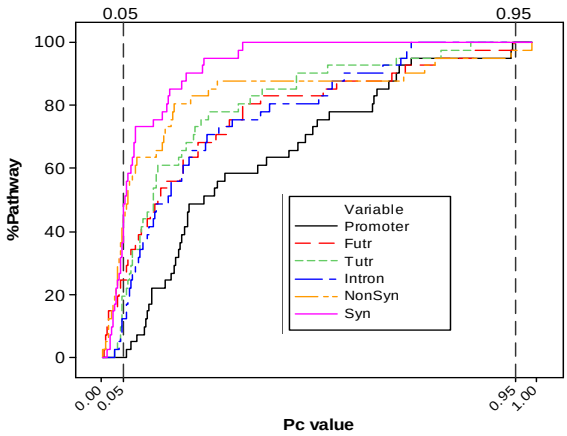


Figure 3.2 10,000-time statistical sampling simulation with random genes of comparable size showed that SNP enrichment across the DRPs are not random.

3.3.2 Potentially functional SNPs in DRPs

In order to elucidate the architecture of SNPs potential functional effect, the detailed distribution of SNPs in the DRPs were analyzed based on their predicted implication for gene functions. Figure 3.3A shows the proportion of potentially functional SNPs in the DRPs based on three functional levels of a SNP's effect: expression regulatory (rSNPs), structural RNA (srSNPs) or protein level (cSNPs). The result shows the apparent multi-level functional role of SNPs in drug-response, with potential effects exerted in all levels of the central dogma of molecular biology; from gene expression and RNA structure levels to protein structure and function levels.

In this detailed SNP architecture, the relatively high proportion of potentially functional SNPs in the DRPs was also revealed (Fig. 3.3A). For the regulatory SNPs (rSNPs) category, the potential effect on transcription factor (TF) binding

sites is highly prominent compared to that of miRNA binding sites. In the structural RNA SNPs (srSNPs) category, a potentially strong influence is exerted in RNA splicing regulatory sites, but not in the other functional prediction categories (codon usage, 3' UTR conservation or nonsense-mediated RNA decay). Furthermore, 10,000-time statistical sampling simulation showed that the high proportion of DRP SNPs in the two functional categories, TF binding and splicing regulatory sites alteration, are among the ones with the lowest trend of statistical *Pc* values (Fig. 3.3B). This suggests that the high proportion of TF binding and splicing regulatory site SNP categories is not due to random chances. On the other hand, in the protein structure and function level, where it is possible to observe coding SNPs (cSNPs), a substantial presence of non-synonymous SNPs was observed to be associated with protein domains or deleterious sites, but not post-translational modification sites (Fig. 3.3A and 3.3B).

This result suggests for the high prominence of regulatory polymorphisms in drug response, which was traditionally less popular than coding SNPs in pharmacogenomics. Furthermore, it also signifies a novel pattern of pharmaco-SNPs functionality, which is now attributed not only by protein variants, but also by the high prevalence of expression regulatory and RNA SNPs.

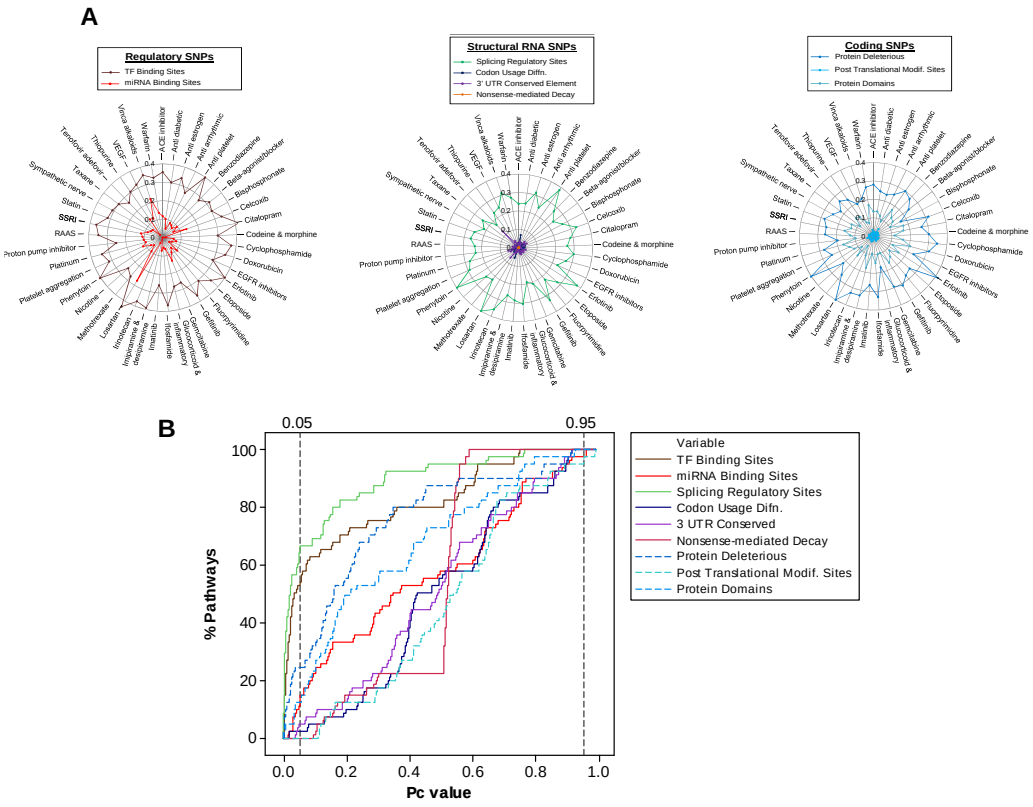


Figure 3.3 Potentially functional SNPs in DRPs. (A) The signature of SNPs with potential function in drug-response is marked by the relatively large scale of regulatory (rSNPs) and structural RNA-affecting SNPs (srSNPs), in addition to coding SNPs (cSNPs). In general, the DRPs carry substantial proportion of SNPs that are predicted to affect gene functional sites. (B) Among the most significant and non-random enrichments (Low *Pc* values) are in SNPs that could affect transcription factor (TF) binding, gene splicing and deleterious amino-acid substitutions.

3.3.3 Expression quantitative loci (eQTL) is linked to rSNP and srSNPs in the DRPs

This part seeks to evaluate whether potentially functional SNPs in DRPs are associated with actual differences in gene expression, a cellular phenotype that can be attributed to genetic variations. A correlation analysis between genotype and gene expression data of HapMap individuals was performed. The expression

data was obtained from gene expression microarray of lymphoblastoid cell lines (LCLs) (Series GSE6536 of the Gene Omnibus Database). The matching genotype of the same individuals was obtained from the 1000 Genomes pilot phase data. A SNP that is associated with differential local gene expression or eQTL is defined as one with false discovery rate (FDR) corrected *P*-value of less than 0.05 following linear regression analysis.

When taking into account SNPs potential function, it was observed that the proportion of SNPs that are associated with gene expression is highest in the rSNPs and srSNPs categories (Fig. 3.4). In addition, we also observed a relatively higher proportion of genes carrying expression-associated rSNPs that alter TF binding sites and srSNPs that alter splicing regulatory sites (Fig. 3.5 and Fig. 3.6). This result further accentuates the raising importance of rSNPs and srSNPs in drug-response. Among the DRPs that carry substantial proportion of this type of genes are those that are responsible for drugs such as the methotrexate, thiopurine and doxorubicin (Fig. 3.6).

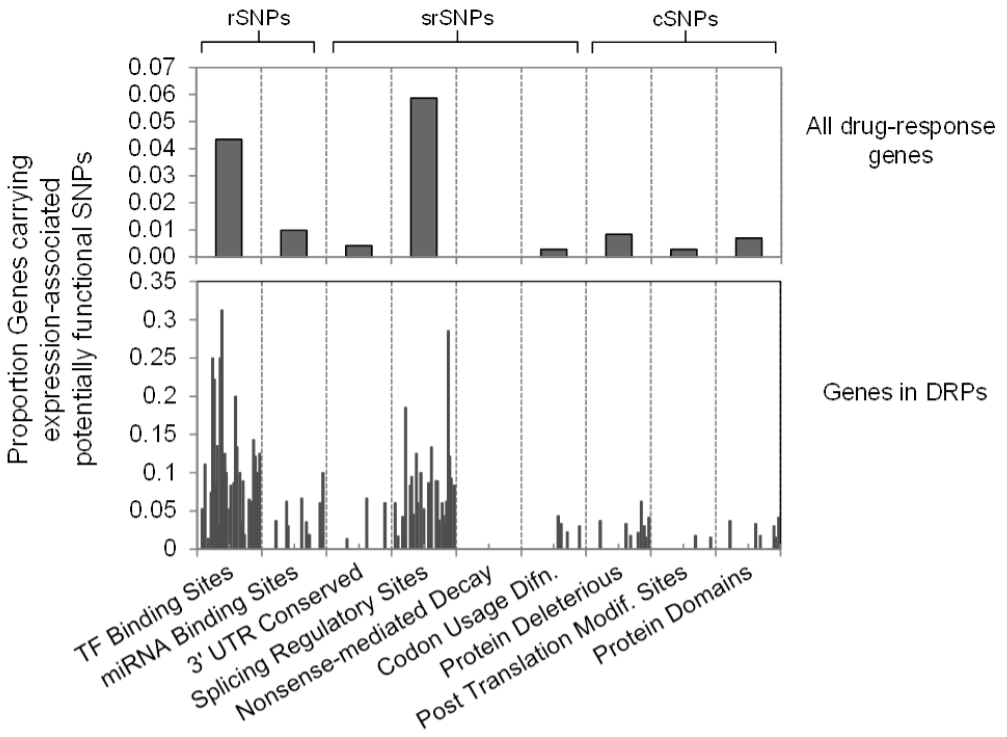


Figure 3.4 Higher proportion of DRP genes carrying TF binding and splicing regulatory site SNPs that are associated with differential gene expression. (Upper panel) The bar represents the collection of 715 drug-response genes. (Lower panel) A bar within the SNP category represents one DRP.

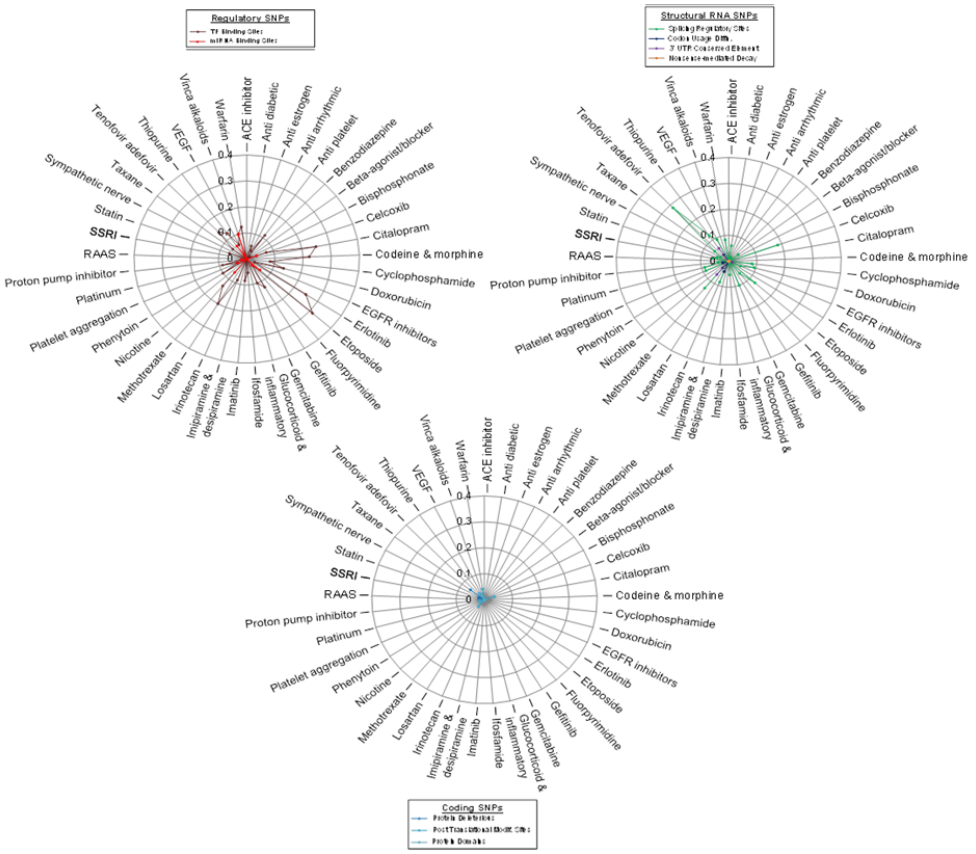


Figure 3.5 Proportion of genes carrying SNPs as expression quantitative loci (eQTL) in DRPs.

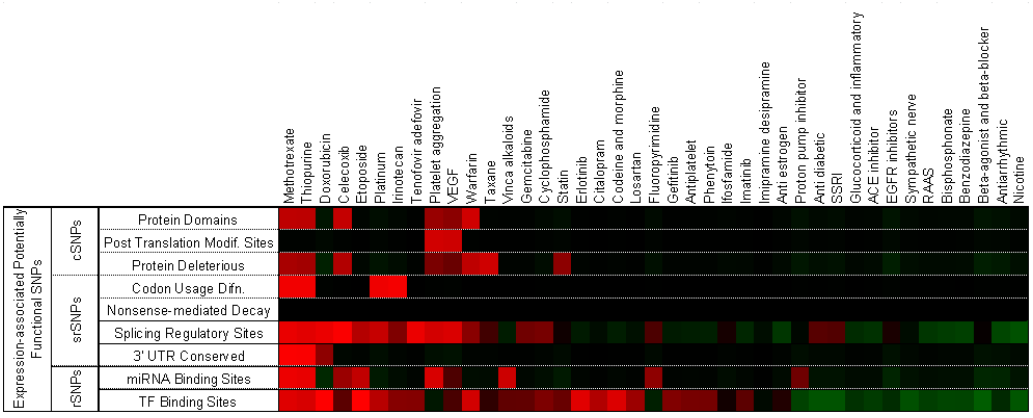


Figure 3.6 Pc values for the proportion of genes carrying SNPs as expression quantitative loci (eQTL) in DRPs.

The result also suggested on the possible co-regulation of drug response by regulatory SNPs that are associated with gene expression. It can presumably take place in DRPs that are unrelated, yet carrying common drug transporters or metabolisers. Out of 41 DRPs, 28 (68%) and 20 (49%) have – in their common genes – contained one or more TF-affecting and splicing-affecting SNPs that are also eQTL in nature, respectively (Appendix 4). Figure 3.7A presents a schematic view of a presumed DRPs co-regulation phenomenon by variants in the two most common genes having expression-associated SNPs in their TF binding sites. eQTL functional SNPs in the promoter of multi-drug resistance gene, *ABCB*1 (rs3747802), as well as phase II metabolising enzyme, *UGT*1*A*1 (rs10929302), could co-regulate 18 different drug pathways through its effect in gene expression regulation.

Figure 3.7 Co-regulation of drug response by common regulatory variants in drug transporters and metabolizers. Expression-associated promoter TF binding site SNPs in ABCB1 and UGT1A1 (A) as well as extremely population-differentiated splicing SNPs in ABCG2 and CYP3A5 (B) serve in regulating multiple DRPs (encapsulated by ovals). These are the most co-shared genes with prospective functional SNPs associated to either differential gene expression or extreme population difference in the DRPs.

3.3.4 High population differentiation in rSNPs and srSNPs of the DRPs

Response to drug therapies are known to vary between different people, especially when they originate from populations of different genetic backgrounds. Hence the subsequent question was whether there is a distinct population differentiation pattern of SNPs within genes of the DRPs. In analyzing the pattern of population differentiation in the DRPs, SNP allele frequencies were derived from the HapMap (Phase 3) and Singapore Genome Variation Project (SGVP) genotype data. The allele frequency data were compiled into four major continental groups: Africans, Europeans, East Asians, and South Asians (see methods). For each SNP, the F_{ST} score was calculated and used to estimate the degree of population differentiation using all the available populations [22]. Based on this data, the extremely population-differentiated SNPs, are the SNPs in the top 5% of the whole-genome F_{ST} distribution. These SNPs in the top 5% are observed to have $F_{ST} > 0.3$ (Fig. 3.8).

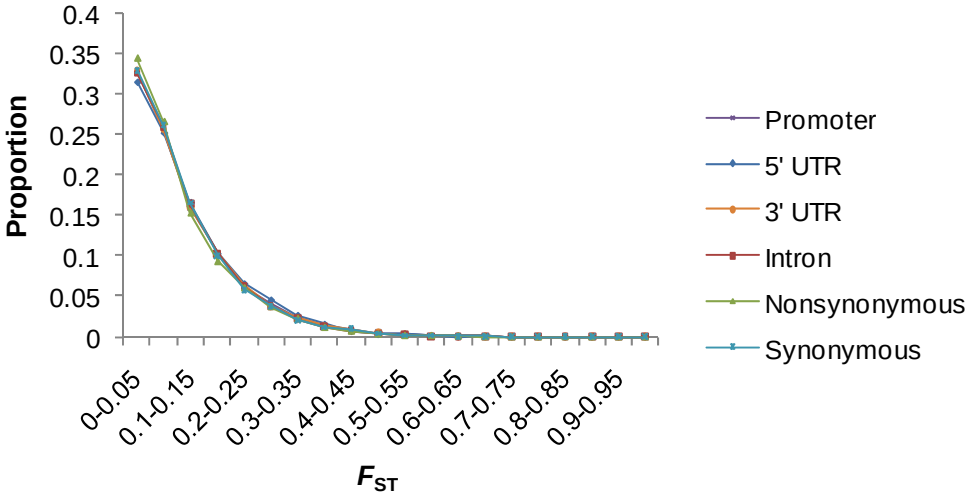


Figure 3.8 The distribution of FST scores of SNPs in the human genome.

It could be observed that the majority (78%) of DRPs carry a substantial proportion of genes that house one or more extremely population-differentiated SNPs, higher than the genome-wide average (proportion genome-wide genes with high-F_{ST} SNPs = 0.27) (Fig. 3.9A). Following 10,000-time statistical random sampling, 11 of the 41 DRPs pose a *Pc* value of less than 0.05. This includes DRPs responsible for the beta-agonist/blocker, antiarrhythmic, bisphosphonate, etoposide, and statin drugs.

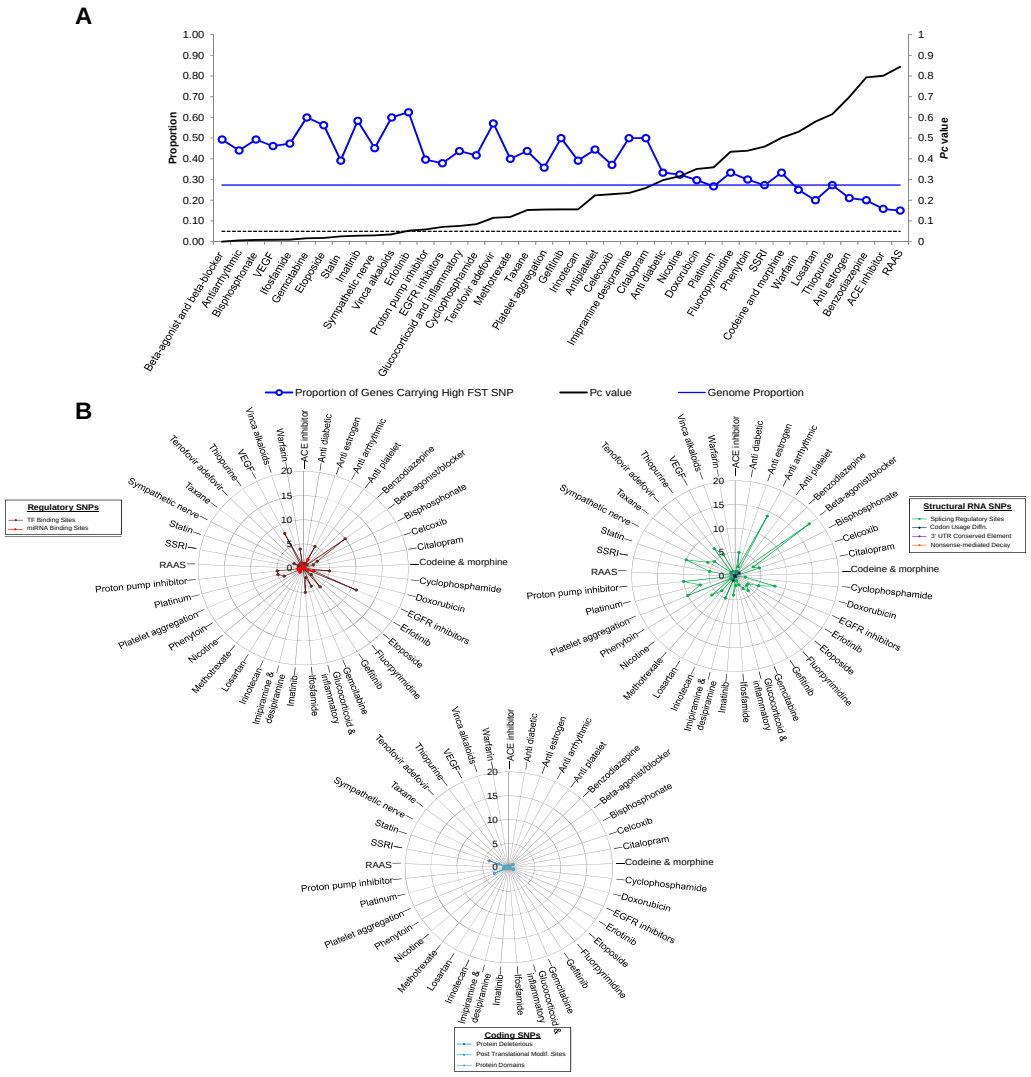


Figure 3.9 Population differentiation in the DRPs. (A) The proportion of genes (blue line) carrying one or more highly differentiated SNPs (in top 5% of distribution) and their *Pc* values (black line) across the DRPs. (B) High magnitude of highly population-differentiated TF binding and splicing regulatory sites-affecting SNPs in DRPs.

Furthermore at the SNP functional level, the magnitude of highly population-differentiated potentially functional SNPs in the DRPs are more pronounced in TF binding site-affecting rSNPs and splicing regulatory site-affecting srSNPs (Figure 3.9B) across the DRPs, including in those associated with beta-agonist/blocker,

antiarrhythmic and statin drugs. However, the gene functional region-directed random simulation showed that the pattern of population differentiation in many DRPs is less uniform (Fig. 3.10). This is despite the relatively more prominent population differentiation in the non-coding category (i.e. Intronic SNPs), but not in the coding SNP category.

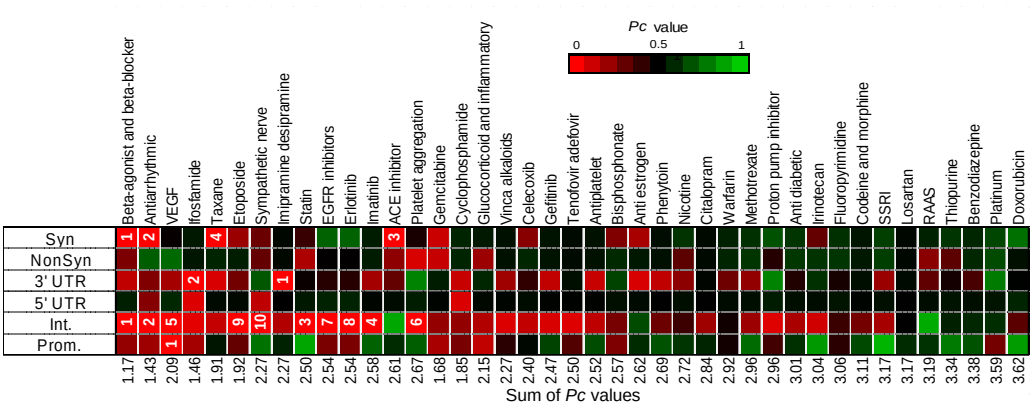


Figure 3.10 High population differentiation can be more obviously seen in the non-protein coding regions such as the Intron and UTR categories. A heat map of DRP *Pc* values obtained from 10,000-time statistical sampling. *Pc* value approaching 0 indicates a significant and non-random proportion of genes carrying highly differentiated SNPs. In each DRP, the numbers in the blocks represent the enrichment rank associated with the specific SNP categories (ie. Syn, NonSyn, 3'UTR, 5'UTR, Intron, and Promoter).

Furthermore, the question whether highly-differentiated regulatory variants in common drug-response genes could serve to co-regulate unrelated drug pathways was also assessed. Common genes that are shared in 32 (78%) of the DRPs are shown to carry extremely population-differentiated 'splicing SNPs', which suggest the potentially high influence of this functional SNP category in drug response (Appendix 5). On the other hand, 17 (41%) of the DRPs carry one or more high-F_{ST} TF binding site variants in their shared common genes. In contrast, only 3 (0.07%) DRPs share highly differentiated SNPs affecting protein deleterious substitution sites. Highly population-differentiated srSNPs in common metabolising enzyme, *CYP3A5* (rs776746), as well as the transporter, *ABCG2* (rs2231164 and rs2725267), could co-regulate 20 different DRPs through their potential effect on RNA splicing (Fig. 3.7B).

3.3.5 Potential translational application and the antiarrhythmic drug as a case

To investigate a possible clinical application of using SNPs architecture information, a pilot DRP priority scoring strategy (see methods) was designed. This score was used to estimate the probable occurrence of drug-response variation in a pathway using the features listed in Table 3.2.

Figure 3.11 showed *Py* scores of all 41 DRPs in this study sorted from highest to lowest. A high *Py* score would signify a higher potential of having a drug-response variation event. In addition, several studies that report on events relating to drug-response variation among different population groups were also mined (Appendix 6). These reports were used to corroborate the *Py* scores. The result demonstrated that as *Py* score increases, the number of pathway without a recognized report on therapy variation decreases (Fig. 3.11). DRP with relatively high *Py* scores would have had reports that indicate on experiencing a response difference, except for one DRP (the VEGF pathway) where literature evidence is yet to be found.

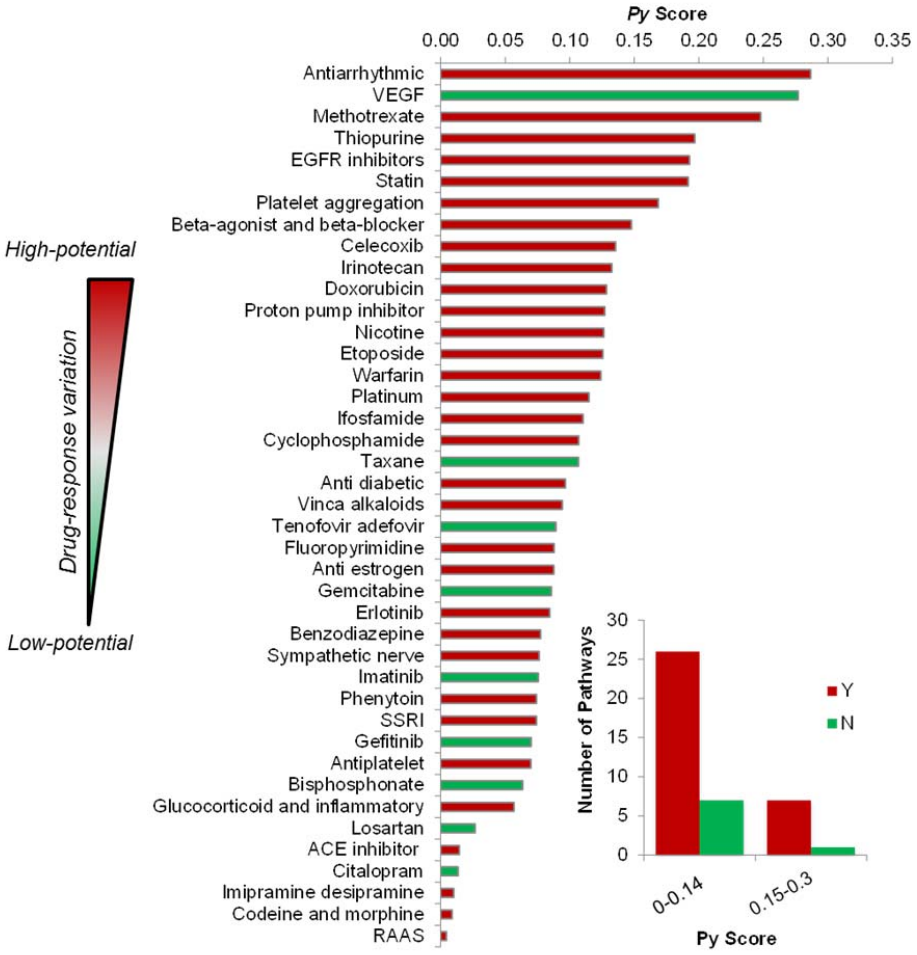


Figure 3.11 The potential implication of human genetic variation to differences in drug response. The graph is sorted from high-to-low potential of having a drug-response difference based on the DRP *Py* scores. (Inset) The number of pathways with literature evidence that corroborate for the presence of drug response variation across different population groups. Red (Y): evidence found or green (N): no evidence found.

A highly potential candidate for clinical application would be in the antiarrhythmic pathway, where *Py score* is the highest (*Py* = 0.29), and where previous studies have reported on population differences in response [23-24]. Table 3.4 provides the list of antiarrhythmic pathway SNPs that might be useful for future pharmacogenetics-based testing. These are SNPs that are relevant for

clinical study not only because of their potential functional implication to altering gene expression and splicing regulatory sites, but also because of the extreme population differentiation signature within them. Four eQTL regulatory SNPs in three genes are also present within the antiarrhythmic pathway (Table 3.5). Interestingly, SNP rs8022091 in *SLC8A3* is the only high-F_{ST} SNP that is associated to eQTL functional SNPs of the entire 715 drug response genes in this study.

Table 3.4 Highly population-differentiated potentially functional SNPs in the Antiarrhythmic pathway.

Gene	SNP	F_{ST}	Populations	Functional Category
ABCC8	rs2077655	0.56	AF-EA/SA/EU	Splicing
	rs2077654	0.51		
	rs12293228	0.34		
ANK2	rs2272229	0.41		
	rs17045935	0.41		
	rs2293324	0.37		
	rs9307389	0.33		
	rs3733615	0.32		
ATP1A1	rs1407716	0.50		TF Site
CACNA1D	rs6766988	0.57		Splicing
KCNJ5	rs10790976	0.31	AF-SA/EU	TF Site
KCNQ1	rs10798	0.33	EA-EU/SA/AF	miRNA Binding Site
	rs8234	0.32		
LMNA	rs505058	0.55	AF-EA/SA/EU	Splicing
	rs547915	0.42		
	rs2485664	0.33	AF-EA/EU	
	rs520973	0.31		
SLC8A2	rs830132	0.36	AF-EA/SA/EU	
	rs830134	0.36		TF Site
SLC8A3	rs8022091	0.31	AF-SA/EU	

Table 3.5 Potentially functional SNPs in the Antiarrhythmic pathway that are associated with differential local gene-expression.

Gene	SNP	Functional Category	r^2	p-value	FDR	Associated transcript	Population
HCN2	rs34830716 rs35926953	TF Site	0.444393	1.06E-06	0.00094	NM_198591 (BSG)	CEU
RYR2	rs2275288	ISRE	0.352726	3.38E-05	0.016806	NM_001035 (RYR2)	YRI
SLC8A3	rs8022091	TF Site	0.248623	5.84E-05	0.008819	NM_182936 (SLC8A3)	CHB & JPT

3.4 Discussion

In this chapter, the global SNP architecture of pathway genes important in drug therapies was deeply analyzed. Using the gene functional region-directed approach, this study reveals the high polymorphic property of the DRPs. Furthermore, it suggests for the prominence of SNPs with potential implication to TF binding, RNA splicing and protein deleterious site. The high presence of regulatory and RNA SNPs was also highlighted, in addition to those of coding SNPs in the DRPs.

Clinical variations in therapy are well attributed to differences in the patients' genetic background [1-2]. Using population genotype data, it has been shown that the drug-response genes are more differentiated than other genes in the human genome [16]. However up to the point when this study was conducted, no report described the relevance of population differences to the global architecture of SNPs functionality in these genes. In this chapter, it is shown that when population differentiation is considered, the magnitude of functional cSNPs in the DRPs is lower. The low frequency diversification of functional cSNPs could suggest that population differences in drug response may less be affected by SNPs acting on a protein structural level. Instead, there is a higher plausibility that these population differences in drug response are more affected by SNPs acting on gene regulation, as shown by the high prevalence of extremely population-differentiated rSNPs residing in TF binding sites as well as rsSNPs residing in

splicing regulatory regions. This result therefore implies the importance of these rSNPs and rsSNPs in drug response.

Gamazon et al has previously shown that chemotherapeutic drug susceptibility-associated SNPs are enriched in expression quantitative trait loci (eQTL) [25]. This is aligned with the results obtained in this chapter, as it was also observed that there are many more DRP genes that house rSNPs and srSNPs associated with changes in gene expression profile. In fact, the result presented in this chapter could further extend the knowledge that was reported by Gamazon et al, by providing a layer of SNPs potential functionality in mind. This implicates the rSNPs and srSNPs and their probable significance in regulating the expression of genes in the DRP. It is valuable for designing future studies because the insight provided here could put more weight on these rSNPs and srSNPs, highlighting their equally important role compared to the cSNPs. In fact, my argument is parallel to the discussion that was presented by Sadee et al [12].

Furthermore, a drug PK/PD process is not an isolated event [19-20], yet many pharmacogenetics studies focused only on certain candidate genes in studying the role of SNPs in inducing drug response variation. This traditional approach would isolate other genes or SNPs within the same pathway, which may be equally crucial as the candidate gene or the candidate SNP itself. Here using the DRP-based approach, it can be shown that the pathways responsible for drug therapies are not only highly polymorphic in nature; but also have high magnitude of expression-associated or population-differentiated rSNPs and srSNPs. Many of these SNPs could have been missed out if one is to use the traditional candidate

gene- or SNP-based approach in pharmacogenetics. Hence in this chapter, it is shown that the PK/PD pathway-based approach in analyzing the genetic basis of drug response is arguably effective, allowing one to gain a deeper understanding of the genetic pattern in these drug-response genes. In another word, this gives ability to have a 'helicopter view' of the polymorphisms residing in these genes before we narrow down to study a particular SNP in detail.

Moreover, with this pathway-based approach, a potential phenomenon in which several unrelated DRPs are possibly being regulated by a selected gene could also be observed. An rSNP in the ABCB1 transporter (rs3747802) and another one in the UGT1A1 metabolizer (rs10929302) are found to be involved in multiple DRPs. Respectively, these SNPs are potentially functional in the TF binding on the promoter region, in addition to being associated with differential gene expression. Furthermore, the rs10929302 or UGT1A1*93 (5' UR -3136G>A) polymorphism has previously been shown to be associated with the different susceptibility of Irinotecan-induced toxicity [26-28]. Based on this result, it is suggested that this same rSNP could also functionally affect pathways associated with other drugs such as etoposide, statin and losartan.

In addition, this phenomenon is also observed in rSNPs that are population-differentiated. One high-F_{ST} rSNP in the *CYP3A5* metaboliser (rs77646) and two high-F_{ST} rSNPs in the *ABCG2* transporter (rs2231164 and rs2725267) have potential functional impact on RNA splicing. The SNP significance is supported by Zeng et al, who in 2011 reported a candidate gene-based study of 211 pancreatic cancer patients (137 of whom have had chemotherapy), where the

rs2231164 (*ABCG2* Intron 14 -46A>G) variant was found to have been associated with survival [29].

Moving forward, it can therefore be expected that when performing functional studies involving a SNP or gene that is commonly involved in multiple drug pathways, one could also expect to also see the SNP effect in several different drug pathways. The reason is because this SNP is involved in multiple DRPs.

There is a potential application from this pathway-based potentially functional SNPs analysis. Hence, the last part of this chapter presented an attempt to prioritize drug pathways based on their potential therapeutic differences. A scoring method was developed, taking account of the DRP SNPs potential functional and population differentiation status. The different weight applied on the variables could assist prioritization using extremities (such as extremely high F_{ST} SNP in a gene) as an underlying factor. As *Py* score increases, there are more pathways that could be corroborated by reports that support on the presence of clinical difference in therapeutic response. In the case of the antiarrhythmic pathway, where *Py* score is the highest, a list of extremely population-differentiated rSNPs and srSNPs was provided, in addition to those that are associated with differential gene expression. Therefore the clinical utility of these SNPs can be explored further.

However, because both the HapMap and Singapore Genome Variation projects employed the tag-SNP approach to genotyping SNPs, limitations exist, particularly whether these 1.4 million SNPs could serve as the best representative

of other SNPs in the drug-response genes, or more, the human genome. In fact, no representative SNP method is ideal in pharmacogenomics except one method: by studying all SNPs in the human genome. This will be the focus of the next chapter.

3.5 References

1. Weinshilboum, R. (2003). *Inheritance and drug response*. N Engl J Med 348, 529-537.
2. Evans, W.E., and Relling, M.V. (1999). *Pharmacogenomics: translating functional genomics into rational therapeutics*. Science 286, 487-491.
3. Eichelbaum, M., Ingelman-Sundberg, M., and Evans, W.E. (2006). *Pharmacogenomics and individualized drug therapy*. Annu Rev Med 57, 119-137.
4. Wang, Z., Wang, B., Tang, K., Lee, E.J., Chong, S.S., and Lee, C.G. (2005). *A functional polymorphism within the MRP1 gene locus identified through its genomic signature of positive selection*. Hum Mol Genet 14, 2075-2087.
5. Letourneau, I.J., Deeley, R.G., and Cole, S.P. (2005). *Functional characterization of non-synonymous single nucleotide polymorphisms in the gene encoding human multidrug resistance protein 1 (MRP1/ABCC1)*. Pharmacogenet Genomics 15, 647-657.
6. Kimchi-Sarfaty, C., Oh, J.M., Kim, I.W., Sauna, Z.E., Calcagno, A.M., Ambudkar, S.V., and Gottesman, M.M. (2007). *A "silent" polymorphism in the MDR1 gene changes substrate specificity.* Science 315, 525-528.
7. Pang, G.S., Wang, J., Wang, Z., and Lee, C.G. (2009). *Predicting potentially functional SNPs in drug-response genes*. Pharmacogenomics 10, 639-653.
8. Wolf, S.J., Bachtiar, M., Wang, J., Sim, T.S., Chong, S.S., and Lee, C.G. (2011). *An update on ABCB1 pharmacogenetics: insights from a 3D model into the location and evolutionary conservation of residues corresponding to SNPs associated with drug pharmacokinetics*. Pharmacogenomics J.
9. Moitra, K., Scally, M., McGee, K., Lancaster, G., Gold, B., and Dean, M. (2011). *Molecular evolutionary analysis of ABCB5: the ancestral gene is a full transporter with potentially deleterious single nucleotide polymorphisms*. PLoS One 6, e16318.

10. Arsenault, J., Lehoux, J., Lanthier, L., Cabana, J., Guillemette, G., Lavigne, P., Leduc, R., and Escher, E. (2010). *A single-nucleotide polymorphism of alanine to threonine at position 163 of the human angiotensin II type 1 receptor impairs Losartan affinity*. Pharmacogenet Genomics 20, 377-388.
11. Boni, V., Zarate, R., Villa, J.C., Bandres, E., Gomez, M.A., Maiello, E., Garcia-Foncillas, J., and Aranda, E. (2010). *Role of primary miRNA polymorphic variants in metastatic colon cancer patients treated with 5-fluorouracil and irinotecan*. Pharmacogenomics J.
12. Sadee, W., Wang, D., Papp, A.C., Pinsonneault, J.K., Smith, R.M., Moyer, R.A., and Johnson, A.D. (2011). *Pharmacogenomics of the RNA world: structural RNA polymorphisms in drug therapy*. Clin Pharmacol Ther 89, 355-365.
13. Pottier, N., Paugh, S.W., Ding, C., Pei, D., Yang, W., Das, S., Cook, E.H., Pui, C.H., Relling, M.V., Cheok, M.H., et al. (2010). *Promoter polymorphisms in the beta-2 adrenergic receptor are associated with drug-induced gene expression changes and response in acute lymphoblastic leukemia*. Clin Pharmacol Ther 88, 854-861.
14. Wang, J., Ronaghi, M., Chong, S.S., and Lee, C.G. (2011). *pfSNP: An integrated potentially functional SNP resource that facilitates hypotheses generation through knowledge syntheses*. Hum Mutat 32, 19-24.
15. O'Donnell, P.H., and Dolan, M.E. (2009). *Cancer pharmacoethnicity: ethnic differences in susceptibility to the effects of chemotherapy*. Clin Cancer Res 15, 4806-4814.
16. Li, J., Zhang, L., Zhou, H., Stoneking, M., and Tang, K. (2011). *Global patterns of genetic diversity and signals of natural selection for human ADME genes*. Hum Mol Genet 20, 528-540.
17. (2003). *The International HapMap Project*. Nature 426, 789-796.
18. Teo, Y.Y., Sim, X., Ong, R.T., Tan, A.K., Chen, J., Tantoso, E., Small, K.S., Ku, C.S., Lee, E.J., Seielstad, M., et al. (2009). *Singapore Genome Variation Project: a haplotype map of three Southeast Asian populations*. Genome Res 19, 2154-2162.

19. Owen, R.P., Altman, R.B., and Klein, T.E. (2008). *PharmGKB and the International Warfarin Pharmacogenetics Consortium: the changing role for pharmacogenomic databases and single-drug pharmacogenetics*. Hum Mutat 29, 456-460.

20. Hewett, M., Oliver, D.E., Rubin, D.L., Easton, K.L., Stuart, J.M., Altman, R.B., and Klein, T.E. (2002). *PharmGKB: the Pharmacogenetics Knowledge Base*. Nucleic Acids Res 30, 163-165.

21. Klein, T.E., Chang, J.T., Cho, M.K., Easton, K.L., Fergerson, R., Hewett, M., Lin, Z., Liu, Y., Liu, S., Oliver, D.E., et al. (2001). *Integrating genotype and phenotype information: an overview of the PharmGKB project. Pharmacogenetics Research Network and Knowledge Base*. Pharmacogenomics J 1, 167-170.

22. Weir, B.S., and Cockerham, C.C. (1984). *Estimating F-Statistics for the Analysis of Population Structure*.

23. Potkin, S.G., Shen, Y., Pardes, H., Phelps, B.H., Zhou, D., Shu, L., Korpi, E., and Wyatt, R.J. (1984). *Haloperidol concentrations elevated in Chinese patients*. Psychiatry Res 12, 167-172.

24. Lin, K.M., Poland, R.E., Lau, J.K., and Rubin, R.T. (1988). *Haloperidol and prolactin concentrations in Asians and Caucasians*. J Clin Psychopharmacol 8, 195-201.

25. Gamazon, E.R., Huang, R.S., Cox, N.J., and Dolan, M.E. (2010). *Chemotherapeutic drug susceptibility associated SNPs are enriched in expression quantitative trait loci*. Proc Natl Acad Sci U S A 107, 9287-9292.

26. Innocenti, F., Undevia, S.D., Iyer, L., Chen, P.X., Das, S., Kocherginsky, M., Karrison, T., Janisch, L., Ramirez, J., Rudin, C.M., et al. (2004). *Genetic variants in the UDP-glucuronosyltransferase 1A1 gene predict the risk of severe neutropenia of irinotecan*. J Clin Oncol 22, 1382-1388.

27. Cote, J.F., Kirzin, S., Kramar, A., Mosnier, J.F., Diebold, M.D., Soubeyran, I., Thirouard, A.S., Selves, J., Laurent-Puig, P., and Ychou, M. (2007). *UGT1A1 polymorphism can predict hematologic toxicity in patients treated with irinotecan*. Clin Cancer Res 13, 3269-3275.

28. Cecchin, E., Innocenti, F., D'Andrea, M., Corona, G., De Mattia, E., Biason, P., Buonadonna, A., and Toffoli, G. (2009). *Predictive role of the UGT1A1, UGT1A7, and UGT1A9 genetic variants and their haplotypes on the outcome of metastatic colorectal cancer patients treated with fluorouracil, leucovorin, and irinotecan*. J Clin Oncol 27, 2457-2465.

29. Zeng, H., Yu, H., Lu, L., Jain, D., Kidd, M.S., Saif, M.W., Chanock, S.J., Hartge, P., and Risch, H.A. (2011). *Genetic effects and modifiers of radiotherapy and chemotherapy on survival in pancreatic cancer*. Pancreas 40, 657-663.

30. Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., and Sirotkin, K. (2001). *dbSNP: the NCBI database of genetic variation*. Nucleic Acids Res 29, 308-311.

31. Stranger, B.E., Forrest, M.S., Dunning, M., Ingle, C.E., Beazley, C., Thorne, N., Redon, R., Bird, C.P., de Grassi, A., Lee, C., et al. (2007). *Relative impact of nucleotide and copy number variation on gene expression phenotypes.* Science 315, 848-853.

32. (2010). *A map of human genome variation from population-scale sequencing*. Nature 467, 1061-1073.

33. Holsinger, K.E., and Weir, B.S. (2009). *Genetics in geographically structured populations: defining, estimating and interpreting F(ST)*. Nat Rev Genet 10, 639-650.

Chapter 4. Population Differentiation Pattern in Individuals of the 1000 Genomes Project

4.1 Introduction

The previous chapter highlighted the result of studying population differentiation pattern of SNPs in conventional drug-response genes. However, there is a limitation that these HapMap tag-SNPs may not be the best representatives of all SNPs in our genome. Hence in this chapter, with available genome-wide data from the 1000 Genomes Project, which at that time had just released its Phase 1 data, I expanded the analysis to a genome scale. This data would cover all SNPs that could be identified from genome sequencing results of thousands of individuals. In addition, the data would eventually cover novel genes that may be important in drug response, in addition to the 'conventional' drug-response genes and SNPs that were covered in the previous chapter.

In this chapter, the main focus is to articulate the novel pattern of genetic differentiation in addition to elucidating their potential functional relevance in the genomes of individuals that participated in the 1000 Genomes project. During the initiation of this thesis, few had deeply investigated this potentially intriguing pattern of population differentiation in our genome as prior to the advancement of DNA sequencing technology, such data was sparse. The effort that is focused here

is based on a novel approach, which to the best of my knowledge, had not been adopted in previous studies. Here, using the 1000 Genomes project (Phase 1) data, a genome-wide scan of pattern of genetic differentiation is conducted in 14 global populations that originated from four different continents: Latin America, Europe, Africa, and East Asia.

I present the identification of 'top chromosome differentiated SNPs' (tcdSNPs) and the genes that house these SNPs, the 'tcdGenes'. Many of these SNPs are also predicted to have a functional significance. The identification of tcdSNPs had also allowed a significant expansion of work that was done in chapter 3 of this thesis, whereby using this data, we can now better elucidate the potential routes in which population genetic differentiation can contribute to population differences in drug response.

Furthermore, it is intriguing to investigate the biological pathways that are affected by population-differentiated genes. Hence towards the last part of this chapter, tcdGenes enrichment in biological pathways was investigated. Because human phenotype differences are determined by the variation of pathways activity, this approach would be extremely relevant in determining the connection between differentiation in population genetic architecture with that of phenotype variation seen across different populations. By identifying pathways that are enriched by tcdGenes, we could point to a possible population differentiation determinant that would eventually play role in phenotype differences.

4.2 Methods

4.2.1 Estimating Population Differentiation from 1000 Genomes Data

In this chapter, genome-wide population differentiation was estimated from publicly available data that was released by the 1000 Genomes project (Phase 1) [1]. This encompasses allele frequency data from 1,092 individuals who are originated from 14 different populations, which was derived by the use of the VCF tool (version 0.1.9). Based on their geographic origins and genetic relatedness, these diverse populations could broadly be clustered based on four different continental origins: Africa, East Asia, Latin America, and Europe.

The African cluster encompasses the ASW (African ancestries from Southwest United States), LWK (Luhya individuals in Kenya) and YRI (Yoruba individuals in Nigeria). The East Asian populations are CHB (Han Chinese in Beijing, China), CHS (Han individuals in Southern China) and JPT (Japanese individuals in Tokyo, Japan). Three populations belong to the Latin American group including the CLM (Columbian in Medellin, Columbia), MXL (Mexicans in Los Angeles, United States) and PUR (Puerto Rican in Puerto Rico). The IBS (Iberian in Spain) lies somewhere between the Latin American and European cluster. Lastly, the European groups consisted of individuals from the CEU (Northern and Western European ancestries in Utah, United States), FIN (Finish in Finland), GBR (British in England and Scotland, Great Britain), TSI (Toscani in Italy).

The SNP allele frequency data was used for the estimation of population differentiation using F_{ST} statistics [2, 3]. The F_{ST} score calculation method

employed in this chapter is identical to the one utilized in the previous chapter, albeit performed in a significantly bigger data size. Here, F_{ST} scores were calculated based on population pair comparisons. This was performed using an 'R' script that was originally constructed for the course of this thesis. The 'R' script was deployed in the LCFG high performance work stations, in addition to the NUS high performance computing cluster across multiple processing nodes.

4.2.2 Identifying maximum-differentiated SNP clusters in the human genome

The greater SNPs coverage in the 1000 Genomes Project Data has opened the possibility of performing a genome-wide analysis of SNPs population-differentiation, in investigating the population differentiation pattern of the human genome. By identifying SNP clusters that contain the 'maximum-differentiated' SNPs, we could identify chromosome regions that could have been subjected to population differentiation or selection process. Here, I have developed a novel approach in identifying tcdSNPs that are located within the SNP clusters that are assosciated with the maximum F_{ST} scores in a given population pair comparison across 23 human chromosomes. An algorithm that is based on a 'roll mean' method was developed to compute the moving average spanning a window size of 15,000 SNPs, which corresponds to the size of the cluster (Fig. 4.1). A cluster carrying SNP associated with the maximum average F_{ST} score could be identified by reading maximum peak region. The algorithm then mined the top 5% of SNPs in this cluster, which carry the highest F_{ST} scores in the respective cluster. These SNPs are then classified as the 'top chromosome differentiated SNPs' (tcdSNPs).

The distribution of F_{ST} scores across the SNP gene functional region that is presented in this chapter was analysed based on two population pairs from different extremes: most similar (CEU-GBR) and least similar (CHS-YRI) populations.

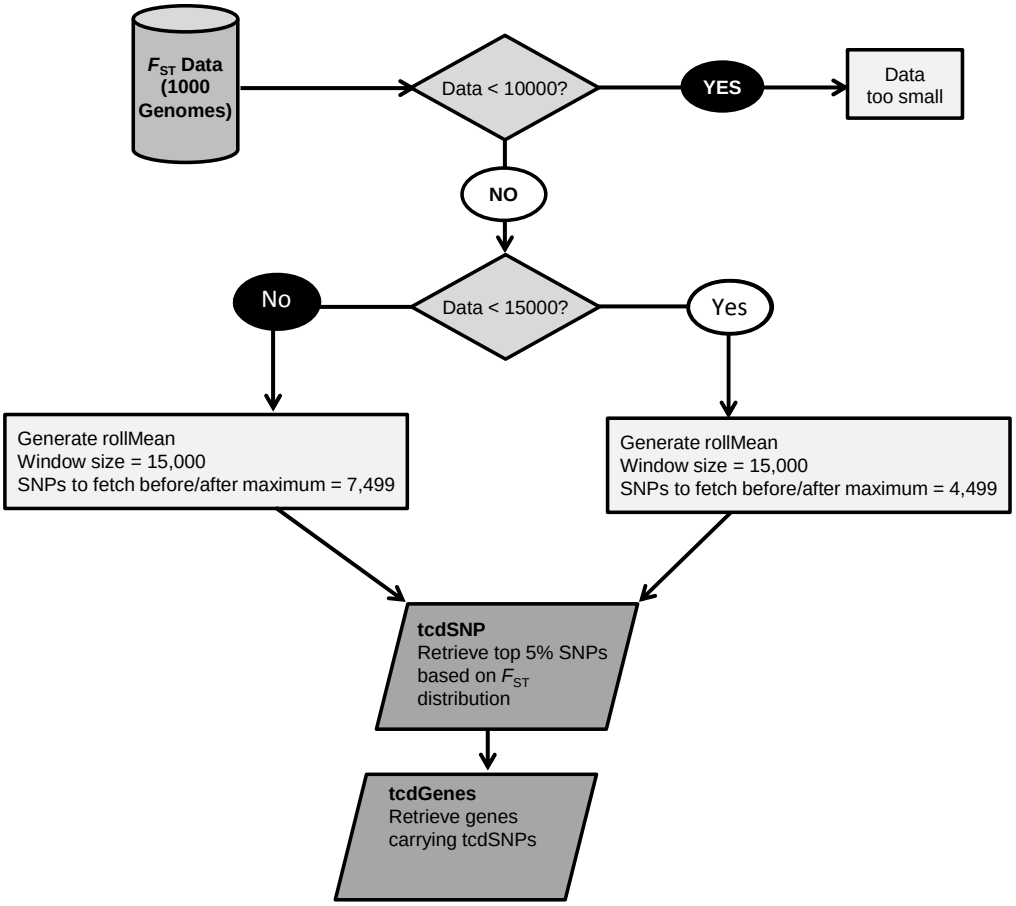


Figure 4.1 Algorithm for finding maximum-differentiated SNP clusters

4.2.3 Genome-wide SNPs mapping to functional gene regions

SNPs that reside within functional gene region are labelled ‘genic’ SNPs whilst those outside of genes are labelled ‘intergenic'. SNPs classification was performed using the functional region-directed approach as explained in the previous chapter. However, unlike what was performed in the previous chapter, mapping the 1000 Genomes SNPs to the functional gene regions is more challenging, which is attributed mostly to the big data size. In this chapter, NCBI dbSNP build 137 was

used as reference, a significant upgrade of SNPs collection to that of build 131. The 'top chromosome differentiated genes' (tcdGenes) are defined as genes that house one or more tcdSNPs.

4.2.4 tcdGenes enrichment in biological pathways

Population-differentiation in genes in biological pathways could potentially result in the variation of phenotypes seen in different population, including drug response variation. To investigate this potential impact of population genetic differences, several enrichment analyses using a diverse collection of biological pathways were performed. In this chapter, the curated pathways information was obtained from the Ingenuity Pathway Analysis (IPA), KEGG [4] and Gene Ontology (GO) [5] databases. With the exception of the IPA pathway analysis, all enrichment analyses were done in the 'R' environment by utilising the clusterProfiler package (version 1.10.0) [6]. The enrichKEGG function was selected for analysing population-differentiated genes enrichment in the KEGG pathways, whilst the enrichGO function was used for the analysis involving the GO database. For each pathway in the database, all functions will compute the enrichment of population-differentiated genes. The 'enrich' functions are modelled based on the hypergeometric distribution. Enrichment is defined when a pathway achieves *P*-value less than 0.00005, a very stringent cut-off that take account the Bonferroni multiple tests error correction. The QIAGEN's Ingenuity® Pathway Analysis (IPA®, QIAGEN Redwood City, www.qiagen.com/ingenuity) was used to conduct the IPA analysis.

4.2.5 Population differentiation in pharmacogenomics pathways

In this thesis, my work encompasses the integration of drug pathways or gene sets from multiple resources, which will be mainly utilized in the next chapter. Briefly for this chapter, the drug-response genes collection was significantly expanded as compared to the previous chapter. In addition to annotated genes information from the PharmGKB database [7], drug-response genes collection were also obtained from the DrugBank [8], ChEMBL (version 13) [9], and the Comparative Toxicogenomics Databases (CTD) [10]. Genes in these pharmacogenomics pathways were previously curated based on their association with a drug and/or other compounds.

Furthermore, a subset of the tcdGenes are linked to these pharmacogenomic pathways. These tcdGenes were then used in pathway enrichment analysis. Each pathway is associated with drugs or compounds that were approved by the US food and Drug Authority (FDA) and the proportion of such genes in all the pharmacogenomics pathways was calculated.

For each pathway, tcdGenes enrichment analysis was performed using random sampling statistical exercise. As performed in the previous chapter, this random sampling simulation involved a 10,000 time iterations of random samples from the genome, with the same number of genes as found in the pathway. These genes would also have to fall within the same size range as the pathway or drug gene set. Subsequently, the multiple sampling iterations allowed the formation of an approximately normal empirical distribution that is unique for each drug gene sets.

Using this distribution, a 'Z-score' was derived according to the following equation.

$$\text{Z score} = \frac{\text{p-eP}}{\text{eSD}}$$

Where:

p = proportion of tcdGenes in drug-response gene set

eP = average proportion of tcdGenes in empirical distribution of the respective drug

eSD = standard deviation of empirical distribution of the respective drug

A high Z-score, particularly one that is higher than 1.96, signifies a non-random observation of high proportion of tcdGenes in the drug-response gene set, whereas a small Z-score defines a non-significant presence of tcdGenes.

4.3 Results

4.3.1 Population differentiation in different world regions

In this chapter, the SNP data that were utilized mostly originated from the 1000 Genomes Project. A total of 31,953,064 SNPs are annotated in the NCBI dbSNP database (build 137). F_{ST} calculations were successful for 22,866,661 SNPs, accounting for 71.56 percent of the total number of SNPs mined from the 1000 Genomes Project. In 28.44 percent of the SNPs, the F_{ST} algorithm encountered computation error which could possibly be attributed to the presence of mono allelic SNPs or variant call error. In comparison to the HapMap SNPs that were utilised in chapter three, which publicised up to 1.4 million SNPs, the 1000 Genomes Project data is 16 times more vast and covered significantly more part of the human chromosomes.

To estimate pairwise differentiation across 14 populations, which equates to 91 population pairwise combinations, the Weir F_{ST} statistics were calculated using SNPs allele frequency. The average F_{ST} between populations that are originated from a more similar geographic region tend to be lower than those that are separated by a further geographic distance (Fig. 4.2). For instance, in the European population groups, average F_{ST} lies between 0.001 and 0.0062. Moreover, not only that the CEU-GBR pair is the least differentiated among the European population, it is also the most similar population pair of all the other populations in the 1000 Genomes Project. On the other extreme, the highest differentiation is observed between the East Asian and African populations. The

CHS-YRI pair has an average F_{ST} of 0.037 and is the most differentiated population pair (Fig. 4.2A). Moreover, the results presented in this thesis suggest the closer genetic distance between the European populations and the Latino populations, but not the East Asian and African population (Fig. 4.2A and 4.2B).

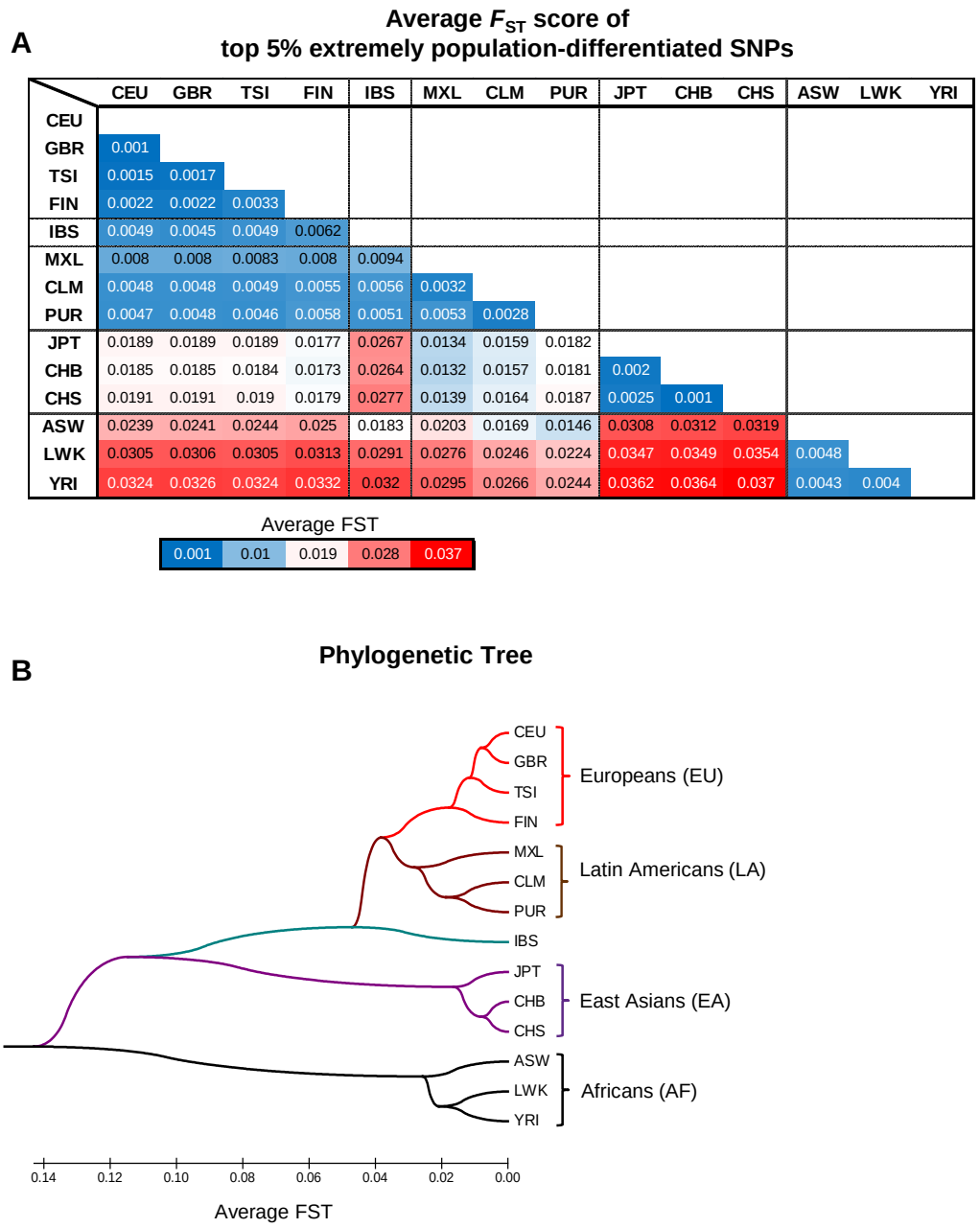


Figure 4.2 Population differentiation pattern in 14 global populations that participated in the 1000 Genomes Project. (A) The average pairwise F_{ST} scores computed from SNPs in the top 5% of the population pair F_{ST} distribution. (B) Population differentiation as seen using a phylogenetic tree that was constructed using the average pairwise F_{ST} scores of the SNPs within the top 5% of the population pair F_{ST} distribution.

Furthermore, within similar geographic region, the African populations (ASW, LWK and YRI) are more diversified, suggesting the stronger presence of 'local differentiation' that leads to more diversification in Africans. On the other hand, there is less local differentiation within the East Asians, Latinos and Europeans populations (Fig. 4.2A and 4.2B). This shows high diversity within the African populations, in addition to the further genetic separation of this populations group compared to the Europeans, Latinos and East Asians. Furthermore, the East Asian group of populations, as shown by the shorter branch length, is less diversified than the other population groups (Fig. 4.2B). This result suggests the presence of lesser genetic diversity in the East Asian populations as compared to that of the Africans. In this study, the IBS population is considered as a 'standalone' population because it is not grouped to any one of the continental population cluster due to its relatively low similarity with the other populations.

4.3.2 F_{ST} scores distribution in human genes

The underlying hypothesis in this chapter is that outside of gene region, there is less potential functional impact that DNA sequence variation could potentially exert, hence intergenic SNPs are more susceptible to population differentiation. To address this hypothesis, this chapter compared the average population differentiation scores between genic and non-genic SNPs, in two most extreme population-pairs. The first one is in CHS-YRI, the two most genetically different

population (average F_{ST} = 0.037), whilst the other pair is the two most genetically similar populations CEU-GBR (average F_{ST} = 0.001) (Fig. 4.2A).

In CHS-YRI, the average F_{ST} score of non-genic SNPs is significantly higher than those in the genic region (*P*-value < 2.2 x E-16) (Fig. 4.3A). Similarly, in the two most similar populations (CEU-GBR), non-genic SNPs also have a significantly higher average F_{ST} compared to genic SNPs (*P*-value < 2.2 x E-16).

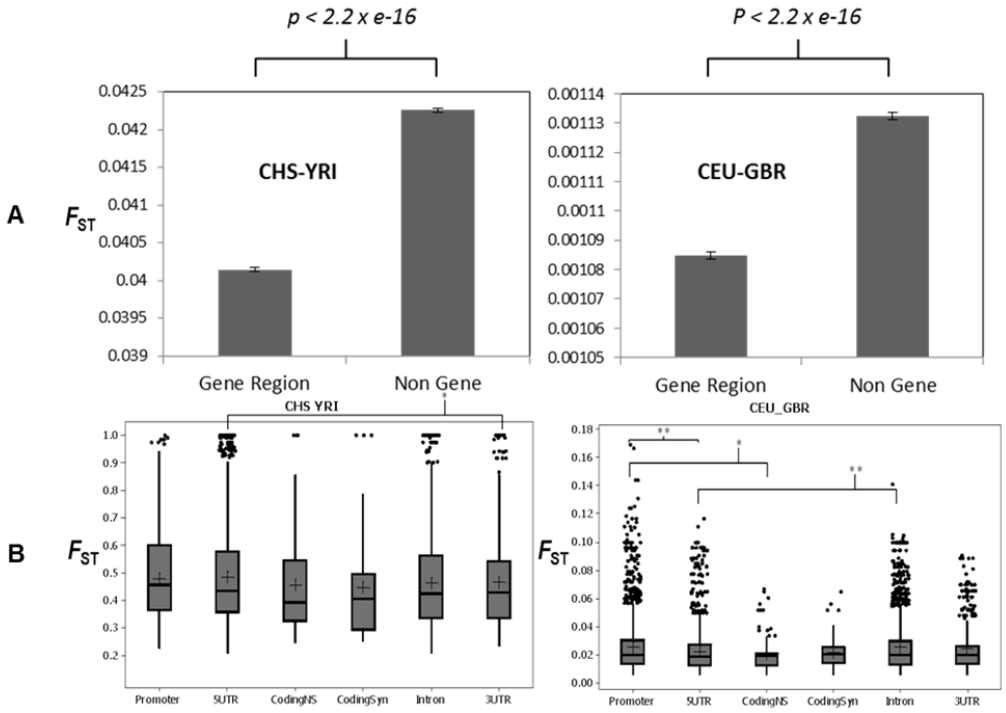


Figure 4.3 Population differentiation across gene functional regions. (A) Significant difference in average F_{ST} scores between the genic and intergenic regions in both CHS-YRI and CEU-GBR population pairs (*P*-value < 2.2 x E-16). The F_{ST} scores distribution across different gene functional regions in these two population pairs are shown in (B).

In gene regions, the SNPs are classified based on their location such as promoter, 5' UTR, 3' UTR, intron as well as the coding non-synonymous and coding synonymous SNPs. In the CHS-YRI pair, the overall distribution of F_{ST} scores across these different categories seems to be no different (Fig. 4.3B). A *t*-test performed between the average F_{ST} scores of the SNPs in the 5' UTR and 3'UTR yielded a *P*-value of 0.039 in the CHS-YRI population pair. In the CEU-GBR pair, where overall F_{ST} scores are much lower than the CHS-YRI pair, significant differences in the F_{ST} scores distribution is observed between SNPs in the

promoter versus 5' UTR, promoter versus coding non-synonymous, in addition to the 5' UTR versus intron comparison.

4.3.3 Genomic signature of 'maximum population differentiation'

Since the 1000 Genomes SNPs data is vastly distributed across the spectrum of the human genome, we could perform a genome-wide scan to find the SNPs that are associated with 'maximum population differentiation'. These SNPs were identified using an algorithm that was constructed based on a roll-mean function. It first computed the moving average F_{ST} scores using 15,000 SNPs sliding window, which provided a significant noise reduction as compared to smaller window size. Subsequently, using these moving average values, group of SNPs with the highest moving average were identified. These groups of SNPs are referred as 'chromosome maximum-differentiated SNPs cluster'.

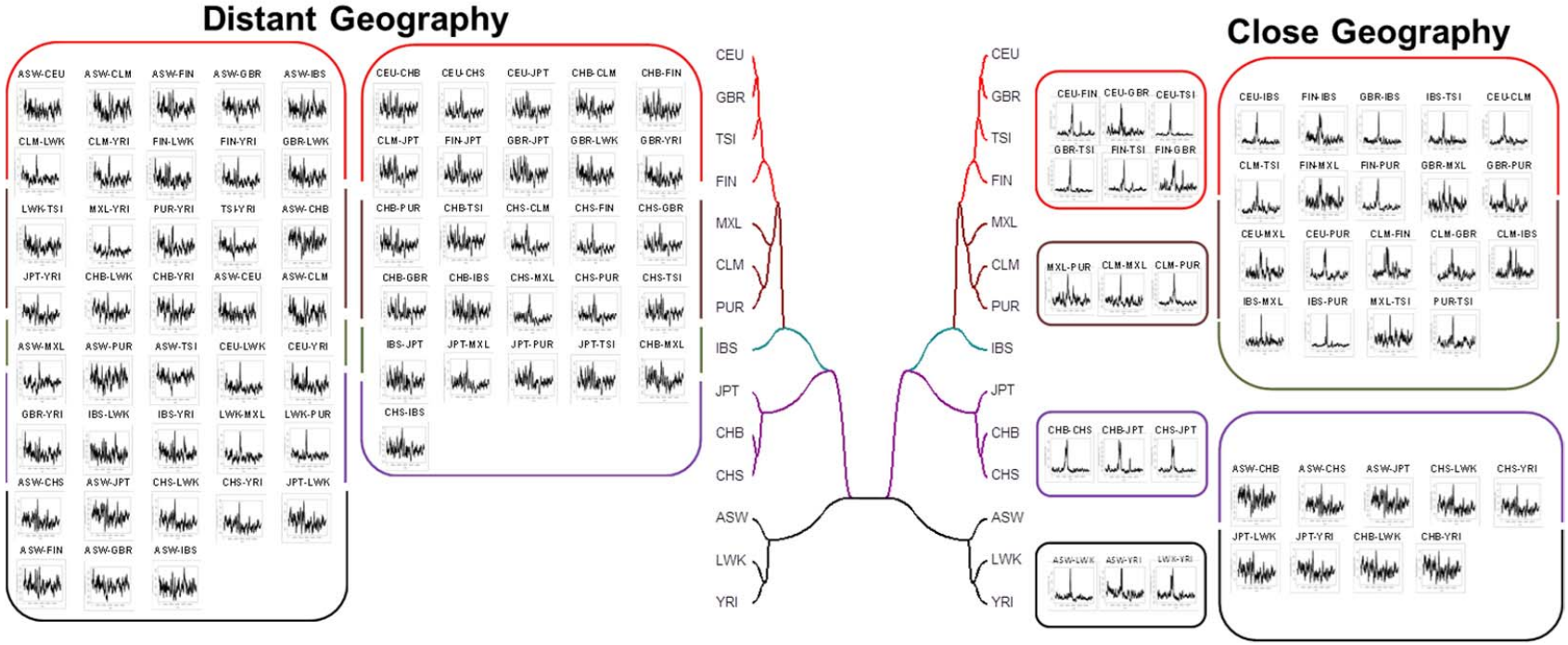


Figure 4.4 The roll mean or moving average analysis of population-pair F_{ST} scores in chromosome 6. Each graph represents the population-pair differentiation pattern of SNPs in chromosome 6. The graphical summary was arranged in accordance to the F_{ST} tree branching (centre), which represents the average degree of differentiation between one population and another population. Here, the population pairwise comparisons were clustered into subgroups based on their general differentiation pattern in the whole genome. The y-axis is the moving average value that was calculated from a sliding window of 15,000 SNPs, whilst the x-axis is the index of SNPs from the 5' to the 3' end of the chromosome.

Within these maximum-differentiated SNPs cluster, the algorithm then fetched SNPs in the top 5% of the F_{ST} distribution of the respective cluster. These SNPs are referred as the 'top chromosome differentiated SNPs' (tcdSNPs). Figure 4.4 displays the results of this analysis in chromosome 6, and figure 4.5 presents the result across eight representative population pairs. Four pairs represent the populations from close (CEU-GBR, CHB-CHS, ASW-LWK, and MXL-CLM) and distant (CHS-YRI, IBS-YRI, GBR-YRI, and YRI-JPT) geographical origins, respectively. In the former group, there is a consistent occurrence of maximum differentiation in the SNPs cluster that is located around the 30 millionth base region, which is approximately at locus 6p21 in chromosome 6 (shown as region between 400,000 and 600,000 SNP indices in Fig. 4.4 and 4.5).

However, in the group of population pairs that are originated from geographically distant location (with average F_{ST} of more than 0.03), this pattern is not observed. This is due to the presence of other high peak regions at these distant population pairs. Hence, this 'signal' is stronger in genetically more similar individuals but is masked by the presence of other genetic differentiation in more distant population groups.

The result may suggest the potential significance of population genetic differentiation in genes located at the 6p21 locus. Morever, because the population differentiation pattern is only seen across populations that are genetically similar, but not in genetically more different individuals, the result could suggests for the presence of population-differentiated genes or chromosome region of functional significance that could be exerted in individual level.

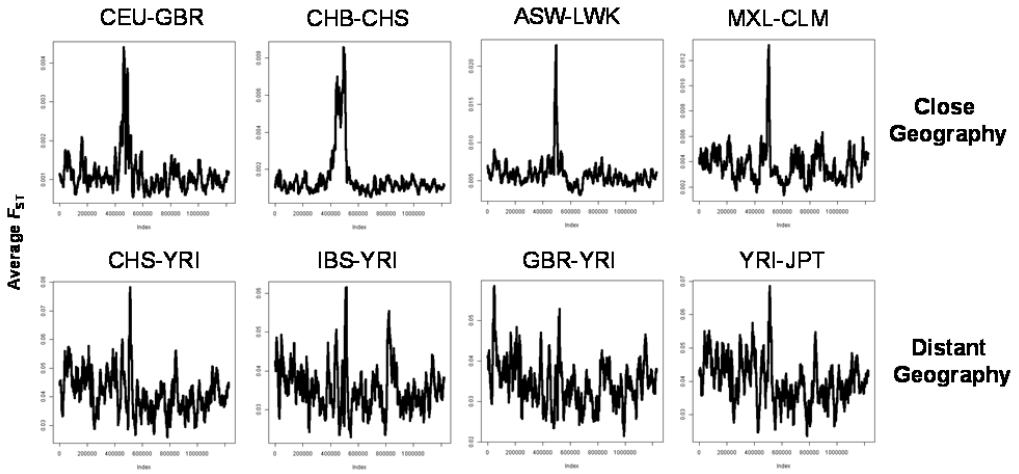


Figure 4.5 The roll mean or moving average analysis for F_{ST} scores in chromosome 6 across representative population pairs. The y-axis is the moving average value that was calculated from a sliding window of 15,000 SNPs, whilst the x-axis is the index of SNPs from the 5' to the 3' end of the chromosome.

4.3.4 Genes in the maximum-differentiated SNPs clusters

The maximum-differentiated SNPs clusters that are unique to the different chromosomes and 91 population-pair combination consisted of 15,000 SNPs. Since these genic SNPs are mapped to NCBI genes, the genes that can be associated with SNPs in these maximum differentiation clusters could be extracted. Through this work, a total of 579,291 tcdSNPs are found to reside within the maximum-differentiated SNPs clusters. On average, a human chromosome contains 25,186.57 tcdSNPs that are located within maximum-differentiated SNPs clusters. Chromosome X has the smallest number of tcdSNP (17,776 SNPs) whilst Chromosome 16 has the highest number of such SNPs

(30,057 SNPs) (Fig. 4.6A). These SNPs have the potential to affect phenotypic variation that we can observe in different human populations.

Out of 579,291 tcdSNPs, 236,679 (41%) can be mapped to a total of 4,355 NCBI-annotated genes, which are referred to as 'top chromosome differentiated genes' (tcdGenes). For chromosome 6, there are 181 annotated genes which carry tcdSNPs, 35 of them are differentiated between the CEU and GBR in chromosome 6 (Appendix 7). For this case, it is observed that there is a high number of tcdSNPs in HLA-C, a chromosome 6 gene that encode the MHC class I heavy chain receptor type of protein. Depending on its mRNA splice form, the HLA-C is potentially affected by ~60 tcdSNPs for the CEU-GBR population pair. And these SNPs are scattered on the 5' upstream and 3' downstream regions, in addition to that of the Exon, Intron and 3' UTR regions. Another interesting observation in chromosome 6 is in the case of the TRIM genes family. Out of 8 TRIM members in this chromosome, 6 genes carry tcdSNPs. These are TRIM 10, TRIM15, TRIM26, TRIM31, TRIM39, and TRIM 40, which could be potentially affected by these high F_{ST} SNPs.

Chromosome 6 however, contains relatively lower proportion of tcdGenes. As shown in Figure 4.6B, it is the smaller chromosomes that have the highest density of such genes. For instance in chromosome 22, there are 257 out of 629 genes (41%) that are classified as tcdGenes (Fig. 4.6B). The larger chromosomes, such as chromosomes 1, 2 and 3, are less dense in tcdGenes with 261, 224 and 165 tcdGenes, respectively (Fig. 4.6B).

To elucidate the genomic population differentiation pattern, the work in this chapter then focused on accessing its significance in three important gene functional regions: the regulatory Promoter and 3' UTR, as well as the protein coding region. Subsequently, in investigating the potential functional significance of such SNPs, the next step was to filter the tcdSNPs that are predicted to be functionally important including in sites that are crucial for transcription factor-binding, miRNA-binding and protein functional domain. By utilizing the potentially functional SNPs (pfSNP) database, it is identified that 48% or 5,394 out of 11,245 tcdSNPs could potentially be functionally important in the coding region (Fig. 4.6C). In the promoter region, there is a higher absolute count of tcdSNPs that could be functionally important such as in affecting transcription factor binding sites, numbering 22,999 SNPs out of a total number of 78,104 promoter tcdSNPs (29%). Moreover in the 3'UTR region, 28% or 4,745 out of 16,956 of tcdSNPs are predicted to be functionally important.

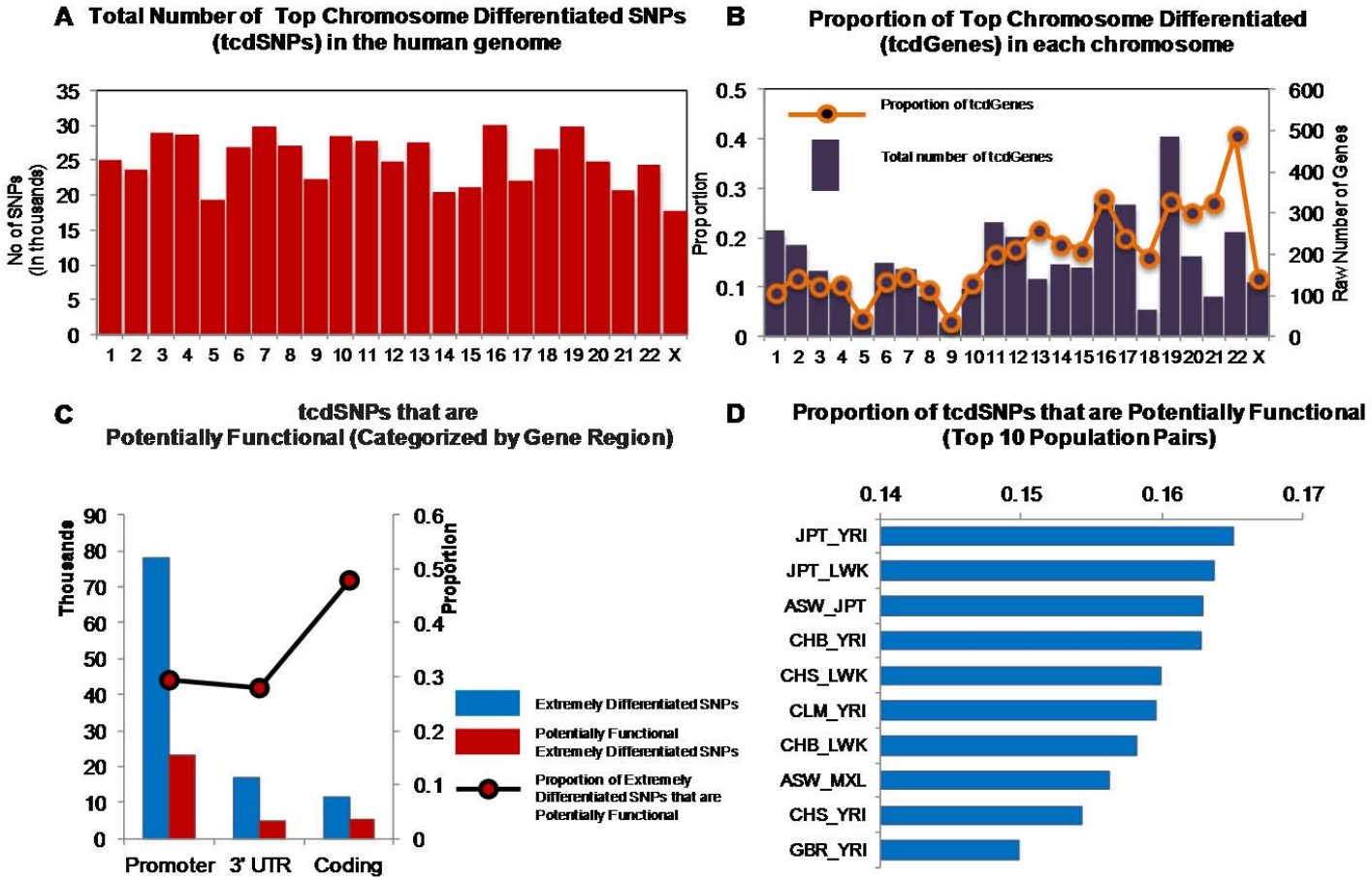


Figure 4.6 SNPs in the maximum-differentiated SNPs clusters. (A) Total number of tcdSNPs residing in the maximum differentiated SNPs clusters in different chromosomes. (B) The proportion of tcdGenes carrying tcdSNPs is displayed by the line chart (primary y-axis). The bar-chart displays the total number of tcdGenes (secondary y-axis). (C) tcdSNPs that are potentially functional in the promoter, 3' UTR and coding gene regions. (D) The proportion of tcdSNPs that is potentially functional across the top 10 population pairs. The top three population pairs involved the Japanese and African populations.

Furthermore, the subject was further examined in different population pairs. With 16.5% of tcdSNPs that are predicted to be functional, the JPT-YRI is associated with the highest proportion of tcdSNPs that are potentially functional (Fig. 4.6D). Furthermore as shown in Figure 4.6D, the JPT-LWK and ASW-JPT, which are both Japanese-African pairs, respectively have 16.4 and 16.2% of the tcdSNPs that are predicted to be functional. The subsequent population pairs with relatively high percentage of potentially functional tcdSNPs are the CHB-YRI (16.3%) and CHS-LWK (16%), which belong to the Chinese-African pair combination (Fig. 4.6D).

In addition, based on the result of this analysis, in East Asians, there is a higher proportion of tcdSNPs that are predicted to affect the promoter region as compared to the other populations (Fig. 4.7). The top seven population pairs with relatively higher proportion of tcdSNPs that are potentially functional in the promoter are the combination between East Asians (CHB, CHS and JPT) versus any of the European (FIN and GBR), Latin American (PUR and CLM) or African (ASW) populations (Fig. 4.7). Furthermore in the 3' UTR region, with the CLM-YRI on top, the proportion of potentially functional tcdSNPs are more uniform in the top 20 population pairs. For the coding SNPs, the top two populations that carry a relatively higher percentage of potentially functional tcdSNPs are the combination of LWK and FIN with Mexican.

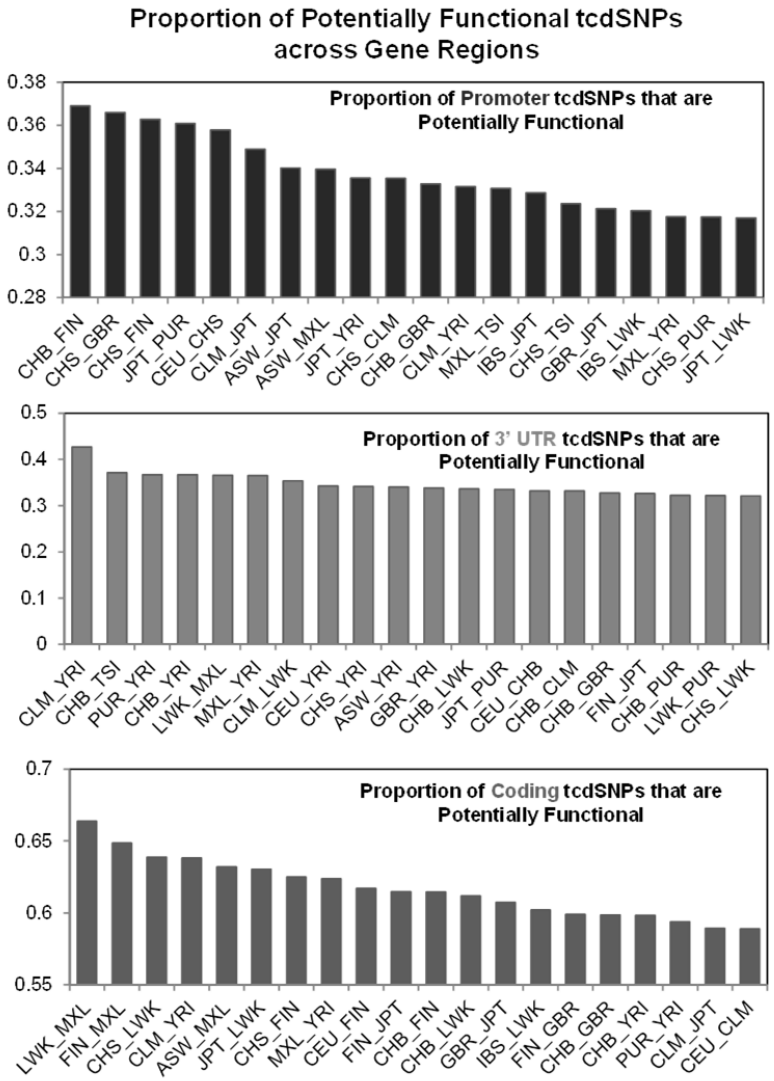


Figure 4.7 The top 20 population pairs based on the proportion of tcdSNPs that are potentially functional in the promoter, 3' UTR and coding regions.

4.3.5 Does size matter?

So far in this chapter, it is observed that there is a high proportion of tcdGenes in chromosomes that is releatively smaller in size. This prompted the question whether there is a correlation between chromosome size and the prevalence of population differentiation in the different chromosomes.

To address this question, an estimate measurement of the prevalence of population differentiation was established. This was done using the 'Differentiated Population-Pairs Ratio' (DPPR). DPPR was obtained by counting the total number of population-pair differentiation that is observed in a chromosome. The raw count was then normalised by the maximum number of differentiated population pairs that could have been possibly observed in that chromosome. Hence, chromosome with large DPPR is postulated to be more susceptible to population differentiation.

As observed in Figure 4.8A, there is a tendency that the smaller chromsomes have relatively high DPPR score. In contrast, it is observed that the larger chromsomes, such as chromosomes 1, 2, 3, 4, and 5, have releatively smaller DPPR scores. Furthermore, in comparison to the other chromosomes, the smallest chromosome 22 has the highest DPPR. This signifies higher prevalence of population-pairs that are affected by population differentiation in this chromosome.

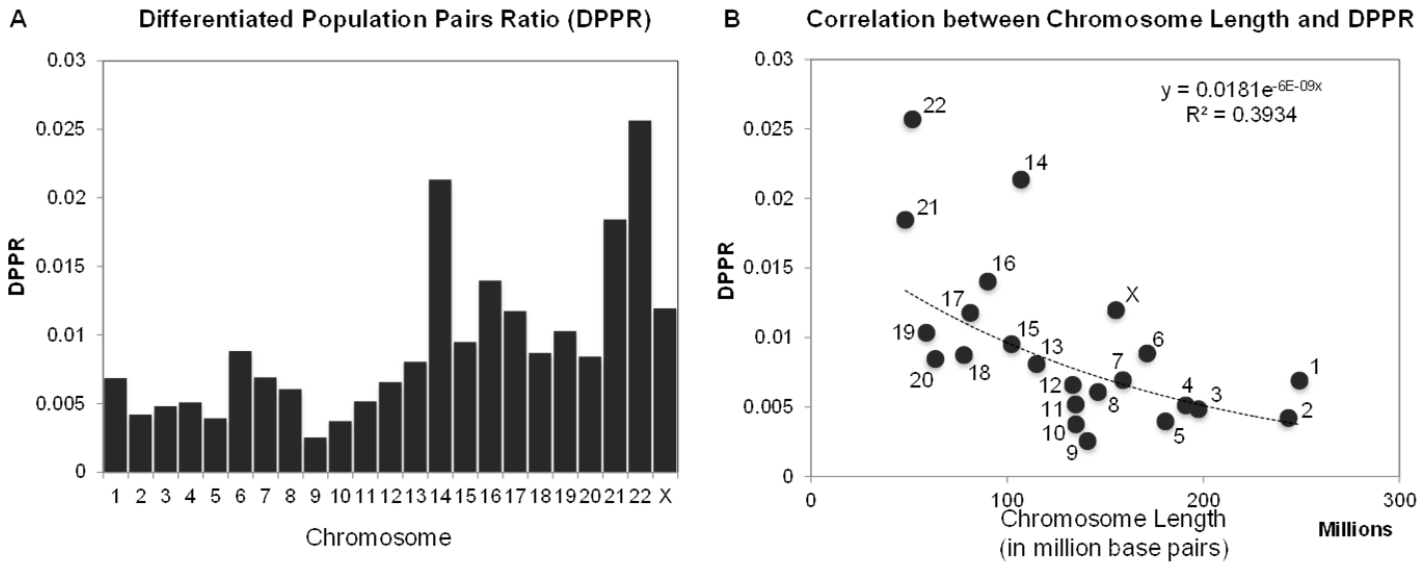


Figure 4.8 Population differentiation and chromosome size - does size matter? Differentiated Population Pairs Ratio (DPPR) is the proportion of the count of observation of highly-differentiated population pairs in a chromosome, normalized by the total number of population-pair differentiation that could have been possibly observed. (A) Smaller chromosomes tend to have larger DPPR as compared to the larger chromosomes. (B) Plotting chromosome size against the DPPR produced a negative correlation fit.

In addition, to eventually look into the possible correlation between chromosome size and the DPPR, a x-y plot was constructed. A negative exponential correlation ($R^2 = 0.3934$) is observed when plotting chromosome size against the DPPR (Fig. 4.8B). Here, it is observed that chromosome 9, which is 141 million basepairs in size, has a DPPR of 0.0025409, the lowest of all chromosomes. In comparison, chromosome 22 is the chromosome with the highest DPPR (0.0256643), which signifies for the more prevalence of population pairs that are affected by the tcdGenes. Chromosome 22 is 51 million basepairs in size.

Figure 4.9 presents the tcdGenes carrying SNPs that are extremely differentiated in 30 or more population pairs, which include the OCA2, GABRG3, HERC2, HLA-DBR1, and HLA-DBR6. The OCA2 and HERC2 genes especially, are known to have been associated with eye colour determination in human.

4.3.6 Enrichment of tcdGenes in pathways

One method for deducing the potential functional significance of population differentiation is to explore the biological pathways that are affected by the tcdGenes. In figure 4.10, the KEGG pathways that are enriched by these tcdGenes are presented. The analysis had taken into account the grouping of populations based on their continental origins. Since the IBS population does not belong to any particular continental cluster (Fig. 4.2B), this population is excluded in this analysis. In Figure 4.10, the African population group is noted to be associated with the highest number of pathways that are significantly enriched by the presence of tcdGenes that are extremely differentiated against the other

populations. These pathways range from one that is associated with *Staphylococcus aureus* infection to autoimmune thyroid disease pathway. One pathway with a significant percentage of tcdGenes only in the African population is the Leishmaniasis pathway. Enrichment of tcdGenes in the cell adhesion molecules (CAMs) pathway on the other hand, is only significant in the Latin Americans. In addition, the data also suggest for the significant enrichment of tcdGenes in olfactory transduction pathways in East Asians, Europeans and Latin Americans, but not in the African population.

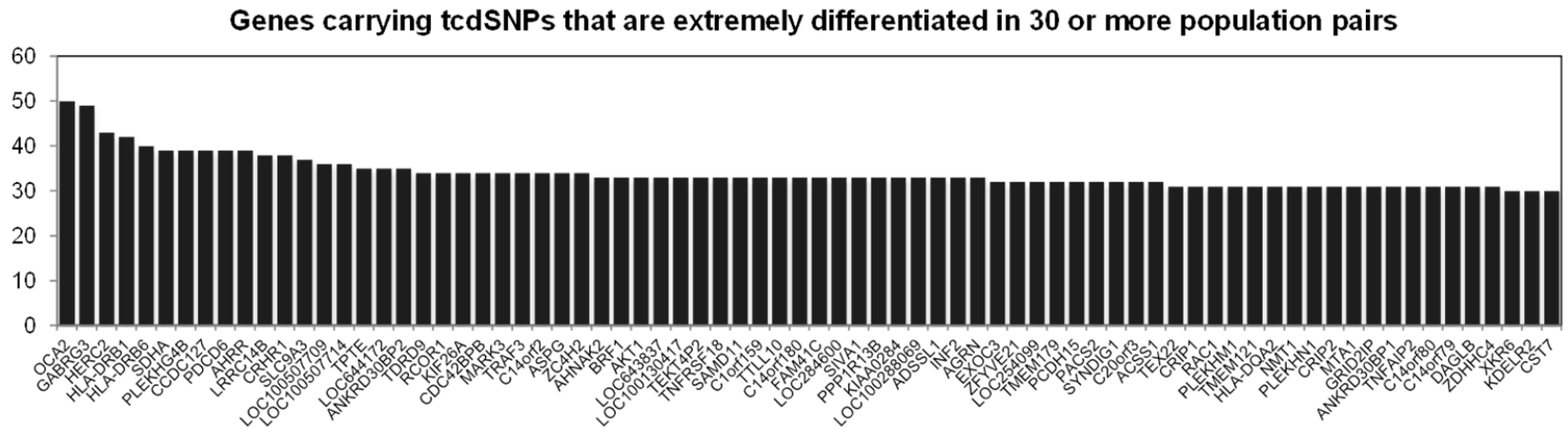


Figure 4.9 Genes with the highest number of population differentiation occurrence.

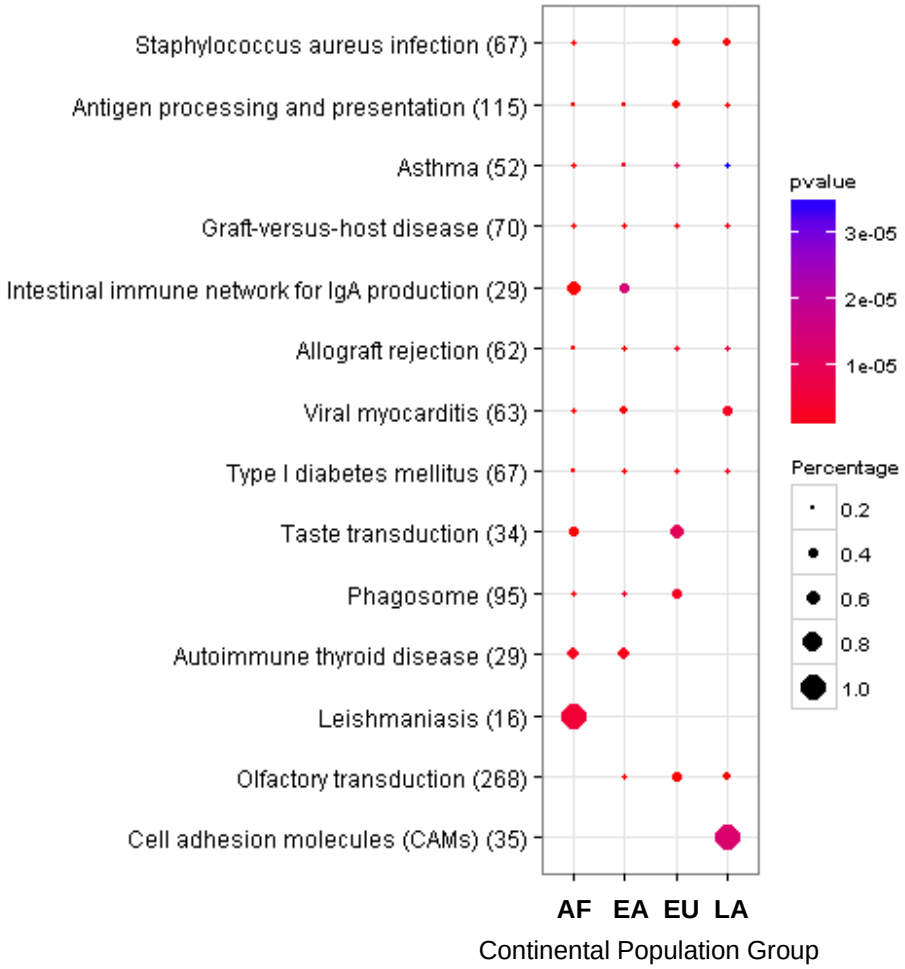


Figure 4.10 Enrichment of tcdGenes in KEGG pathways. Each column represent the significant presence of tcdGenes that are extremely differentiated between one group of populations and the rest of the population groups. The population groups are AF (Africans - ASW, LWK and YRI); EA (East Asians - JPT, CHB, and CHS); EU (Europeans - CEU, GBR, TSI, and FIN); and LA (Latin Americans - MXL, CLM and PUR). The size of the circle corresponds to the proportion of tcdGenes in the pathway whereas the color signifies the enrichment *P*-value.

In addition, Figure 4.11 shows the results of GO molecular functions enrichment analysis. Based on this observation, there are two molecular function groups which are differentiated only in one continental population group. In East Asians for instance, the nucleoside and antigen binding-related molecular function categories are significantly enriched by tcdGenes (shaded purple). For the Latin Americans group, such enrichment can be observed in the molecular functions categories that are relevant to carbohydrate binding; transferase activity; transferring acyl groups; serine-type endopeptidase activity; transferase activity; transferring acyl groups other than amino-acyl groups; protein C-terminus binding; and serine-type peptidase activity (shaded brown). Some if not all of the genes in these group of molecular functions were previously thought to have been more conserved, rather than extremely population-differentiated. Several additional GO annotations that are generally enriched by tcdGenes include the MHC class II receptor and the general umbrella of the molecular function GO annotation itself (shaded green).

The next part of the enrichment analysis takes account canonical pathways collection that are available at the Ingenuity Pathway Analysis (IPA) platform. Here, the antigen presentation pathway, with more than 60 percent of it genes associated with tcdSNPs, is observed to have the lowest statistical *P*-value (Fig. 4.12A). This is followed by the allograft rejection and OX40 signalling pathways. Another notable observation is that in this study, most (8 out of top 10) of the top pathways with significant enrichment of tcdGenes, are relevant to the immune

system. This suggests the importance of population differentiation in these 'immune genes'.

To further assess the pathways enrichment, a deeper stratification based on the three SNPs functional gene region categories were conducted. These are SNPs that are predicted to be functional in the promoter, 3' UTR and coding region, including their potential functional significance in the respective category. The result presented in this part of this chapter suggests a constant enrichment of the Antigen Presentation Pathway, OX40 Signalling Pathway and Allograft Rejection Signalling Pathway by tcdGenes carrying potentially functional tcdSNPs in the promoter (Fig. 4.12B), 3' UTR (Fig. 4.12C) and coding regions (Fig. 4.12D). There is a consistent enrichment pattern, regardless of the SNPs functional categories, suggesting the potential importance of population differentiation in this group of pathways, which is typically relevant to the immune system and is predominated by the HLA genes family.

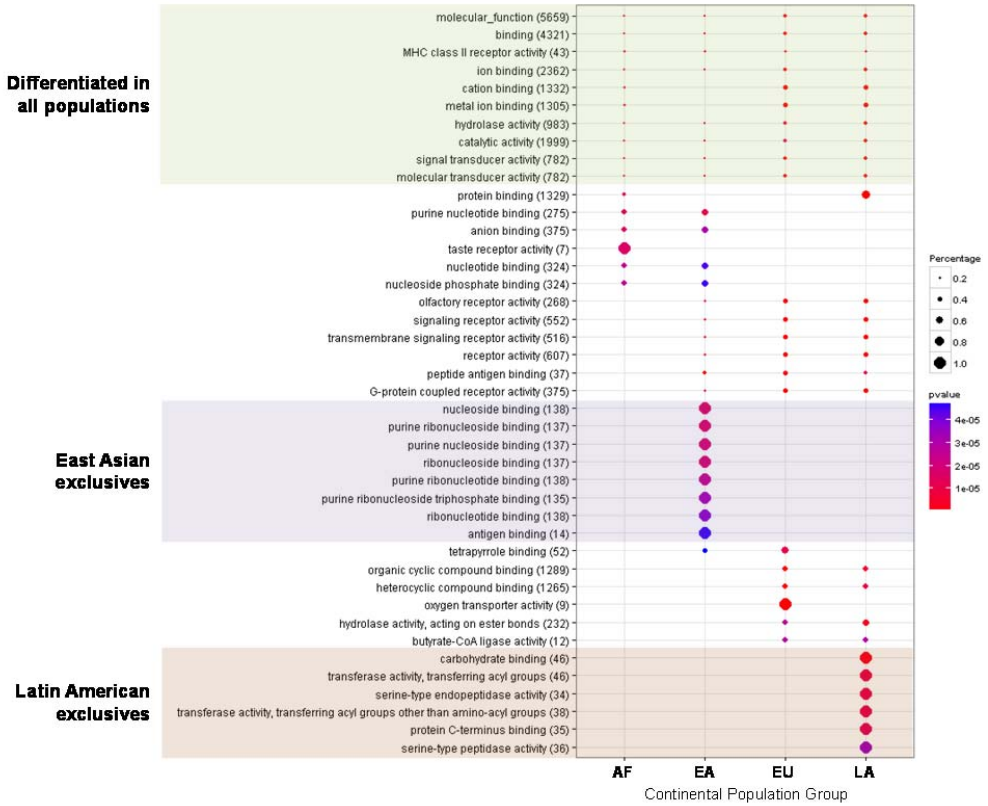


Figure 4.11 Enrichment of tcdGenes based on GO Molecular Function Annotation

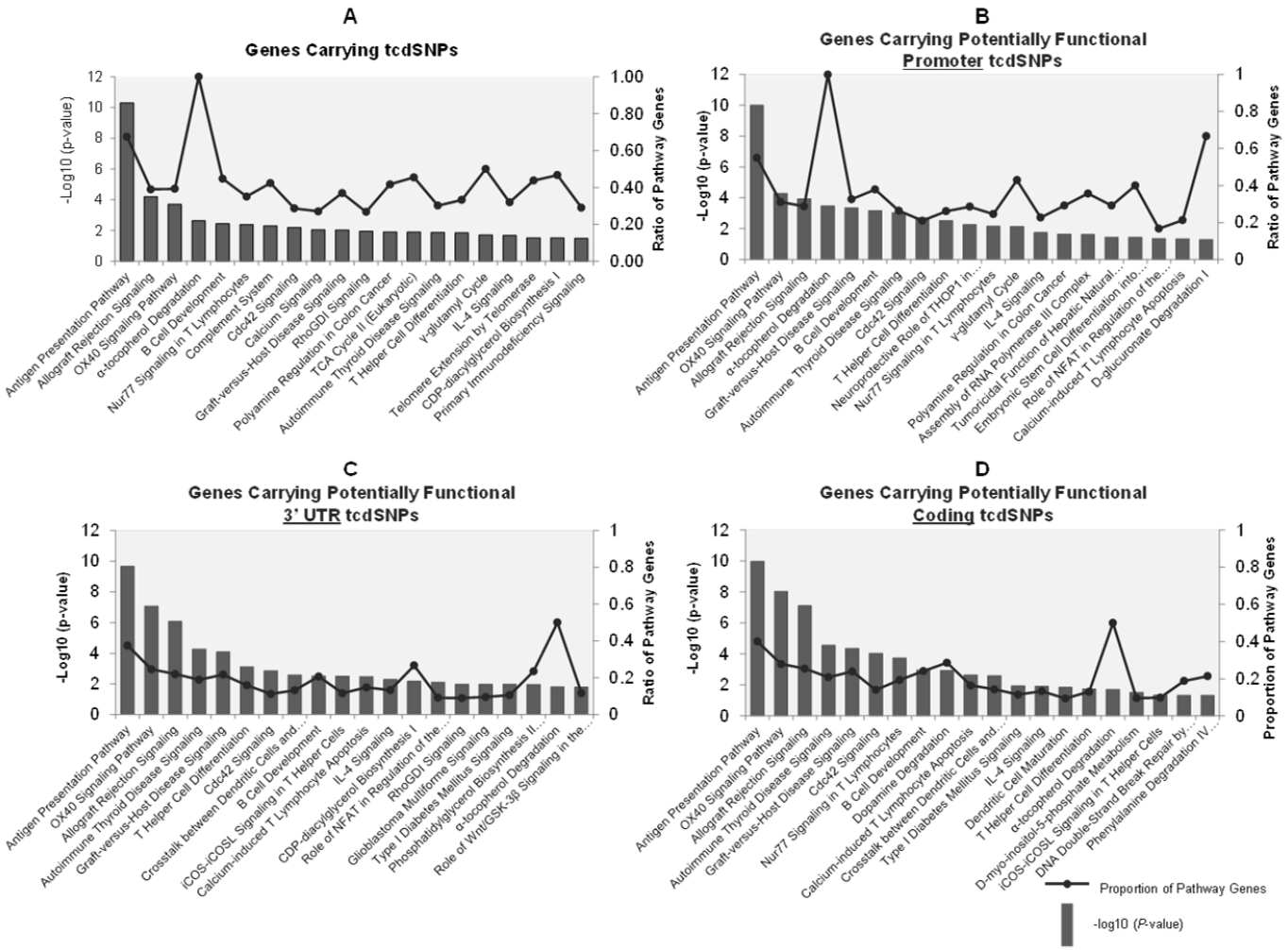


Figure 4.12 Enrichment of tcdGenes in canonical pathways. The bars signify the enrichment *P*-value (primary y-axis) and the line depicts the proportion of such genes in the pathway (secondary y-axis). The analysis was performed with pathways that are made available by the QIAGEN's Ingenuity® Pathway Analysis using all genes carrying tcdSNPs (A), in addition to genes carrying tcdSNPs that are also predicted to be functional in the promoter (B), 3' UTR (C) and coding (D) regions.

4.3.7 Pharmacogenomics utility

The results obtained in this chapter provided a backbone data for the subsequent part of this thesis, in attempting to identify SNPs that are relevant for population differentiation of drug response. In this section of the chapter, I would like to present a pilot project that can bring us closer towards translational pharmacogenomics.

Here, the focus was to investigate population differentiation of genes that are central in pharmacogenomics. In doing so, the first step was to explore whether there are any drug-response gene sets that are associated with high presence of tcdGenes. Therefore, the pilot work involved investigation of the tcdGenes enrichments in a wide array of drug-response gene sets. For each set, a statistical random sampling with 10 thousand times iteration was conducted. The background genes were obtained from the genome, which consisted of 22.5% tcdGenes. In Figure 4.13, drugs/compounds with z-score of more than 1.96, which are enriched by tcdgenes are displayed. A drug with high z-score is implied to have a non-random enrichment of tcdGenes. Top scoring dugs/compounds (z-score > 3) include selenium, vitamin E, bortezomib, arsenic trioxide, vorinostat, cisplatin, decitabine, pentagastrin, and eight additional drugs. In these drugs, the proportion of tcdGenes range from 25 to 100 percent of the drug-response gene set. This is relatively higher than the percentage of tcdGenes in the genome background (22.5%), which is associated with a z-score of 0.016.

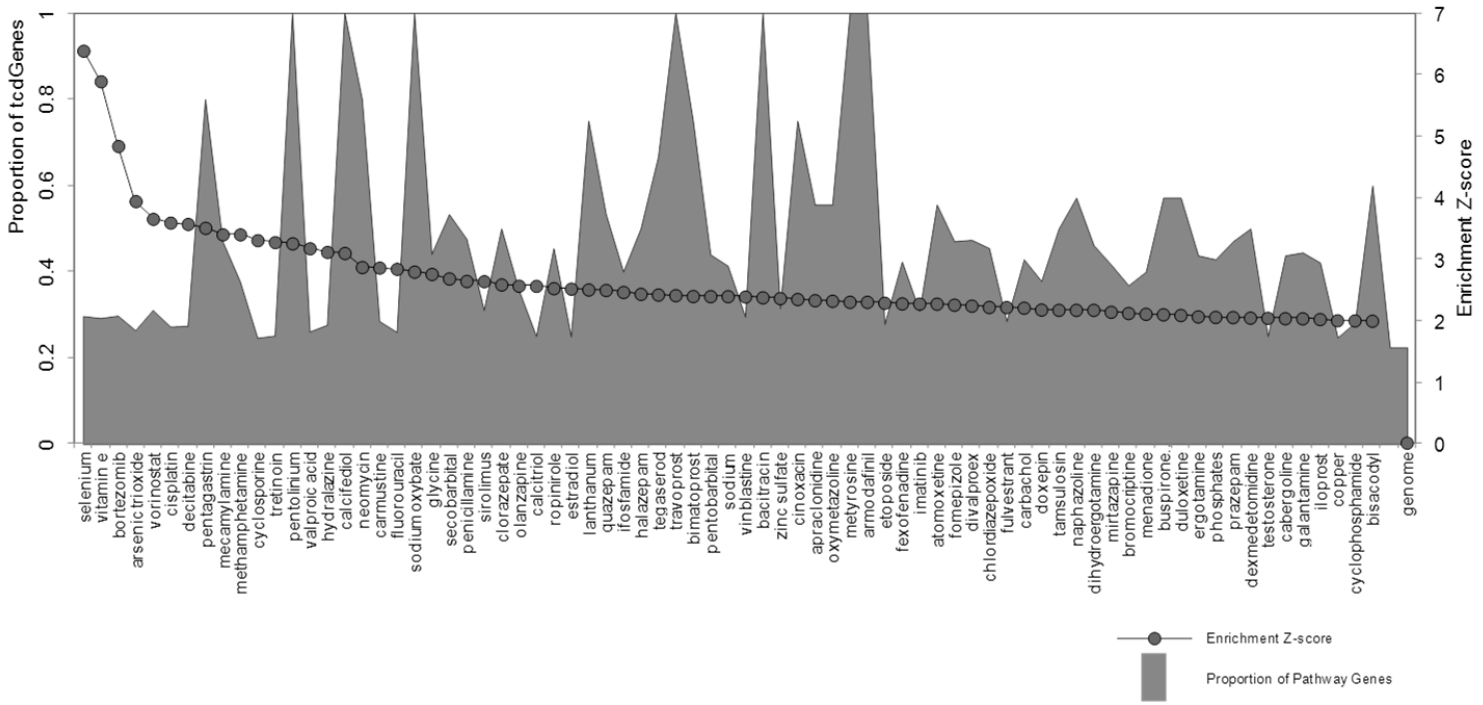


Figure 4.13 Enrichment of tcdGenes in drug pathways.

The top scoring compound, selenium, has a central role in a number of enzymatic reactions in which it acts as a cofactor. Among the sources of this essential micronutrient include yeast breads, meats, poultries, fishes, eggs, and milk [11]. In this study, out of 1,394 genes that can be associated with selenium, 414 are classified as tcdGenes.

4.4 Discussion

The presence of genetic polymorphisms is an important factor that determines variation in phenotype. Having a substantial knowledge of the population differentiation pattern of the human genome is essential, before applying this genomic architecture for pharmacogenomics purposes. One of the main focuses of this chapter was to obtain genome-wide population differentiation data, which can be utilized for multiple purposes, including pharmacogenomics. Unlike other study predecessors, this thesis chapter did not focus on using the tag SNPs or candidate genes approaches in finding population genetic differentiation signature in human. Instead, the focus was to elucidate the differentiation pattern using most SNPs that have been identified in the human genome. This was achieved by computing the F_{ST} scores of almost 23 million SNPs in the human genome, based on allele frequency data that has been released by the 1000 Genomes Project.

In addition, because the 1000 Genomes project employed a next-generation genome sequencing technology, it allowed the examination of most if not the

entire SNPs list from the thousands individuals who participated in this study. This is a significant improvement over the tagging SNPs approach as every SNP could now 'democratically represent itself', rather than be represented by a sample of tag SNPs. Hence, with the exception of mono-allelic SNP or those that are associated with variant call error, no SNP is left out in this study. This would then open up more possibility in finding SNPs that contribute to population differentiation. In general, these SNPs are scattered all around the genome, both within functional genes and non-genic regions.

Moreover, in studying the population genetics factor behind drug response variation, having an in depth knowledge of the human genome background would ensure that no single gene is neglected. Traditionally, pharmacogenomics focuses in conventional genes list that are believed to have been associated with drug response. However, with the availability of the genome sequences of thousands of individuals, we are entering a new era in science and medicine, in which there is a potential translational utility of personal genomic information that may have arisen from non conventional drug-response genes, which are genes that have not been traditionally included in pharmacogenomics studies. In achieving this, the initial step was therefore to first elucidate the whole genome genomic pattern that contributes to population differences in phenotype. This chapter of my thesis examined human population differentiation in the genetic level, and it opened more paths for further examinations on the specific genes, pathways or networks that could potentially affect a whole range of phenotype differences, including drug response.

One important contribution of this chapter is in the discovery of novel knowledge surrounding population diversity. Here, it is observed that there is relatively less differentiation within the East Asian or European populations. In contrast, this 'local differentiation' is greater among the Africans as compared to the other population groups. Furthermore, the approach that was adapted in this chapter is hypothesis free, which involved a non-guided process of scanning the entire genome for SNPs clusters that contain the maximum population-differentiation. A significant advantage of this approach is the non-biased identification of the chromosome regions that are extremely differentiated in the respective population. It also has the potential to decrease false negative rate, as tcdSNPs were identified in all chromosomes. In this chapter, the 30 millionth base region of chromosome 6 or a region surrounding loci 6p21 is identified to be a hotspot of population differentiation, particularly between populations that are closely related. The presence of population-differentiated SNPs in closely related populations suggests that it potentially has more significance in determining differentiation in individual level, but not in the population level. In population pairs that are more genetically different, this extreme differentiation signature is probably masked by the presence of other high F_{ST} SNPs that would eventually be differentiated anyway, due to natural selection that takes place in parallel with geographic divergence during early human migration. Here, the most genetically similar population pair, the CEU-GBR, was used as a representative in studying the genes that are potentially affected by tcdSNPs in this 30 millionth base region of chromosome 6. In this study, 35 genes were identified, including the HLA-C and

TRIM gene families, which highlight the possible importance of these genes in determining inter-individual differences.

Moreover, a comparison of the population differentiation pattern of SNPs between the genic and intergenic regions could support the existing knowledge in the field. It is observed that the intergenic region, which does not affect protein structure, is more inclined to carry greater magnitude of population genetic differentiation as compared to the gene region. This supports the result of the previous chapter (see 3.3.4) where a high population differentiation in regulatory regions was observed. Nonetheless, it is worth noting that the statistical significance could possibly been achieved as a result of the relatively large sample size that was involved in the statistical *t*-test calculation.

It is noted that the results presented in this chapter suggest that there is a generally higher proportion of tcdSNPs that are potentially functional in the coding region despite the lower number of SNPs compared to that of the non-coding region. And because they can potentially affect protein function, through the variation in protein 3D structure, domain functional and post-translational modification, the finding could further suggests on the potential bridge between genetic differentiation and the variation that could be seen in phenotype. However, a general conclusion on this phenomenon is not advisable at this current stage because as shown by the subsequent result, the routes at which these genetic differences are translated to variation in gene functions can vary depending on the population. For example, East Asians have more potentially functional tcdSNPs in the promoter region, such as those residing on transcription factor binding sites. In

comparison, the Latin Americans, including the MXL and CLM, have higher proportion of potentially functional tcdSNPs in the coding region.

There is also a possible link between population differentiation and chromosome size. Using the differentiated population-pairs ratio (DPPR), it is observed that the smaller chromosomes are subjected to more population differentiations. This could be related to the chromosome recombination rate, which could be dependent on chromosome size [12]. Nonetheless, at the current phase, deducing a biological scenario from this observation will be too premature as we need to further investigate on the potential causal-and-effect relationship between the population differentiation phenomenon and chromosome size.

Enrichment of tcdGenes in pathways that contribute to human phenotype would open up the link between genetic differences and phenotype variation. The pathways for olfactory transduction and immune system-related functions, which are associated with inter-individual differences in smell and defence mechanism, respectively, are enriched by tcdGenes. The identification of olfactory transduction and immune system-related pathways is not unexpected because these pathways are already known to have great genetic variability [13-15]. And this finding could indeed act as a 'positive control' that the approach employed here is effective in finding tcdGenes. Intriguingly, during the course of work in this chapter, it is also observed that pathways that have been traditionally thought to be associated with conserved genes were indeed found to be enriched by tcdGenes. These enrichments are observed in the cellular mechanism- and

metabolic-related pathways including the Cell Adhesion Molecules (CAMs), Nucleoside Bindings and Transferase Activities.

Alleles that pose greater reproductive and/or survival advantage in different geographical regions would overtime be positively or negatively selected. In this thesis, using population differentiation pattern, it is possible to elucidate more patterns that corroborate this assumption. SNPs in the maximum-differentiated clusters can potentially determine differences of phenotypes in different population. In this chapter, OCA2 is identified as the gene that has the greatest number of observation of population-pair differentiation. Its role in the production of skin colour determinant is well-studied [16]. Because skin colour is a phenotype that is highly differentiated between human populations, this finding serves as a concrete example on the effectiveness of the hypothesis-free approach in identifying tcdSNPs that are extremely differentiated between diverse populations. Furthermore, another gene that is identified to have a high number of population-pair differentiations is the HERC2. Similar to OCA2, this gene is involved in determining skin colour, in addition to hair and eye pigmentation, which are phenotypes that are associated with population or racial differences [17, 18]. The effectiveness of this approach in identifying tcdSNPs is further strengthened by the identification of such SNPs in the HLA loci, which contain important immune system genes and is well known for its highly variable nature. As a matter of fact, most pathways that are enriched by tcdGenes, including those that contain functionally important tcdSNPs, are immune system-associated pathways.

It is important to note however that despite this intriguing observation of population differentiation in human genes, phenotype differences are attributed to multifactorial components. Besides inherent factor such as genetic polymorphisms, external factors such as socio-economic influence, culture and other environmental factors are known to affect human differences. Studying all these factors at the same time however, is challenging, hence it is best to answer this question a single factor at the time [19, 20]. Due to this reason and because it is a more constant variable, studying the genetic factors behind phenotype difference is therefore still the most viable option. Genetic polymorphisms are known to have been associated with differences in foetal development, immune system and environmental response [21]. It is therefore in the interest of my thesis to deeply investigate the novel pattern of population differentiation in the human genomes.

Theoretically, the differences in SNPs allele frequency are caused by the variation selection pressures during human migrations. This phenomenon contributed to the allele frequency differentiation of human populations [21-23], which have been associated with the variation of phenotype in different individuals. Hence, everyone is different. We carry different traits that are manifested in differences of phenotype, including our physical shape, disease susceptibility and drug response. The differences that we see across individuals are in turn seen on a population level, where population differentiation in human phenotype is seen as a common phenomenon.

Based on the findings that are presented in this chapter, we could observe the complexity of population genetic variation in our genome. As also shown in the previous chapter, decoding the genome is not a one layer effort as it involves a multilayer work with imagination as the limit. Therefore currently, many of our efforts in understanding the pattern of population differentiation in the genome would raise more questions for future studies. The strategy is not to bring too much complexity as with current limitation, it is not possible to tackle all layers at the same time. In this chapter of my thesis, it is possible to discover a novel pattern of population genetic differentiation using a simplified strategy of scanning the genome for maximum population-differentiated SNPs cluster. This contributes one significant foundational layer to the field, which can serve as a module for another layer. I envision that in the not-so-distant future, genetic polymorphisms can be used as an individual's molecular identity, which include early recognition of drug response profile of a person based on SNPs information.

The pilot pharmacogenomics work was conducted to serve this purpose, where the data that is generated in this chapter was brought forward closer to translational impact. As drug therapy often yielded a highly varied clinical outcome, identifying drug-response gene sets or pathways that are enriched by tcdGenes could provide a new method of prioritizing drugs that have higher potential to be associated with response differences. It has been reported in many drugs that SNPs are a factor that contribute to the variation of drug metabolism, transport and efficacy [24].

For the last part of this chapter, the pilot pharmacogenomics work had preliminary identified enrichment of tcdGenes in several widely-used drugs or compounds. This include selenium, vitamin E, bortezomib, cisplatin, and decitabine. Cisplatin has particularly been reported to be associated with inter-individual or population differences in response [25-27]. This shall therefore bring a spotlight on the potential of this population genomics approach in finding the genes that would eventually contribute to exerting phenotype differences across populations. And it will be especially significant in pharmacogenomics where population differentiation is a major concern. In the next chpater, I will utulize the knowledge that has been generated in this chapter to elucidate the genomic basis of population differences in drug response.

4.5 Reference

1. Abecasis, G.R., et al., *An integrated map of genetic variation from 1,092 human genomes.* Nature, 2012. **491**(7422): p. 56-65.

2. Weir, B.S. and C.C. Cockerham, *Estimating F-Statistics for the Analysis of Population Structure.* Evolution, 1984. **38**(6): p. 1358-1370.

3. Holsinger, K.E. and B.S. Weir, *Genetics in geographically structured populations: defining, estimating and interpreting F(ST).* Nature reviews. Genetics, 2009. **10**(9): p. 639-50.

4. Kanehisa, M., et al., *The KEGG resource for deciphering the genome.* Nucleic acids research, 2004. **32**(Database issue): p. D277-80.

5. Ashburner, M., et al., *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.* Nature genetics, 2000. **25**(1): p. 25-9.

6. Yu, G., et al., *clusterProfiler: an R package for comparing biological themes among gene clusters.* Omics : a journal of integrative biology, 2012. **16**(5): p. 284-7.

7. Whirl-Carrillo, M., et al., *Pharmacogenomics knowledge for personalized medicine.* Clinical pharmacology and therapeutics, 2012. **92**(4): p. 414-7.

8. Wishart, D.S., et al., *DrugBank: a knowledgebase for drugs, drug actions and drug targets.* Nucleic acids research, 2008. **36**(Database issue): p. D901-6.

9. Gaulton, A., et al., *ChEMBL: a large-scale bioactivity database for drug discovery.* Nucleic acids research, 2012. **40**(Database issue): p. D1100-7.

10. Mattingly, C.J., et al., *The Comparative Toxicogenomics Database (CTD).* Environmental health perspectives, 2003. **111**(6): p. 793-5.

11. Rayman, M.P., *The importance of selenium to human health.* Lancet, 2000. **356**(9225): p. 233-41.

12. Jensen-Seaman, M.I., et al., *Comparative recombination rates in the rat, mouse, and human genomes.* Genome Res, 2004. **14**(4): p. 528-38.

13. Hasin-Brumshtein, Y., D. Lancet, and T. Olender, *Human olfaction: from genomic variation to phenotypic diversity.* Trends Genet, 2009. **25**(4): p. 178-84.

14. Young, J.M., et al., *Extensive copy-number variation of the human olfactory receptor gene family.* Am J Hum Genet, 2008. **83**(2): p. 228-42.

15. Satija, R. and A.K. Shalek, *Heterogeneity in immune responses: from populations to single cells.* Trends Immunol, 2014. **35**(5): p. 219-29.

16. Sulem, P., et al., *Genetic determinants of hair, eye and skin pigmentation in Europeans.* Nature genetics, 2007. **39**(12): p. 1443-52.

17. Branicki, W., U. Brudnik, and A. Wojas-Pelc, *Interactions between HERC2, OCA2 and MC1R may influence human pigmentation phenotype.* Annals of human genetics, 2009. **73**(2): p. 160-70.

18. Eiberg, H., et al., *Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the HERC2 gene inhibiting OCA2 expression.* Human Genetics, 2008. **123**(2): p. 177-187.

19. Gingeras, T.R., *Origin of phenotypes: genes and transcripts.* Genome Res, 2007. **17**(6): p. 682-90.

20. McCarthy, M.I., et al., *Genome-wide association studies for complex traits: consensus, uncertainty and challenges.* Nat Rev Genet, 2008. **9**(5): p. 356-69.

21. Barreiro, L.B., et al., *Natural selection has driven population differentiation in modern humans.* Nature genetics, 2008. **40**(3): p. 340-5.

22. Balaresque, P.L., S.J. Ballereau, and M.A. Jobling, *Challenges in human genetic diversity: demographic history and adaptation.* Human molecular genetics, 2007. **16 Spec No. 2**: p. R134-9.

23. Campbell, M.C. and S.A. Tishkoff, *The evolution of human genetic and phenotypic variation in Africa.* Current biology : CB, 2010. **20**(4): p. R166-73.

24. Bachtiar, M. and C.L. Lee, *Genetics of Population Differences in Drug Response.* Current genetic medicine reports, 2013. **1**(3): p. 162-170.

25. O'Donnell, P.H., et al., *Population differences in platinum toxicity as a means to identify novel genetic susceptibility variants.* Pharmacogenet Genomics, 2010. **20**(5): p. 327-37.

26. Watanabe, A., M. Taniguchi, and S. Sasaki, *Induction chemotherapy with docetaxel, cisplatin, fluorouracil and l-leucovorin for locally advanced head and neck cancers: a modified regimen for Japanese patients.* Anticancer Drugs, 2003. **14**(10): p. 801-7.

27. Huang, R.S., et al., *Identification of genetic variants contributing to cisplatin-induced cytotoxicity by use of a genomewide approach.* Am J Hum Genet, 2007. **81**(3): p. 427-37.

Chapter 5. Elucidating the Genomic Basis of Drug Response Variation with Population-differentiated SNPs

5.1 Introduction

As introduced in the initial part of this thesis, drug response variation is common. A drug that is effective in one population may not be equally beneficial when prescribed to a population with different background; or worse if it could pose adverse drug reaction (ADR). Moreoever as seen in previous chapters, population differentiation in the human genome could manifest as a factor that contribute to population differences in drug response.

Using the 1000 Genomes Project data, the results of scanning for genomic population differentiation patterns were presented in Chapter 4. These patterns are derived from the identification of top chromosome differentiated SNPs (tcdSNPs) and the corresponding top chromosome differentiated genes (tcdGenes). These tcdSNPs when found in genes that are responsible for drug response, are potential factors that could affect drug response variation. Hence in this chapter, using the tcdGenes that was obtained from the previous chapter, the focus is narrowed down to the genes that have been reported to be associated with drug response. The work in this chapter aimed to unleash this potential benefit by utilizing these tcdGenes information and developing a prototype pharmacogenomics application.

In doing so, drugs that potentially have strong population differentiation profile will be identified using their respective gene sets.

The work in this thesis chapter involves massive integration work of various data types. Among them are tcdSNPs and tcdgenes information, in addition to pharmacogenomics annotation that allow linking population differentiation information of SNPs and genes to various drugs/compounds. In general, the first type of information consisted of 'big-data genomics'. This includes SNPs population differentiation scores, which is measured by the F_{ST} statistics [1] that was derived from individuals who participated in the 1000 Genomes Project. The second breadth of information contains drug/compound names that were mined from four major databases namely: the PharmGKB [2], DrugBank [3], ChEMBL [4], and Comparative Toxicogenomics Database (CTD) [5]. In addition to drug/compound names, this information includes annotation of drug classifications which was derived from the WHO Anatomical Therapeutic Classification (ATC) system. The drug-response gene sets were then created by inter-linking these datasets, which allowed the formation of an integrated resource called the 'PharmaSNP'.

This knowledge integration could serve as a connecting bridge between population genomics and pharmacogenomics, in delivering its potential translational application. With this approach, it is possible to identify the drugs that are linked to significant number of population-differentiated genes (tcdGenes) and elucidate the gene sets that could possibly affect drug response variation in different populations. The subsequent part of this chapter hence presented the

result of drugs clustering based on their population differentiation profile. This profile was derived from enrichment analysis of tcdGenes that are linked to these drugs.

5.2 Methods

5.2.1 The 'next generation' pharmacogenomics genes and the PharmaSNP resource

In this chapter, in addition to using conventional drug-response genes that are annotated by the PharmGKB database, the gene-drug association information were also obtained from data mining of three other databases namely: DrugBank [3], Comparative Toxicogenomics Database (CTD) [5], and ChEMBL [4]. This has resulted in the production of 'next generation' pharmacogenomics genes that encompasses a much greater number of genomic genes and variants that could possibly play important role in pharmacogenomics.

This next generation pharmacogenomics genes collection is stored in the SQL database engine and is publicly available in the PharmaSNP resource that was constructed in the course of producing this thesis. The PharmaSNP, a PHP-based web portal, contains information on genes that have been reported to be associated with drug activities such as its pharmacokinetic, pharmacodynamic, toxicity and other cellular effects. It is accessible at http://bit.ly/pharma-snp and will be elaborated further in the next chapter.

In addition, the drug data mining that was performed for the purpose of this chapter yielded an extensive list of drugs and compounds. Here, in order to be closer towards its translational application, it was decided to adopt an inclusion criteria. Only drugs or compounds that have been approved by the U.S. Food and Drug Administration (FDA) were included in the analysis. The list of approved drugs was obtained from the FDA Approved Drug Products with Therapeutic Equivalence Evaluations, which is also referred as the Orange Book [6]. For classification purposes, drug names were also associated with the various drug types in accordance to the WHO Anatomical Therapeutic Classification (ATC) system. These drugs are classified based on the relevant organ or system localization of the drug, in addition to other properties including chemical, pharmacological and therapeutic information.

5.2.2 Population genetic differentiation in drug-response genes

Population genetic differentiation was estimated using SNPs allele frequency data that have been presented in Chapter 4, which was obtained from the 1000 Genomes Project (Phase 1) [7, 8]. In this study, the SNPs allele frequency is derived from 1,092 unrelated individuals who are originated from 14 global populations as described in the previous chapter. For each SNP, pairwise population differentiation scores were calculated by the use of F_{ST} statistics [1].

In this analysis, SNPs that are considered to be extremely population-differentiated were identified. These SNPs, which are referred as top chromosome differentiated SNPs (tcdSNPs), were identified using an algorithm that detects

SNPs clusters carrying the highest moving average pairwise F_{ST} score across chromosomes as of conducted in the previous chapter. Moreover as explained in the previous chapter, top chromosome differentiated genes (tcdGenes) are defined as those that carry one or more tcdSNPs. Subsequently, a subset of these tcdGenes is found in the PharmaSNP database and these would be the drug-response genes that are defined as extremely population-differentiated.

5.2.3 Random sampling enrichment analysis

For all FDA-approved drugs/compounds that can be associated with one or more drug-response genes, the enrichment of tcdGenes was analyzed. The drug z-score was obtained by performing a 10,000 time random sampling iterations with genome genes that are within similar size range. Briefly, for each sampling set, the proportion of tcdGenes found in the random sample was recorded. The process was repeated 10,000 times, and based on these iterations, an empirical distribution specific to the drug in question was generated. The drug z-score was derived from the relative position of the observed tcdGenes proportion in this empirical distribution, which can be calculated with the following equation.

$$\text{Z score} = \frac{\text{p-eP}}{\text{eSD}}$$

Where:

p = proportion of extremely population-differentiated genes in drug-response gene set

eP = average proportion of tcdGenes in empirical distribution of the

respective drug

eSD = standard deviation of empirical distribution of the respective drug

A drug that is significantly enriched by tcdGenes typically has a z-score of more than 1.96 or within the 0.05 percentile of its empirical distribution. On the other hand, a drug that is absent by the presence of such genes will typically have low z-score.

5.2.4 Drugs and population cluster analysis

The population genetic data is potentially useful in pharmacogenomics, where a drug population differentiation in response could possibly be predicted using the enrichment z-scores obtained in the above random sampling exercise. Subsequently, a two dimensional matrix was produced using the z-scores of multiple drug-response gene sets in 91 population pairs. With this matrix, a heatmap that is ordered based on hierarchical clustering of both the drug z-scores and population pair differentiation was generated. This was done by utilizing the heatmap.2 function available in the gplots R package [9].

In order to optimize the method, the heatmap and cluster generation was performed in drugs that have no null z-score value. Furthermore, because these z-scores are derived from a standardized empirical distribution that was generated from each set of random samples, no additional scaling was conducted for the heatmap generation.

The word cloud visualization was performed in wordle.net by utilizing the non-duplicated names of the drug classes.

5.3 Results

5.3.1 Drug-response genes repertoire

The drug and drug-response genes collection were obtained from mining four data sources namely Chembl 13, PharmGKB, Comparative Toxicogenomics Database (CTD), and DrugBank. Figure 5.1 shows the results of this data mining, in which we obtained 10,902 unique drug/compound names in total. In descending order the CTD, DrugBank, Chembl 13, and PharmGKB contained 5,900; 4,184; 2,863; and 194 drugs or compounds information, respectively (Fig. 5.1). With 997 drug names, the overlap is greater between the CTD and Chembl 13 than the rest of the databases.

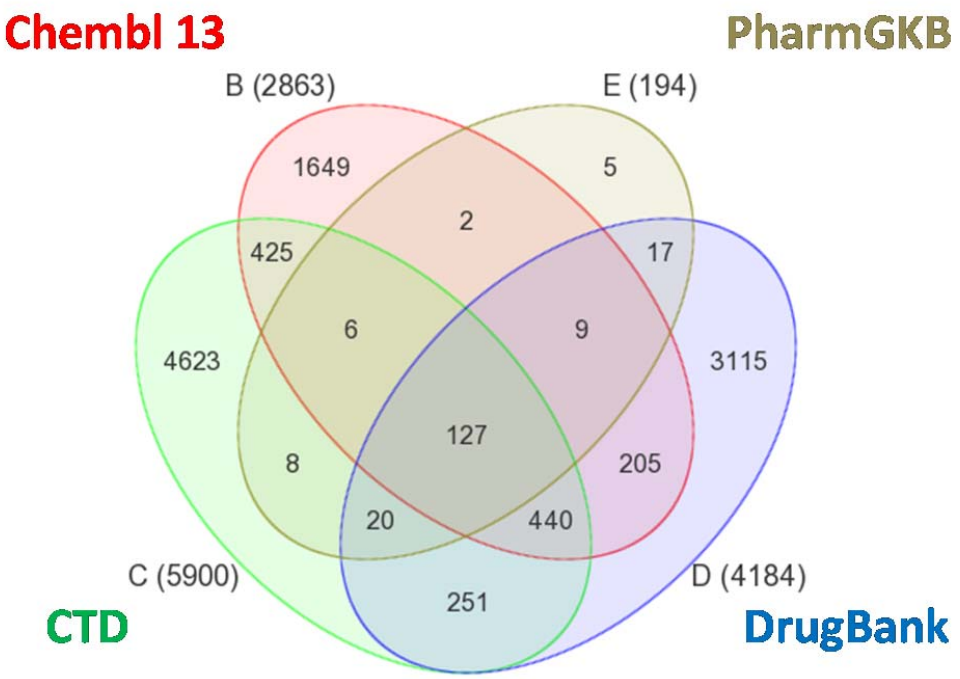


Figure 5.1 Drug names obtained from mining four major databases: Chembl 13, PharmGKB, Comparative Toxicogenomics Database (CTD), and Drug Bank.

Out of 10,902 drugs in total, 1,511 are approved by the US FDA. These 'approved drugs' were utilized for analysis of population differentiation pattern in this study. From this drug data mining, we were able to collect 16,357 genes that are somewhat associated with drugs or compounds available at the four data sources. Similarly, the majority of these genes (16,185) are available at the CTD.

As shown in Figure 5.2, the top 10 compounds with the most number of gene sets are cyclosporine (4,188 genes), tretinoin (2,681 genes), estradiol (1,899 genes), calcitriol (1,755 genes), arsenic trioxide (1,733 genes), copper (1,427 genes), valproic acid (1,426 genes), selenium (1,394 genes), vitamin E (1,369), and fluorouracil (1,165 genes).

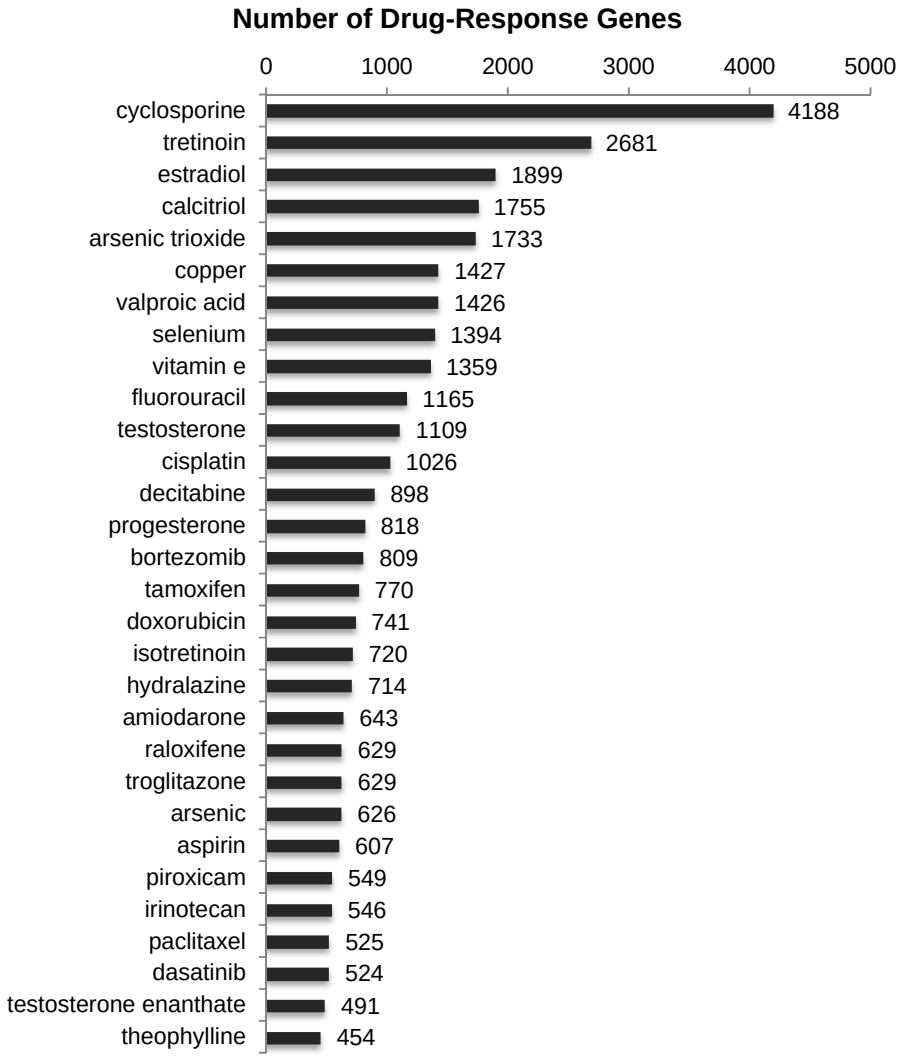


Figure 5.2 Drugs with the highest number of gene sets.

5.3.2 Enrichment of tcdGenes in drug response gene sets

Population differentiation in drug-response genes were estimated by using the F_{ST} scores of SNPs in these genes. Drug-response genes are considered to be extremely population-differentiated if they carry one or more tcdSNPs. These

genes are referred as ‘tcdGenes’. Out of 14,166 drug-response genes that were successfully mined from the four databases, 3,108 (21.9%) are considered to be extremely population-differentiated. Furthermore, with 91 populations that were involved in the 1000 Genomes project, it is possible to calculate differentiation across 91 pairs of population. Because population differentiation was calculated based on a population pair approach, it is possible to identify the genes that are extremely population-differentiated in a specific population pair.

In order to access the significance of tcdGenes in the drug-response gene sets, a random sampling simulation was performed. In the previous chapter, the CHS-YRI and CEU-GBR pairs were identified as the two most distant and most similar populations, respectively. Figure 5.3 presents the top 20 drugs-response gene sets that are enriched by tcdGenes that are extremely differentiated in the two extreme population pairs CHS-YRI and CEU-GBR. In the CHS-YRI pair, where population genetic differentiation is the greatest, sincalide, azithromycin, nalidixic acid, levofloxacin, and menthol are among the drugs or compounds that are enriched by population-differentiated genes. 10,000 times random sampling analysis in these five drugs yielded a z-score between 5.07 and 8.21 in menthol and sincalide, respectively.

On the other hand, random sampling analysis with tcdGenes that are extremely differentiated in the CEU-GBR population pair yielded different conclusion. In this population, enrichment of extremely population-differentiated genes is highest in phytonadione, bacitracin, consyntropin, alprostadil, and ammonium

chloride (Fig. 5.3). The z-scores of these five drugs range from 4.24 in ammonium chloride to 8.3 in phytonadione.

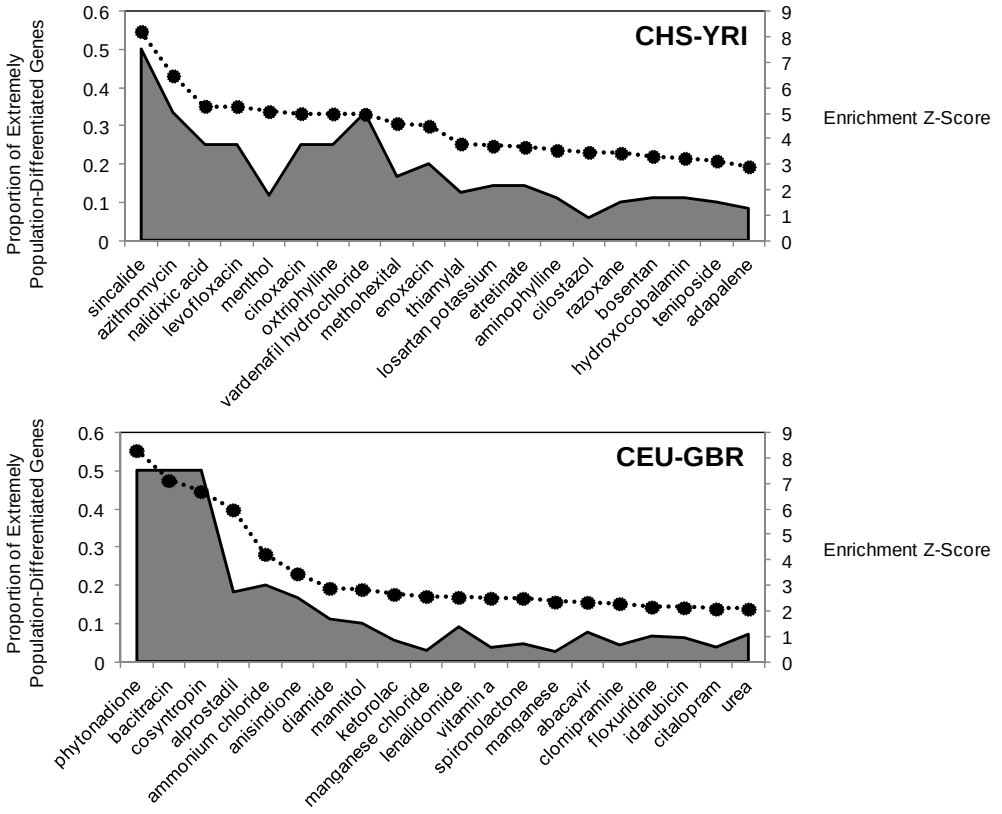


Figure 5.3 The top 20 drugs that are enriched by extremely population-differentiated genes in the two most distant and most similar populations of CHS-YRI and CEU-GBR, respectively. The proportion of extremely population-differentiated genes is presented by the shaded area that corresponds to the primary y-axis. The enrichment z-score is presented by the dotted line that corresponds to the secondary y-axis. A z-score greater than 1.96 typically signifies an enrichment of extremely population-differentiated genes in the particular drug.

5.3.3 Drugs clustering based on population differentiation profile

In identifying the drugs or group of drugs with similar population differentiation profile, this part of my thesis presents a hierarchical cluster analysis. The first step involved grouping the populations into their continental groups. In this regard, the

ASW, LWK and YRI were grouped into Africans (AF), whilst the JPT, CHS and CHB were categorized as East Asians (EA); CEU, GBR, TSI, and FIN as Europeans (EU); in addition to the MXL, CLM and PUR which are grouped into Latin Americans (LA). The IBS population, due to its relatively low similarity to any of the continental group (as seen in section 4.3.1), was excluded in this analysis.

Following categorization of these populations into four continental groups, a matrix was constructed, with the continental population pairs and drugs positioned as the matrix's columns and rows, respectively. Figure 5.4 shows a heatmap that was generated by utilizing this matrix data as input. Drugs with null z-score value in one or more population pair, which could be due to insufficient population genetics data, would automatically be excluded from analysis. In this heatmap that consists of 141 drugs, both axes were arranged with hierarchical clustering method.

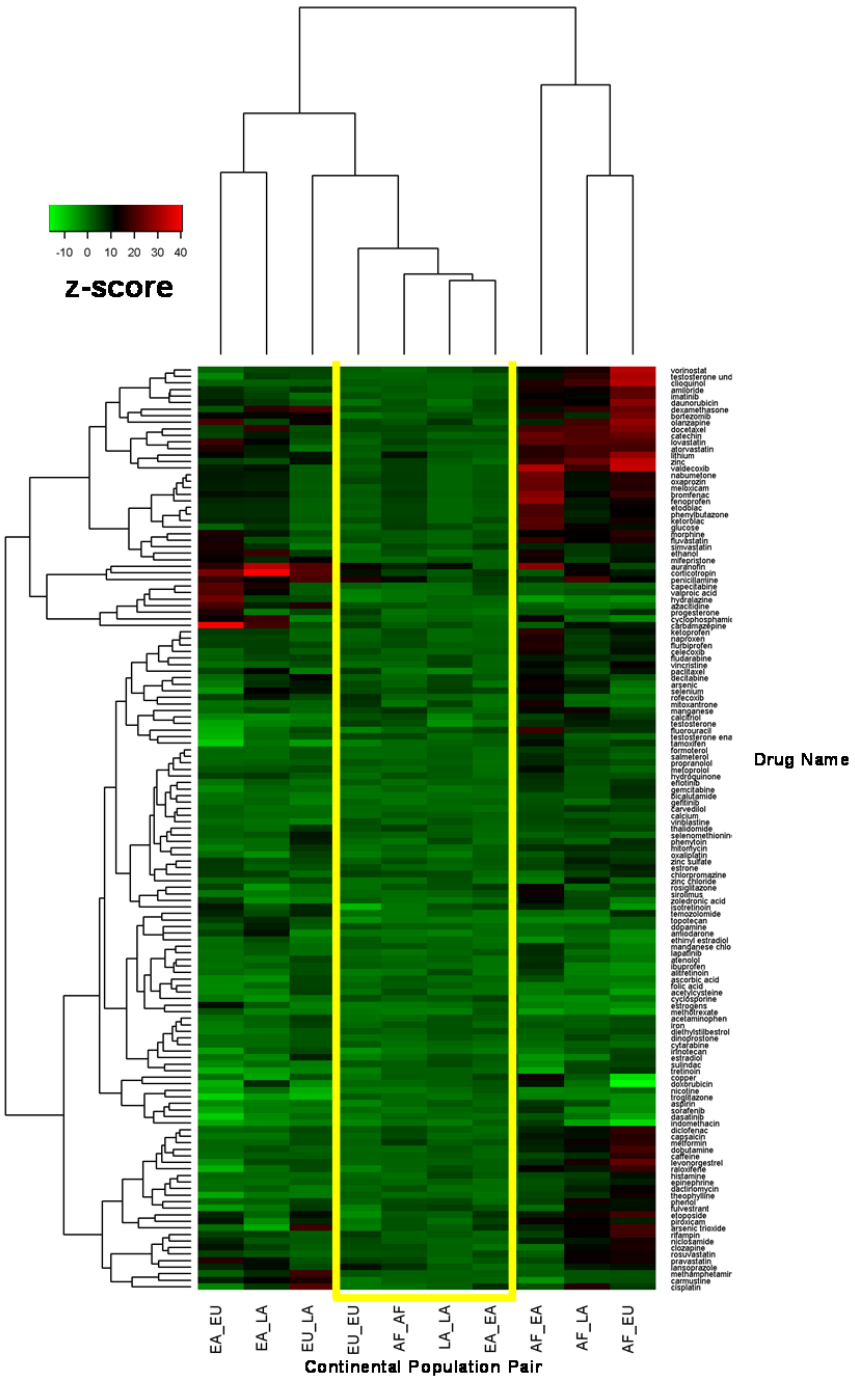


Figure 5.4 A clustered heat-map generated using the enrichment z-scores of 141 drugs. Each row in the heat-map represents a drug population differentiation profile, with z-scores corresponding to a drug name. These z-scores represent the degree of enrichment of tcdGenes that are linked to these drug, in the respective population pairing (column).

From this result, it can be observed that within continental group comparisons, which are consisted of closely related population pairings, most if not all drugs are observed to have relatively low z-scores (green colour cells inside yellow colour box in the heatmap). With the exception of few compounds such as lithium, auranofin and corticotropin, most drugs are not found to be enriched by tcdGenes that are differentiated among the EU-EU, AF-AF, LA-LA, and EA-EA within continental population pair comparisons. On the other hand, comparisons between population pairs that originate from two different continents, such as the EA-EU, EA-LA, EU-LA, AF-EA, AF-LA, and AF-EU, showed a number of drugs that are significantly enriched by tcdGenes (outside of yellow colour box in heatmap). Notably, this enrichment is more frequent in the comparisons involving the African populations, such as in the AF-EA, AF-LA and AF-EU group comparisons.

By taking the above observation into consideration, a second clustering step was performed. In this step, the hierarchical clustering took account only of the data involving different continental pairs, which resulted in the inclusion of 173 drugs. As observed in Figure 5.5, cluster analysis with six continental population groups (EA-EU, EU-LA, EA-LA, AF-EA, AF-LA, and AF-EU), yielded two distinct groups of drugs based on their population genetic differentiation profile.

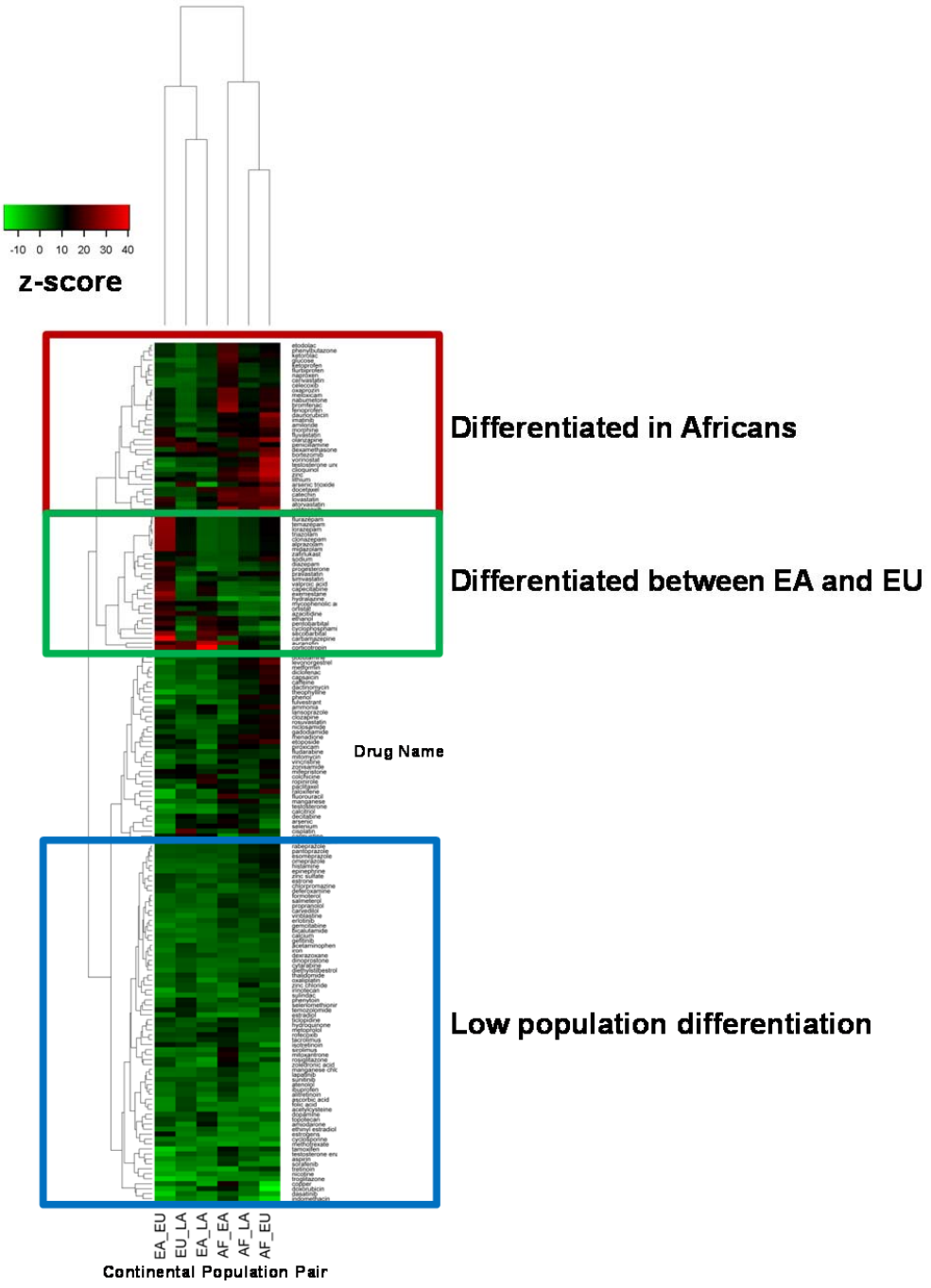


Figure 5.5 A heat-map generated after a second step of cluster analysis. This clustered heat-map was generated by including only the z-scores representing enrichment of tcdGenes that are specific to different continental population pairs.

As shown in Figure 5.5 (inside red colour box), the first cluster consisted of 34 drugs having a strong differentiation profile between the Africans and other continental groups, which are the AF-EA, AF-LA and AF-EU pairs (Fig. 5.6). Among the drugs that have extreme population differentiation profile between Africans and other continental populations are valdecoxib, atorvastatin, catechin, and docetaxel. Furthermore, it can also be observed that when comparing the African populations with East Asians, Europeans or Latin Americans, there are more drugs that are associated with tcdGenes enrichment between the African and European populations (Figure 5.6A, AF-EU column inside yellow box). Such population differentiation affect a number of frequently prescribed drugs such as imatinib, amiloride, morphine, fluvastatin, olanzapine, dexamethasone, as well as the four drugs mentioned above.

Furthermore, using these 34 drug names, it is possible to observe the frequency of drug class that is linked to these drugs. Drug class information is in accordance to that of the WHO drug Anatomical Therapeutic Chemical (ATC) classification system. Here, the frequency of which a drug class appears in the data is presented using a "word cloud" visualization. In this diagram, the bigger the word size the more frequent the drug class appears in the data. As can be observed in Figure 5.6B, in the first level of the ATC classification system, which categorizes drugs based on their anatomical main group, drugs that play role in the musculo-skeletal system are the most frequently observed in the cluster that shows population difference between Africans and other populations. In the therapeutic (Level 2) and chemical (Level 3) subgroup classification, the Antiinflammatory and

Antirheumatic Products drug category are more dominant than other drugs such as Antineoplastic Agents.

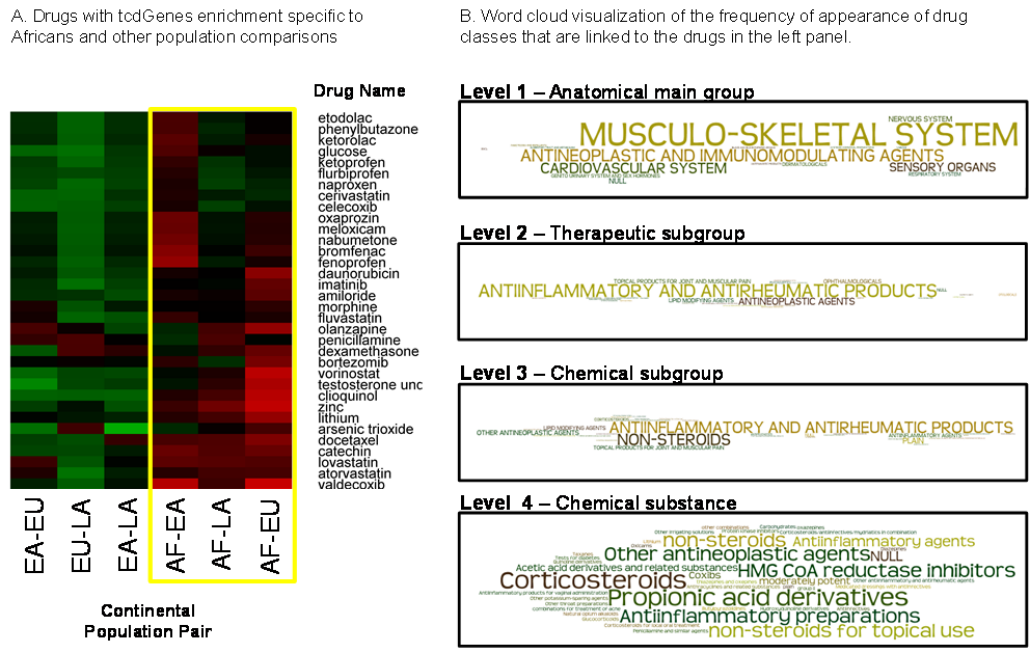


Figure 5.6 Drugs with a strong differentiation profile between the Africans and other continental groups. This cluster is consisted of 34 drugs that are differentiated between either the AF-EA, AF-LA or AF-EU continental population pair (A). The right panel (B) is a word cloud that is used to visualize the frequency of appearance of the drug classes that are associated with these 34 drugs. The word size is influenced by frequency of appearance of the drug class.

As shown in Figure 5.5 (inside green colour box), in addition to the drug-response gene sets with tcdGene enrichment in Africans, it is also observed that there is a second cluster, which consisted of 28 drug-response gene sets that are enriched by tcdGenes differentiated between East Asians and Europeans (Figure 5.7A). This drug cluster includes several benzodiazepine derivatives (oxazepam, flurazepam, temazepam, lorazepam, triazolam, clonazepam, alprazolam, and midazolam) and other drugs such as simvastatin, pravastatin, capecitabine, ethanol, pentobarbital, carbamazepine, and corticotropin. In the WHO ATC level 1 anatomical main group classification, the majority of drugs in this cluster are categorized under

Nervous System category, which is then followed by the Antineoplastic and Immunomodulating agents. Figure 5.7B also shows that the Psycholeptics as well as the Hypnotics and Sedatives drug class are shown to be prominent in the level 2 and level 3 classifications, respectively.

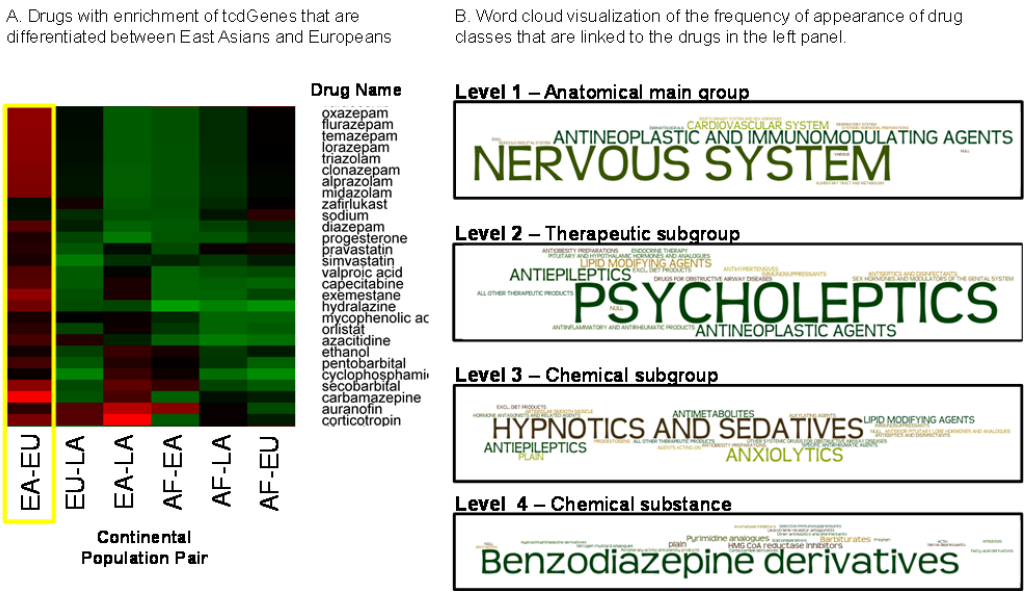


Figure 5.7 Drugs that are enriched by tcdGenes differentiated between East Asian and European. This cluster is consisted of 28 drugs (A), many of them are categorized under Nervous System drug category (B).

Outside of these two clusters, there are 72 drugs with low population differentiation profile (Figure 5.5, inside blue colour box). As a matter of fact, the majority of drugs involved in this analysis belong to this group (Figure 5.8A). In this cluster, with an exception of low differentiation between Africans and East Asians, no significant population differentiation profile is observed in any of the other continental populations. This cluster is dominated by drugs acting under the Antineoplastic and Immunomodulating Agents in the level 1 WHO ATC anatomical main group classification. The presence of Antineoplastic Agents therapeutic class further dominates the level 2 and level 3 drug classifications, respectively. Among the chemical substances that occur more frequently in this

cluster are the Protein kinase inhibitors, Proton pump inhibitors and Other antineoplastic agents (Fig. 5.8B).

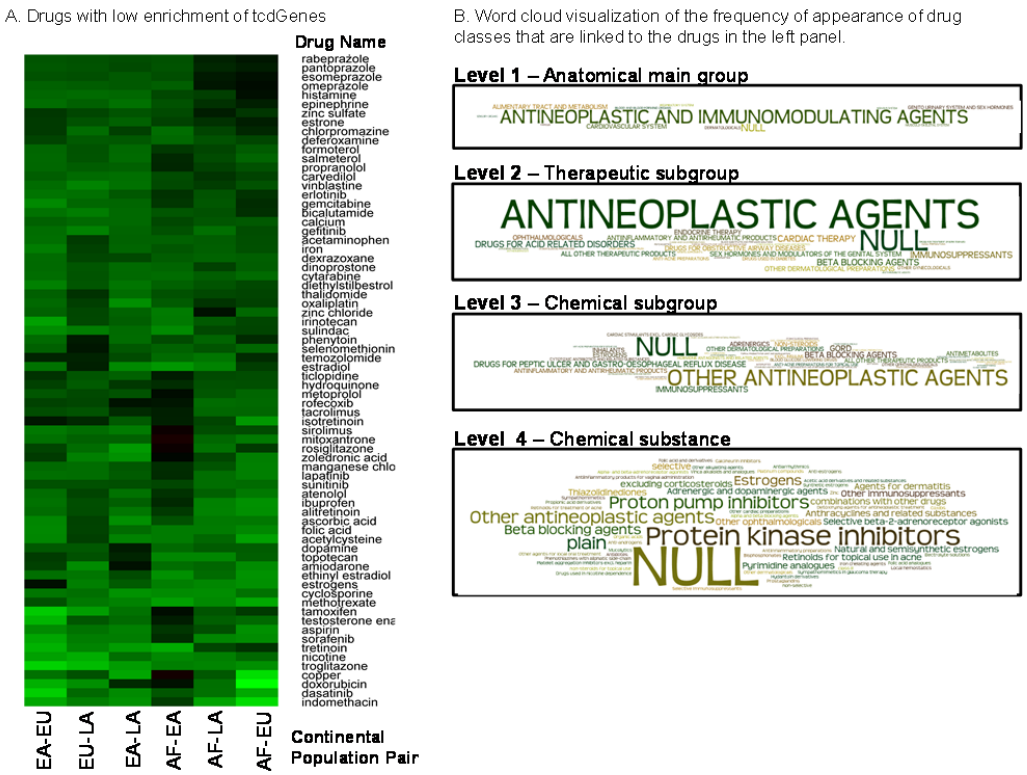


Figure 5.8 Drugs with relatively low enrichment of tcdGenes, which signifies weak population differentiation profile.

5.3.4 Genes and SNPs linked to drugs with high population differentiation profile

It is postulated that drugs with gene sets enriched by tcdGenes would be more likely to have population differences in response. Based on their enrichment z-scores, the drugs with the strongest tcdGenes enrichment across the six different continental population pairs are shown in Table 5.1. In descending order from the highest to lowest z-score, the list includes: Penicillamine (z-score = 6.35 – AF-LA

/ LWK-CLM; z-score = 3.98 – EA-LA / CHS-MXL), Valdecoxib (z-score = 6.12 – AF-EA / LWK-JPT), Secobarbital (z-score = 5.69 – EA-EU / JPT-FIN), Daunorubicin (z-score = 4.45 – AF-EU / YRI-FIN), and Docetaxel (z-score = 4.18 – EA-LA / CHS-MXL). The distribution of F_{ST} scores in the tcdGenes of these drugs are presented in Figure 5.9.

Table 5.1 Drugs with the strongest tcdGenes enrichment across six continental population pairs.

Drug Cluster	Drug Name	Z-score	Population Pair	Continental Population Pair
AF-others	Penicillamine	6.35	LWK-CLM	AF-LA
AF-others	Valdecoxib	6.12	LWK-JPT	AF-EA
EA-EU	Secobarbital	5.69	JPT-FIN	EA-EU
AF-others	Daunorubicin	4.45	YRI-FIN	AF-EU
AF-others	Docetaxel	4.18	CHS-MXL	EA-LA
AF-others	Penicillamine	3.98	TSI-CLM	EU-LA

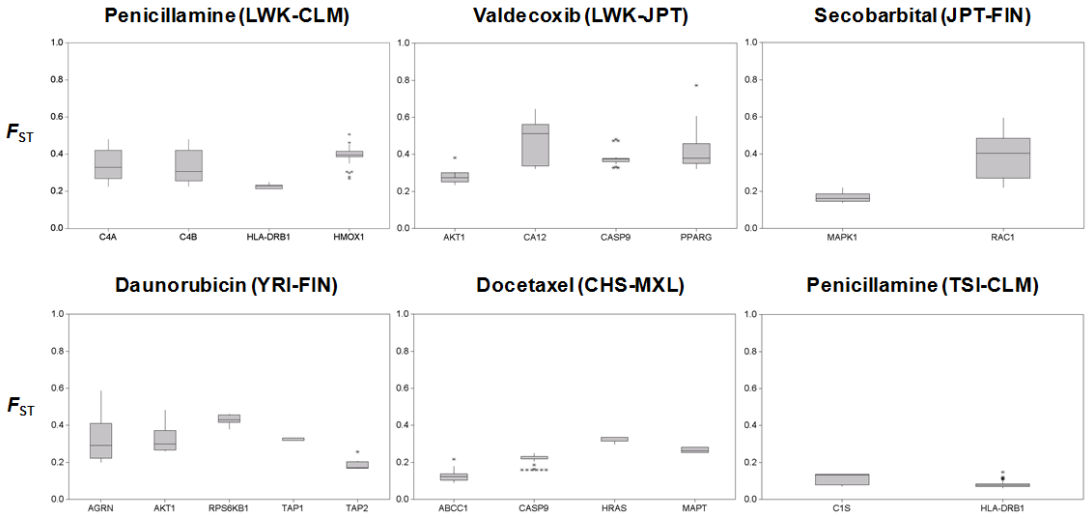


Figure 5.9 Distribution of F_{ST} scores associated with tcdSNPs residing in genes that are linked to drugs with the highest tcdGenes enrichment.

5.4 Discussion

One primary aim of this work was to accumulate and integrate population genomic knowledge that can serve to elucidate the genetic basis of population differentiation in drug response. Gene sets containing genes that have previously been associated with drug response were collected. These genes are involved in pharmacokinetics (PK) and/or pharmacodynamics (PD) pathway. Furthermore, the work in this thesis chapter was primarily conducted in parallel to the construction of a pharmacogenomics resource containing drug population differentiation information. In developing this resource, a total of 10,942 drug/compound data from four different databases were integrated. Out of this total number, 1,511 are approved for therapeutic usage by the FDA. This knowledge accumulation is essential in pharmacogenomics, whereby a drug potential population differentiation profile is constructed based on its drug-response gene information. Hence, a comprehensive resource that contains information of most if not all known drug-genes relationship is useful, particularly for elucidating the drug population genetic profile.

Traditionally pharmacogenomics studies have adopted a conservative approach. Most remain focused on studying specific candidate genes that not only have been commonly reported to play significant role in drug response, but also associated with variants or SNPs that have high population allele frequency differences [10-14]. These conventional pharmacogenomics studies had put much emphasis on studying absorption, distribution, metabolism, excretion (ADME) genes, which

are important to a drug pharmacokinetics. In this thesis however, there is an attempt to expand the genes collection to include other possible drug-response genes, particularly in a drug pharmacodynamics pathway, where little if any information is usually available for most drugs.

In this thesis chapter, with 4,188 genes in the database, cyclosporine is identified as drug with the greatest number of links to drug-response genes. This is then followed by tretinoin and estradiol, with 2,861 and 1,899 genes, respectively. The inclusion of data from the CTD, a non-conventional pharmacogenomics resource, expanded the collection of genes, with those that have been shown to be associated with toxic responses. Moreover, the massive compound information that is available in the CTD database would also expand the drug collection with compounds that are found in food supplements, which may interfere with a patient's response to therapy. It is hoped that the inclusion of these CTD genes could increase the possibility in finding more novel pharmacogenomics genes that are not only linked to a drug or compound-induced toxicity, but also identified to be population-differentiated. Using this novel approach, a significant number of drug response genes were collected, many of them are linked to the CTD, a toxicogenomics resource [5].

Several instances of non-drug/compound information that is not found in conventional pharmacogenomics resources are copper, selenium and arsenic. For example copper, which could be toxic due to its ability in generating reactive oxygen species [15], has 1,472 gene information links, the sixth highest of all compounds.

Subsequently, with this information available, drug-response genes that are population-differentiated could be identified. This was done by integrating the drug information database with those of tcdSNPs and tcdGenes data that were generated in chapter 4. In chapter 4, genes that carry top chromosome differentiated SNPs (tcdSNPs) are classified as tcdGenes. A drug pharmacogenomics profile, including information on enrichment of tcdGenes that could be linked to the drug was generated. Here, it is shown that there are 3,108 out of 14,166 drug-linked genes that were identified to carry tcdSNPs. This is equivalent to 70% of 4,355 tcdGenes in the human genome. One possible reason for this high proportion is because of the environmental response role that is attributed to the biological function that is associated with these tcdGenes. Population differentiation, as estimated by the presence of high F_{ST} tcdSNP, could arguably be advantageous in genes that are "exposed to environmental" pressures. This contributed to our body's defense mechanism against a wide range of xenobiotic substances.

Furthermore, because different geographic regions would probably exert a different range of pressures, human migration and genetic drift may leave a distinct population differentiation 'footprint' in these tcdGenes. The analysis of the top 20 drugs that are enriched by tcdGenes in the two most distant (CHS-YRI) and most similar populations (CEU-GBR) shows a different profile. Between the CHS and YRI, the proportion of tcdGenes is highest in genes that are linked to sincalide, azithromycin and nalidixic acid. On the other hand in the CEU-GBR pair, such proportion is highest in genes that are linked to phytonadione,

bacitracin and cosyntropin. Due to this population genetic differentiation, the drugs that are significantly enriched by these tcdGenes across diverse population pairings are believed to carry greater potential in having response differences.

Nonetheless, one challenge that was encountered in the analysis of 91 population pairs and 1,034 drugs/compounds was in finding a method that can delicately reveal a novel pattern of population differentiation based on genomic data. In this thesis chapter, I presented a two-step hierarchical cluster approach on how we could identify a potential drug response pattern using their population genetic differentiation profiles. The initial attempt at clustering consisted of 141 drug population differentiation profile. In this result, there exist significant presence of drugs that are linked to low population differentiation profiles, particularly between closely-related populations (populations originated from the same continental root). Moreover, because the cluster analysis and heat map generation were set to exempt null data, some drugs, more specifically those that have insufficient population genetics data for generating z-score using complete random sampling analysis in closely-related populations, were not included in this initial clustering. For this reason, it was decided to proceed with the second-step hierarchical clustering, in which z-score data from closely-related populations were excluded. The second tier clustering resulted in the hierarchical grouping of 173 drugs, 34 of which have strong differentiation profile between Africans and other populations. Interestingly, there is a high presence of drugs acting in the musculoskeletal system in the cluster that shows population genetic differentiation between Africans and other populations. In addition, 28 other

drugs were also noted to be in a cluster that is identified to have population differences between East Asians and Europeans. Note that the clustering methodology itself is not new; however, the application of such approach in grouping drugs based on their population differentiation profile is potentially novel. Using such technique, we can now identify the drugs that can have potential response differences between two different populations.

In the next step, using their z-scores, drugs with the strongest tcdGenes enrichment across six continental pairs were identified. Among these drugs, Penicillamine, Daunorubicin and Docetaxel are included in the WHO List of Essential Medicines [16]. Whilst both Daunorubicin and Docetaxel serve as anticancer agents that work by interfering with mitotic process and DNA replication, Penicillamine is a disease modifying anti-rheumatoid drug [17-19]. Here, it is identified that the gene set linked to Daunorubicin is enriched by tcdgenes that are differentiated between the YRI and FIN populations, which belong to the African and European group, respectively. In 2007, Huang et al reported that for Daunorubicin, there is significant difference (p-value < 0.05) in the effect of drug-induced toxicity between cell lines originated from African and European descent as measured by the drug's IC_{50} [20]. Both the African and European cell lines were from Yoruban and CEPH populations that participated in the HapMap project. These are indeed the same source of populations that were later be included in the 1000 Genomes Project, which were also used as the primary population genomic data source of this thesis.

Furthermore, based on the result presented in this chapter, Docetaxel is identified as the drug with the highest enrichment of tcdGenes that are differentiated between Southern Chinese (CHS) and Mexicans (MXL). Millward et al had previously reported a significant difference in the Docetaxel response rate between Asians and Caucasians [21]. Moreover, it was found that ethnicity could act as response predictor. In their analysis, it was observed that in contrast to 31% Caucasian patients who have had response to Docetaxel, there was significantly more Asian patients (65%) who responded to the anticancer drug ($P = 0.01$). When considering the relative genetic closeness between Caucasians and Latin Americans (as shown in Chapter 4), the finding that was reported by Millward et al is parallel to the z-score analysis involving Docetaxel.

The results that are presented in this thesis chapter could potentially be applied in a drug development pipeline. As discussed in the earlier section of my thesis, because a drug that is effective in one individual may not equally be effective in a different person, tailoring drug prescription will have an increasing clinical and economical benefits in the future. This will be especially more significant in certain drug cases that are associated with the occurrence of adverse drug reactions (ADRs). It is the long term objective that can be achieved by expanding the approach that is developed in this study. Variation in the way patient's response to drug, including ADRs susceptibility has been reported to be a significant contributing factor to the increasing cost of healthcare. It was reported that the cost of treating drug-related hospitalization in the US had reached $136 billion [22-24].

To the pharmaceutical industry, the occurrence of ADR in several patients or clinical trial volunteers could potentially be damaging to a drug portfolio, leading to a potential loss of revenue. For the individual patient, suffering an ADR is equivalent to adding an extra layer of burden rather than cure. By identifying the group of individuals who are more susceptible to an ADR or toxic reaction, using their population information as proxy, we could potentially prevent ADR before it happens.

Furthermore, although drug response differences are attributed to various factors, genetic polymorphisms are by far still one of the most promising factor that can be used to group patient populations. One reason is because of its constant and heritable nature, where a person genetic profile is constant for his or her life span. Nonetheless, despite the potentially effective utilization of SNPs that are associated with drug response differences in developing a more personalized medicine, cost is still a hurdle. For instance, if it must tailor new drug development to all variant types in different individuals, drug companies would incur an even greater price-tag into its already burgeoning cost in developing a drug. This translates to higher cost of medicine.

Nonetheless, the future of applying personalizing medicine still depends on the success of clinical or industrial prototyping of this approach, which harnesses an up-to-date knowledge of population genomic information. An attempt to address this challenge is presented in the next chapter, where I conclude this thesis with a presentation of the PharmaSNP resource. It is a prototype of one stop

pharmacogenomics portal that is aimed to utilize the results generated in this thesis and deliver its potential for clinical and/or industrial application.

5.5 References

1. Weir, B.S. and C.C. Cockerham, *Estimating F-Statistics for the Analysis of Population Structure.* Evolution, 1984. **38**(6): p. 1358-1370.
2. Hewett, M., et al., *PharmGKB: the Pharmacogenetics Knowledge Base.* Nucleic Acids Res, 2002. **30**(1): p. 163-5.
3. Wishart, D.S., et al., *DrugBank: a knowledgebase for drugs, drug actions and drug targets.* Nucleic acids research, 2008. **36**(Database issue): p. D901-6.
4. Gaulton, A., et al., *ChEMBL: a large-scale bioactivity database for drug discovery.* Nucleic acids research, 2012. **40**(Database issue): p. D1100-7.
5. Mattingly, C.J., et al., *The Comparative Toxicogenomics Database (CTD).* Environmental health perspectives, 2003. **111**(6): p. 793-5.
6. Administration, F.a.D., *Approved Drug Products with Therapeutic Equivalence Evaluations*, 2014.
7. *A map of human genome variation from population-scale sequencing.* Nature, 2010. **467**(7319): p. 1061-73.
8. Abecasis, G.R., et al., *An integrated map of genetic variation from 1,092 human genomes.* Nature, 2012. **491**(7422): p. 56-65.
9. Gregory R. Warnes, B.B., Lodewijk Bonebakker, Robert Gentleman, Wolfgang Huber Andy Liaw, Thomas Lumley, Martin Maechler, Arni Magnusson, Steffen Moeller, Marc Schwartz, Bill Venables, *gplots: Various R programming tools for plotting data*, gplots, Editor 2014, The R Project for Statistical Computing: http://cran.r-project.org/.
10. Fung, K.L. and M.M. Gottesman, *A synonymous polymorphism in a common MDR1 (ABCB1) haplotype shapes protein function.* Biochim Biophys Acta, 2009. **1794**(5): p. 860-71.
11. Pottier, N., et al., *Promoter polymorphisms in the beta-2 adrenergic receptor are associated with drug-induced gene expression changes and*

response in acute lymphoblastic leukemia. Clin Pharmacol Ther, 2010. **88**(6): p. 854-61.

12. Li, J., et al., *Global patterns of genetic diversity and signals of natural selection for human ADME genes.* Human molecular genetics, 2011. **20**(3): p. 528-40.
13. Aithal, G.P., et al., *Association of polymorphisms in the cytochrome P450 CYP2C9 with warfarin dose requirement and risk of bleeding complications.* Lancet, 1999. **353**(9154): p. 717-9.
14. Bachtiar, M. and C.L. Lee, *Genetics of Population Differences in Drug Response.* Current genetic medicine reports, 2013. **1**(3): p. 162-170.
15. Brewer, G.J., *Copper toxicity in the general population.* Clin Neurophysiol, 2010. **121**(4): p. 459-60.
16. Organization, W.H., *WHO Model List of Essential Medicines*, 2013.
17. Box, V.G., *The intercalation of DNA double helices with doxorubicin and nogalamycin.* J Mol Graph Model, 2007. **26**(1): p. 14-9.
18. Wadhwa, S. and R.J. Mumper, *D-penicillamine and other low molecular weight thiols: review of anticancer effects and related mechanisms.* Cancer Lett, 2013. **337**(1): p. 8-21.
19. Ward, J.R., *Role of disease-modifying antirheumatic drugs versus cytotoxic agents in the therapy of rheumatoid arthritis.* Am J Med, 1988. **85**(4A): p. 39-44.
20. Huang, R.S., et al., *Effect of population and gender on chemotherapeutic agent-induced cytotoxicity.* Mol Cancer Ther, 2007. **6**(1): p. 31-6.
21. Millward, M.J., et al., *Docetaxel and carboplatin is an active regimen in advanced non-small-cell lung cancer: a phase II study in Caucasian and Asian patients.* Ann Oncol, 2003. **14**(3): p. 449-54.
22. Becquemont, L., *Pharmacogenomics of adverse drug reactions: practical applications and perspectives.* Pharmacogenomics, 2009. **10**(6): p. 961-9.
23. Johnson, J.A. and J.L. Bootman, *Drug-related morbidity and mortality. A cost-of-illness model.* Arch Intern Med, 1995. **155**(18): p. 1949-56.

24. Bond, C.A. and C.L. Raehl, *Adverse drug reactions in United States hospitals.* Pharmacotherapy, 2006. **26**(5): p. 601-8.

Chapter 6. The PharmaSNP Resource of Integrative Pharmacogenomics

6.1 Introduction

In this thesis, one key focus is to accumulate new knowledge surrounding human population genomic differentiation, particularly those that are important in pharmacogenomics. It is also equally important to integrate this data with the existing breadh of knowledge, such as SNP, gene and drug information. However this alone, would not probably exert much translational significance, especially if the information is only stored for the benefits of a specific individual or group. Hence in this last chapter, I would like to present the final deliverable of this thesis, a pharmacogenomics online resource which publicizes the information that has been accumulated in this thesis.

In delivering translational impact, the results that have been presented in the previous chapter are packaged and presented in this 'PharmaSNP' resource, an online portal that aims to provide population genetics information of various drugs and compounds. The resource, which is currently in its beta format, integrated drug, gene and SNPs data that are relevant to elucidating the genomic basis behind population differences in drug response.

6.2 Developing the PharmaSNP resource

6.2.1 Data storage

The PharmaSNP resource includes various data that are relevant for profiling population differentiation pattern of various drugs/compounds. These data can be generally classified or grouped based on the type of information that is stored. Similar types of information are stored in a network of tables that are interlinked using the relational database approach. For 'big-data' computation such as in the calculation of F_{ST}, the Microsoft SQL Server (Microsoft Corporation, Washington, USA) relational database was utilized, where a database connection to the 'R' statistical programming environment was done. On the other hand, MySQL (Oracle Corporation, California, USA) was used for the purpose of publishing the data in PharmaSNP website. Table 6.1 summarizes these various data based on their respective information class.

Table 6.1 Data that are integrated in the PharmaSNP resource.

Information Class	Detail Content
Drug Population Differentiation	Z-score signifying enrichment of tcdGenes in drug-response gene set
Drug Classification	Drug classification based on WHO Anatomical Therapeutic Cemical (ATC) classfication system
Drug-response Genes	Drug-response genes data accumulated from PharmGKB, Chembl13, CTD, and Drug Bank
tcdGenes	Top chromosome differentiated genes (tcdGenes) information including gene name, NCBI geneID, chromosome, and population-differentiation status
tcdSNPs	Top chromosome differentiated SNPs (tcdSNPs) information including SNP rsID, mRNA-based location info, and population-pair F_{ST} score

6.2.2 Web development

The PharmaSNP website was developed based on a modular basis (Table 6.2). Whilst the first part consists of a data storage module as described in the above section, the second module involves a PHP-based query and data presentation applications, which connects and presents data on the worldwide web. Because PHP is a server-side scripting language, all queries that are performed in PharmaSNP are done within the NUS server. For the purpose of PharmaSNP, the Sriptcase (Netmake, Brazil) web development platform was used to develop the PHP applications. This includes the interactive database query functions, summary tables, graphical summaries and detailed information tables that are available in the PharmaSNP resource.

Table 6.2 Three major modules that form the PharmaSNP resource.

Module	Platform
Data Storage	Microsoft SQL Server, MySQL
Server-side Scripting (Web database query)	PHP
Web Content Management System (CMS)	Wordpress

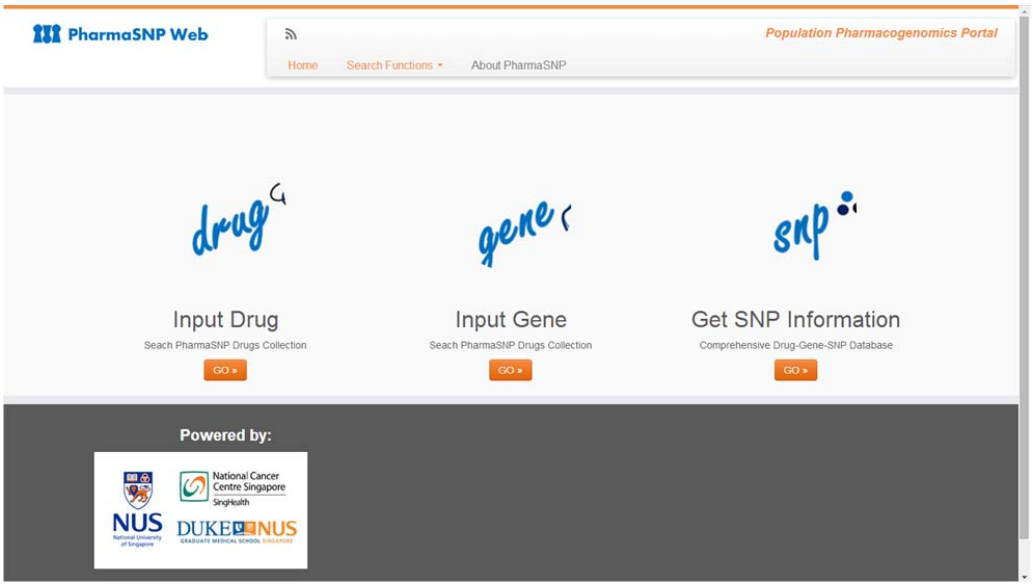


Figure 6.1 The PharmaSNP web resource allows users to query the PharmaSNP database from three initiation points. The website is accessible at **http://bit.ly/pharma-snp**

The last module was developed using Wordpress, a content management system (CMS) platform that is specially used to manage the PharmaSNP website content, allowing the creation of a user-friendly front end (www.wordpress.org). Wordpress is built based on MySQL and PHP, hence allowing straight-forward structural integration for deploying the PharmaSNP website. This last module, which is the most utilized CMS in the world, packages the PharmaSNP resource into a presentable website, allowing seamless navigation of the online resource. In addition, to serve mobile device users, this module can also adapt to various screen resolutions as it implements a responsive frame design.

6.3 Utilizing PharmaSNP

The prototype release of the PharmaSNP resource is currently accessible through the following link http://bit.ly/pharma-snp. When utilizing the beta version of PharmaSNP, one could initiate a database query from three different starting points (Fig. 6.1). The first one involved querying the PharmaSNP database with a drug name, whilst the second entry point would be useful for a database search using a gene name. The third initiation point allows PharmaSNP users to obtain detailed SNP information that is associated with SNPs linked to drug of interest.

6.3.1 Search Drug in PharmaSNP Collection

In the first initiation point, one could search the PharmaSNP database using a drug name and/or drug type (Fig. 6.2). In addition, information surrounding the population pair of interest could also be accounted in the database query by

highlighting the relevant population(s). Entering this information into the search form would allow one to obtain the population differentiation profile of a drug of interest based on the 1000 Genomes data. The following figures that are presented in this thesis were captured when a user initiated a database query using the drug name that contain the word ‘statin’. The same user was interested to explore the population differentiation profile of statin between the Luhya population in Webuye, Kenya (LWK) and Mexican Ancestry from Lost Angeles, USA (MEX).

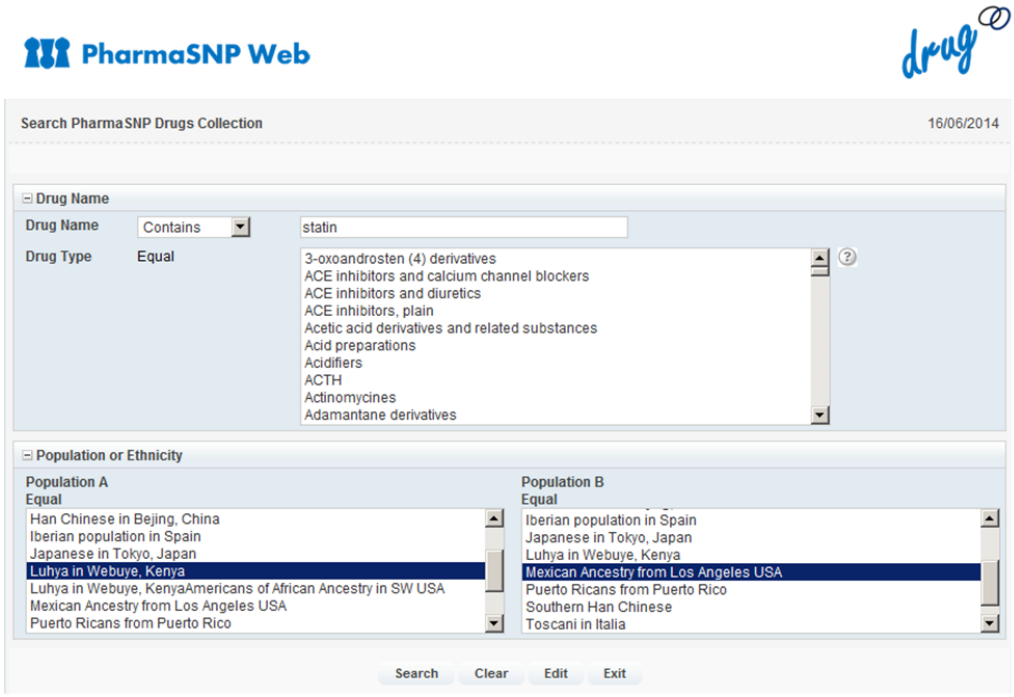


Figure 6.2 Search drug from PharmaSNP collection. Users are able to search the drug collection that is available at the PharmaSNP database.

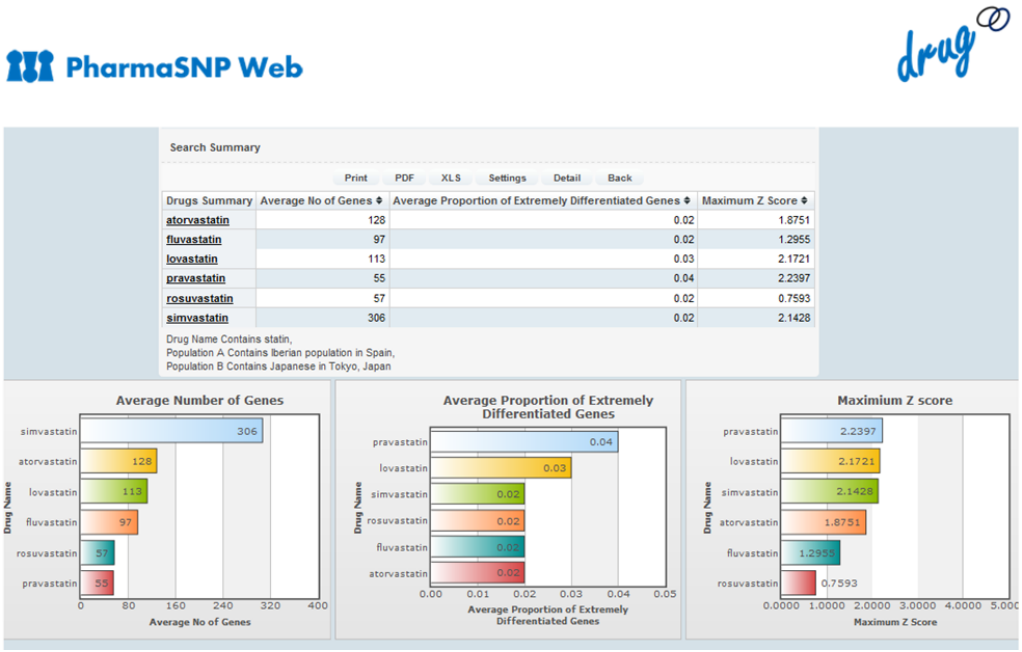


Figure 6.3 Summary page that appear once a user submitted a drug search in the database.

Following a search initiation, the PharmaSNP interface would then bring the user to a summary page, where all drug names containing the word 'statin' are presented (Figure 6.3). In this example, PharmaSNP summarizes population genetics data from six statin drugs: atorvastatin, fluvastatin, lovastatin, pravastatin, rosuvastatin, and simvastatin. In each drug, a summary is presented for the average number of genes that are associated with the drug, the average proportion of extremely population-differentiated genes and the maximum z-score signifying enrichment of such genes. In addition, an interactive barchart that summarizes the information is provided at the bottom of the summary table. When clicked, the individual bar in the chart will direct the user to more detailed population differentiation information of the drug of interest.

Figure 6.4 shows what happened after clicking 'atorvastatin' on the bar in the chart. Here, the users are taken into a more detailed table view of the population genetic differentiation profile of atorvastatin. In this case, PharmaSNP would present information that is relevant to atorvastatin including the total number of genes associated with atorvastatin and the proportion of extremely population-differentiated genes, which is alongside the population names. For instance, the first row of this search results describes that atorvastatin, as an HMG CoA reductase inhibitor, is linked with 128 genes in the PharmaSNP database. Out of 128 genes, 3% carry SNPs that are extremely differentiated between Luhya population (LWK, which belongs to the African continent group) and Mexican (MEX, which belongs to the Latin American continent group). Moreover, because

the z-score of 2.5993 is above the significant threshold of 1.96, these extremely population-differentiated genes are considered to be significantly enriched.

Both the summary and detailed table results that are generated by the PharmaSNP database search could be exported in PDF, table, XML or other text file formats. The user is also able to customize the display of the search results, in accordance to a more specific requirement. This includes advance results sorting based on a chosen field, selecting the columns-to-display and performing a quick search from the result that has been generated.

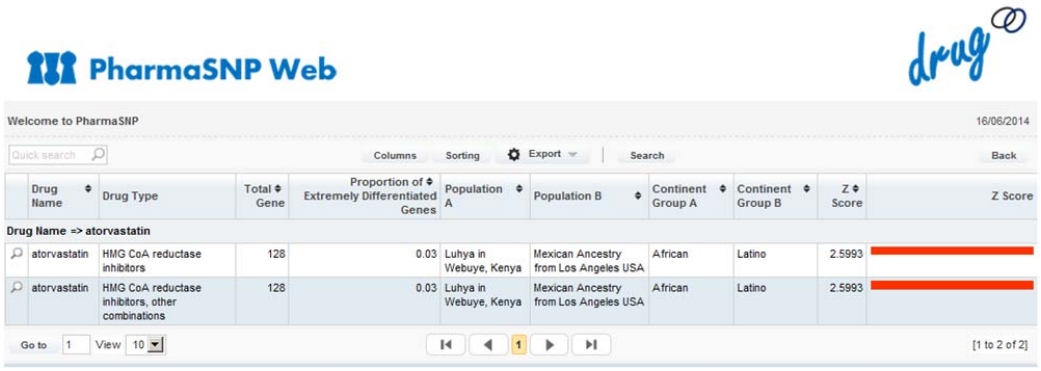


Figure 6.4 Detailed table view of the drugs' population genetic differentiation profile.

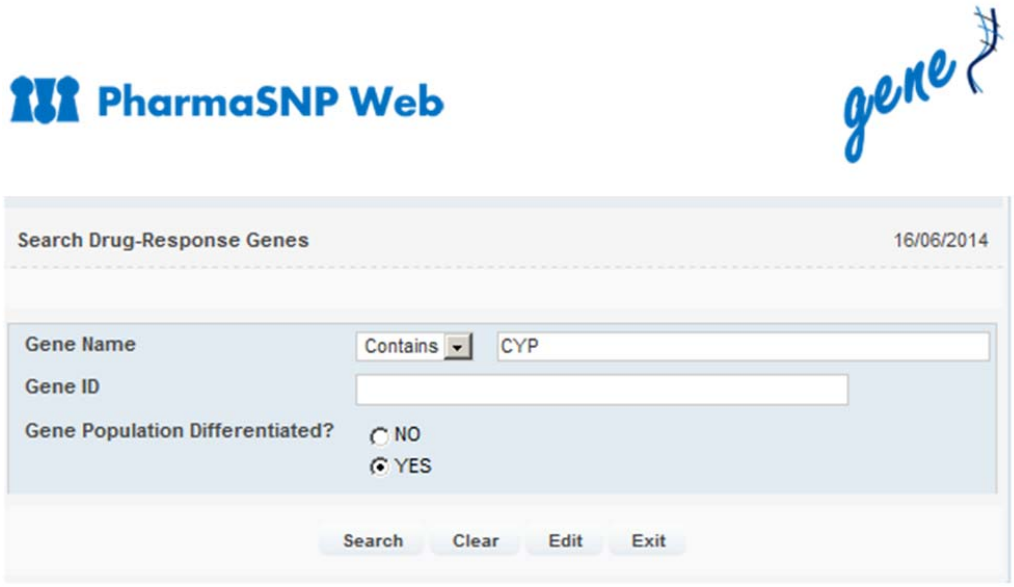


Figure 6.5 The interface that allows one to search the PharmaSNP gene collection.

6.3.2 Search Gene in PharmaSNP Collection

While the first initiation point allows the user to search using a drug name, the second initiation point, will receive a gene name or an NCBI gene ID as input to query the PharmaSNP database. The user could also opt to choose to present a result only if the gene(s) of interest is extremely population-differentiated (Fig. 6.5).

In this particular example, using the PharmaSNP gene query interface, the user has attempted to search for pharmacogenomics information that is relevant to any gene containing the keyword 'CYP' (Fig. 6.5). The user also chose the option to show only the genes that are population-differentiated. Figure 6.6 shows what the PharmaSNP resource would return when this search was submitted. The table summarizes all the CYP genes that are extremely differentiated, in addition to all the drug names that are linked to these tcdGenes. For instance in the first row, the gene CYP2D6, with an NCBI gene ID of 1565 is extremely differentiated between CHB (Han Chinese in Beijing, China) and the CLM (Columbian in Medellin, Columbia). Moreover this gene, is linked to the drug bromfenac, which is an anti-inflammatory agents, non steroids type; as well as to buspirone, an Azaspirodecanedione derivatives (row 2); estrone, a natural and semisynthetic estrogen (row 4); and more.

PharmaSNP Web

Search Drug-Response Genes 16/06/2014

Quick search | Columns | Sorting | Export | Back

Gene Name	Gene ID	Drug Name	Drug Type	Gene Population Differentiated?	Pop A	Pop B
CYP2D6	1565	bromfenac	Antiinflammatory agents, non-steroids	YES	CHB	CLM
CYP2D6	1565	buspirone	Azaspirodecanedione derivatives	YES	CHS	TSI
CYP4F12	66002	ketoconazole	Imidazole and triazole derivatives	YES	CHS	CLM
CYP2D6	1565	estrone	Natural and semisynthetic estrogens, plain	YES	CHB	TSI
CYP4F12	66002	ketoconazole	Imidazole derivatives	YES	JPT	YRI
CYP4F3	4051	phenol	Antiseptics	YES	CEU	JPT
CYP2D6	1565	celecoxib	Other antineoplastic agents	YES	CHB	TSI
CYP4F12	66002	ketoconazole	Imidazole and triazole derivatives	YES	CHB	PUR
CYP2D6	1565	risperidone	Other antipsychotics	YES	CHS	TSI
CYP2D6	1565	letrozole	Aromatase inhibitors	YES	CHB	FIN
CYP2D6	1565	tegaserod	Drugs acting on serotonin receptors	YES	CHS	IBS
CYP4F12	66002	ketoconazole	Imidazole and triazole derivatives	YES	ASW	JPT
CYP4F12	66002	bortezomib	Other antineoplastic agents	YES	JPT	PUR
CYP2D6	1565	ibuprofen	Other cardiac preparations	YES	CHB	TSI
CYP2D6	1565	melarsoprol	Arsenic compounds	YES	CHB	IBS
CYP2D6	1565	atomoxetine	Centrally acting sympathomimetics	YES	CHS	IBS
CYP2D6	1565	fenoprofen	Propionic acid derivatives	YES	CHS	TSI
CYP4F12	66002	ketoconazole	Imidazole and triazole derivatives	YES	FIN	JPT
CYP1A2	1544	ciprofloxacin	Fluoroquinolones	YES	CEU	TSI
CYP2D6	1565	naproxen	Antiinflammatory products for vaginal administration	YES	CEU	CHS

Go to 1 View 20 1 2 3 4 5 [1 to 20 of 21239]

Figure 6.6 The search result following submission of the keyword 'CYP'.

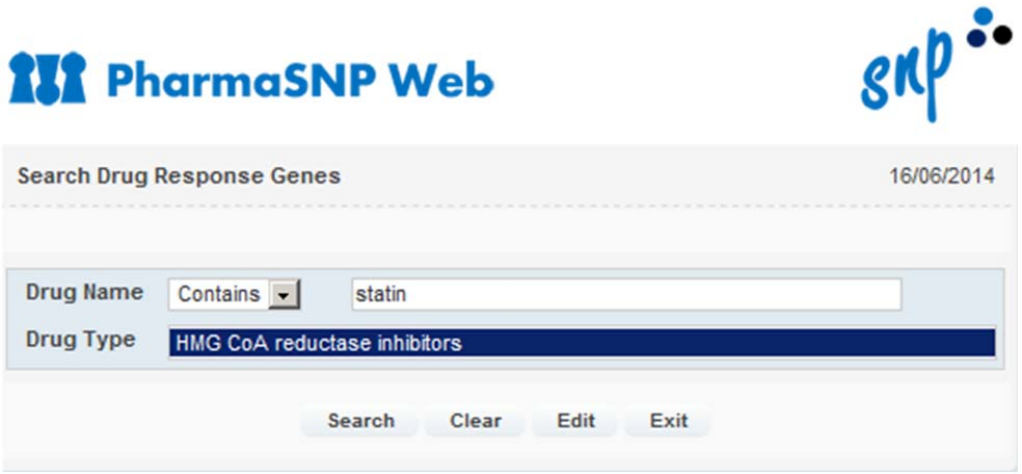


Figure 6.7 To search for the SNPs that are linked to a drug of interest, a user may enter the drug name.

6.3.3 Obtaining Pharmacogenomics SNP Details

The third interface of the PharmaSNP resource integrates the drugs, genes and SNPs data in the database, which allow users to obtain a comprehensive list of genes and detailed SNPs information that could be linked to a drug of interest. In this illustration, the same user attempted to search pharmacogenomics information that are relevant to the keyword of interest, 'statin', which is an HMG CoA reductase inhibitors (Fig. 6.7). Note that due to the massive data integration that is involved, the interface currently serves as a beta version, in which only a subset of the results is shown by the query engine.

When the user entered statin as the keyword, PharmaSNP beta would present a sample result that encompasses the drug named lovastatin. The positive sign next to the drug name would allow the user to expand the results in which a sample population genetics pair differentiation analysis between Luhyans (LWK) and Mexican (MEX) is displayed (Fig. 6.8). In this example, the proportion of extremely population-differentiated genes between these two populations is 0.06, with a significant enrichment z-score of 5.1832. The user could then further expand the table to display the extremely population-differentiated genes that are involved in this particular drug, in addition to the population pair information. Lastly, clicking the expansion side in the gene would then bring the user to a table that displays detailed information of SNPs that reside in the gene of interest (Fig. 6.9).

When the same user expanded the information on the PPARG genes, the SNPs within the gene are displayed in addition to their positions in accordance to the RNA splicing variants. In the first row instance, a PPARG SNP with reference SNP identifier (rs#) of rs709154, is seen to have an F_{ST} score of 0.3065 based on its allele frequency differentiation between the JPT (Japanese) and YRI (Yoruban) populations. Within the PPARG splice variant identified as NM_005037.5, the same SNP could be localized in the Intron 5 region, 1,369 away from the 3' intron-exon junction. SNP rs709154 is associated with an A to T nucleotide substitution and due to its location, is not considered as a promoter SNP. Using this information, a position-centric nomenclature of rs709154 could then be derived, which is I-1369A/T.

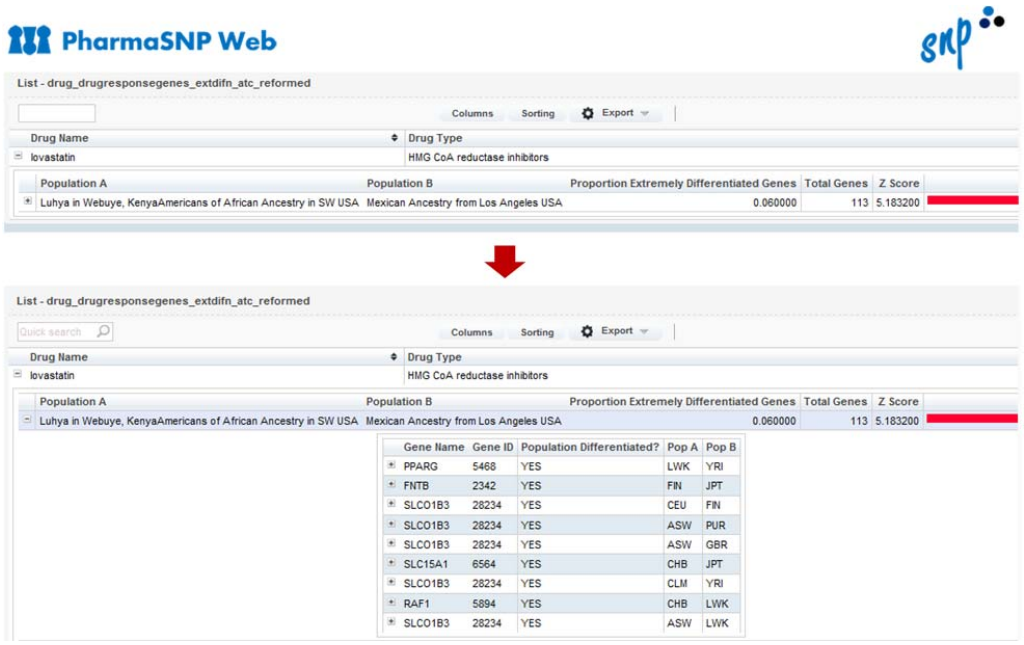


Figure 6.8 The summary page presenting the search results. The user could expand the results to display the extremely population-differentiated genes that are involved in the drug of interest.

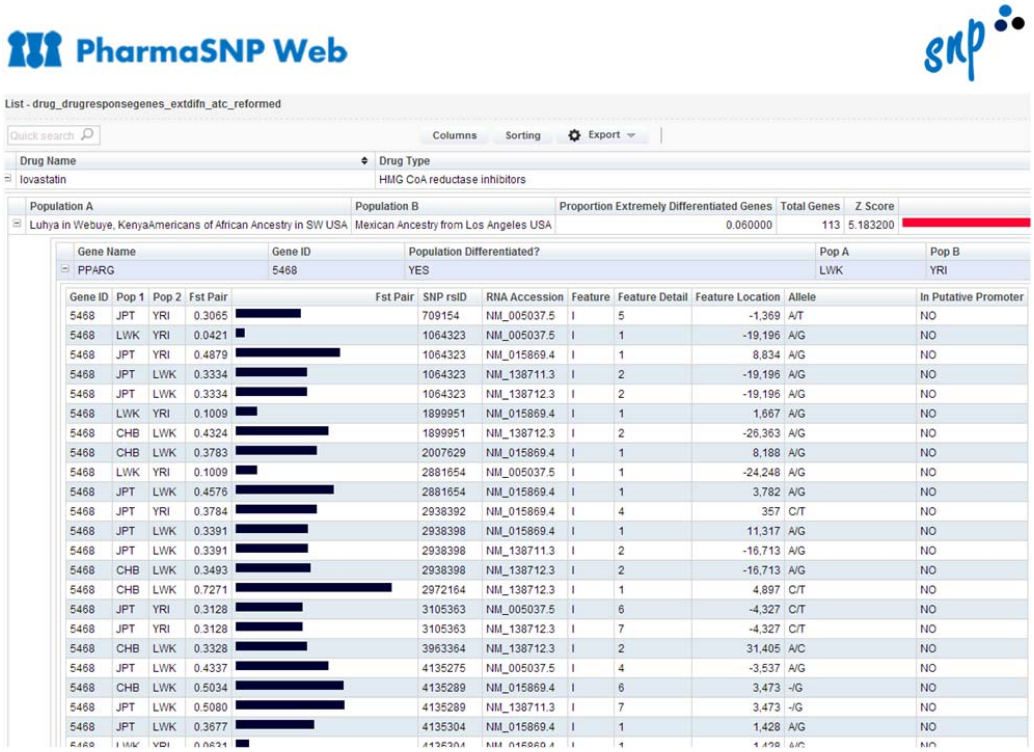


Figure 6.9 To view the SNPs in the gene of interest, another expansion step could be performed.

Significance and Future Direction

Genetic Diversity, a Great Strength

Genetic diversity is a great strength. Once deciphered, the information stored in our genetic sequence could serve as predictor of population differences in phenotype. As illustrated in this thesis, genetic diversity in the form of SNPs in drug-response genes, could explain and have the potential to predict response profile. By identifying the PharmaSNPs that are linked to a set of drug-response genes, a drug population differentiation profile could be constructed (Chapter 6).

The Impact

For the pharmaceutical industry, ADR occurrence can potentially damage a drug portfolio, leading to a potential loss of revenue. For the patient, an ADR occurrence can add extra layer of burden rather than cure. The results presented in this thesis can hopefully attack these problems from a population genomic perspective. By utilizing the PharmaSNP resource, a drug population genetic differentiation profile can be obtained. This allows the early implementation of ADR prevention strategy, by identifying the group of drugs or individuals that are more susceptible to drug toxicity. The aim is to improve treatment quality and concurrently reduce the cost of medicine.

Moreover, in a drug development pipeline, this novel approach may prove useful in identifying the population that could potentially be more susceptible to ADR. I propose two future approaches that can potentially be employed before

conducting a new drug trial. The first relies on structural similarity with an existing drug, whilst the second takes into account any existing or new drug-response genes information that could be linked to the drug in trial.

In the first approach, a new drug can be linked with a set of drug-response genes of an existing drug that is structurally similar. One would be able to obtain the population genetic differentiation profile of a new drug using information from its existing counterpart. The population genomic differentiation data in PharmaSNP, in addition to the drug-response genes information can be used to obtain a drug population genetic differentiation profile. The second approach will hypothetically yield a more accurate prediction as drug-response genes information is derived from *in vitro* experiment involving the new drug. For instance, functional toxicogenomic studies are to be initially conducted, which may yield new genes information. The PharmaSNP resource could then provide the SNPs that can be linked to these genes. Using this information, drug population genetic differentiation profile can then be obtained.

Therefore, the information that is derived from PharmaSNP could arguably be used before considering to test a drug in one or more populations of interest. It may also reduce the potential occurrence of ADR during trial, by identifying the population group that is potentially more susceptible to ADR. This may also increase the probability of success in developing potentially block buster drugs.

These two proposed approaches can hopefully contribute to improving the drug development pipeline, which can reduce the cost that is associated with multi-

populations drug trial. The next step following the completion of this thesis is to continue developing the PharmaSNP prototype, in addition to optimizing new algorithm that can better provide a drug population differentiation profile using this population genomics information.

Nonetheless, whilst it is tempting to conduct a census-scale DNA sequencing in profiling all individuals in a country, ethical and economic reasons will still provide a continuous hindrance. It is for this reason that in this field, there is a greater focus on the study of individual representatives that are originated from different population backgrounds [1-3]. The availability of population genetic data could eventually serve as a proxy to the individuals that belong to that population. And this can be used to infer the inter-individual differences that are seen across diverse human populations, an approach that was adopted by the HapMap [1], Environment Genome Project [3], and more recently, the 1000 Genomes Project [2]. One primary assumption is that within the same population, individuals are likely to carry more similar gene variants compared to those who are originated from a different population.

This study is also associated with limitations, such as in the potential inheritance of error from various external data sources. Addressing these concerns would require involvement of more data speacialists who would be able to thoroughly assess each data before their interrogation in the Pharma-SNPs database.

The Art in Medicine

I would like to end this thesis with a little bit of history. The word 'medicine' is derived from '*medicus*', a latin word that means 'physicians' [4]. For centuries, the field has touched upon one of the most basic human necessities: to have a good health, hence good life. The word '*Ars medicina*' was then used, which means the 'art of healing'. And more recently, parallel to the accumulation of new facts and technology in diagnosing and treating patients, our perception of the word 'medicine' has been influenced by one important component: science. In fact, the current English Oxford Dictionaries recorded medicine as "the science or practice of the diagnosis, treatment, and prevention of disease" [5]. I find it interesting that the word 'art' is not included anymore in today's decscription. It signifies a significant degree of evolution in the field to a more knowledge-based method of treating diseases.

The art component however, shall not be entirely diminished. I believe that as medicine progresses, our ability to understand and apply human art can help to effectively apply and deliver new scientific discoveries. Therefore, medicine did and plausibly still does require a great deal of 'art' in its application. In fact, it is still arguably a combination between science and art [6]. Indeed, this is what provides gravity to my thesis. Here, I presented the scientific accumulation and integration of novel knowledge that in the long run, can hopefully be utilized in developing new utility for improving therapy.

As we move forward, pharmacogenomics will and can propel new medical innovations; especially in the development of personalized medicine. All in all,

the intent is to benefit the consumers, be it the scientists who are investigating, the doctors who are applying this knowledge, or the patients who are receiving the products. In the last chapter of my thesis, I attempted to package the study results into a prototype online resource. It is an art in progress. I believe that the results generated from the work presented in this thesis can humbly expand the medicine arena to a wide range of new possibilities.

References

1. *A haplotype map of the human genome.* Nature, 2005. **437**(7063): p. 1299-320.
2. Abecasis, G.R., et al., *An integrated map of genetic variation from 1,092 human genomes.* Nature, 2012. **491**(7422): p. 56-65.
3. Livingston, R.J., et al., *Pattern of sequence variation across 213 environmental response genes.* Genome Res, 2004. **14**(10A): p. 1821-31.
4. Grandy, J.K., *Medicine, History of. Encyclopedia of Time: Science, Philosophy, Theology, & Culture. SAGE Publications, Inc*, Thousand Oaks, CA: SAGE Publications, Inc. 843-846.
5. *Oxford Dictionaries*. [cited 2014; Available from: http://www.oxforddictionaries.com/.
6. Panda, S.C., *Medicine: science or art?* Mens sana monographs, 2006. **4**(1): p. 127-38.

Appendices

Appendix 1. SNP centric summary of results from association studies assessing the link between ABCB1 coding region SNPs and drug pharmacokinetics or response

No.	SNP (Amino Acid Substitution)	Drug Class/Indication	Drug(s) Association	[Ref] Population (cohort size)*	[Ref] Minor Allele Association Notes^^
1	E3/61A>G (N21D)	Immunosuppressants	Yes	[1] *In vitro* - cell lines.	[1] Modulate CsA intracellular accumulation effect.
			No	[2] White (73).	[2] No influence on CsA pharmacokinetics.
		Opiate analgesic	Yes	[3] NS - Switzerland (276).	[3] Influence methadone plasma level.
		Others/Mixed Substrates	Yes	[1] *In vitro*- cell lines.	[1] Increase of BODIPYL-FL-paclitaxel intracellular accumulation
			No	[4] *In vitro* - cell lines.	[4] N21D, F103L, S400N, A893S, and A998T on the PK of a range of cytotoxic derivatives.
		No report for the following drug classes/indications		Antiepileptic Drugs (AED), Antibiotics, Antidepressants, Anti Cancer Agents, Cardiac Glycoside, Antiretroviral Therapy (HIV), Statins.	
2	E5/266C>T (M89T)	Anti Cancer Agents	Yes	[5] *In Vitro* - cell lines.	[5] Increased resistance to daunorubicin, doxorubicin, valinomycin, or actinomycin D.

		No report for the following drug classes/indications	Antiepileptic Drugs, Antibiotics, Antidepressants, Cardiac Glycoside, Antiretroviral Therapy (HIV), Immunosuppressants, Opiate analgesic, Statins.		
3	**#ns7 E8/554G>T (G185V)**	Anti Cancer Agents	Yes	[6] *In vitro* - multidrug-resistant cell lines.	[6] G185V confer changes in vinblastine and colchicine specificity
		No report for the following drug classes/indications	Antiepileptic Drugs, Antibiotics, Antidepressants, Cardiac Glycoside, Antiretroviral Therapy (HIV), Immunosuppressants, Opiate analgesic, Statins.		
4	**E12/1199G>A (S400N)**	Anti Cancer Agents	Yes	[7] *In vitro*, cell lines; [8] NS - Sweden (51); [9] Taiwanese (59) [10] *In vitro* - cell lines; [11] NS - USA (85); [12] *In vitro* cell lines; [7] NS - USA.	[7] 1199G>T, reduced resistance, variety of drugs; [8] 1199G>T/A = paclitaxel response; [9] 2677GG and 3435CC = docetaxel side effects, no effect from -41A>G, -145C>G, 1236C>T; [10] 1199A = increased resistance; [11] IVS9 -44A>G = SN-38 PK ^^ [12] 2005C>T = lower resistance to paclitaxel and etoposide; [7] T allele is associated with reduced resistance to vinblastine, vincristine, paclitaxel, and doxorubicin in Leukemia patients. Patients carrying the A allele exhibited increased resistance.
			No	[13] *In vitro*, cell lines.	[13] Does not affect resistance to doxorubicin
		Antiretroviral Therapy (HIV)	Yes	[14] *In vitro* - cell lines.	[14] Increased cellular uptake and permeability of HIV protease inhibitors
		Immunosuppressants	Yes	[15] Finnish Caucasian (103), East African (1).	[15] 1199G>T, 1236C>T, 2677G>T/A, 3435C>T haplotypes = CsA PK
		Others/Mixed Substrates	No	[4] *In vitro* - cell lines.	[4] N21D, F103L, S400N, A893S, and A998T on the PK of a range of cytotoxic derivatives.
		No report for the following drug classes/indications	Antiepileptic Drugs, Antibiotics, Antidepressants, Cardiac Glycoside, Opiate analgesic, Statins.		

5	**#s6 E13/1236C>T (G412G)**	Antibiotics	Yes	[16] Chinese (18).	[16] 1236CC = lower cloxacillin Cmax, higher oral clearnace, lower urinary excretion.
		Anti Cancer Agents	Yes	[17] African American and Caucasian (NA); [18] Caucasian (63), Asian (2); [19] NS - Germany (112).	[17] MTX toxicity; [18] Irinotecan dose; [19] 1236CC = temozolamide response -assesed with other SNPs.
		Immunosuppressants	Yes	[20] Han Chinese (103).	[20] Dose adjusted conc. following renal transplant.
		No report for the following drug classes/indications		Antiepileptic Drugs, Antidepressants, Cardiac Glycoside, Antiretroviral Therapy (HIV), Opiate analgesic, Statins.	
6	**E17/1985T>C (L662R)**	Anti Cancer Agents	Yes	[5] *In vitro* - cell lines.	[5] Increased resistance to daunorubicin, doxorubicin, valinomycin, or actinomycin D.
		No report for the following drug classes/indications		Antiepileptic Drugs, Antibiotics, Antidepressants, Cardiac Glycoside, Antiretroviral Therapy (HIV), Immunosuppressants, Opiate analgesic, Statins.	
7	**E17/2005C>T (R669C)**	Anti Cancer Agents	Yes	[5] *In vitro* - cell lines; [12] In Vitro - cell lines.	[5] Increased resistance to daunorubicin, doxorubicin, valinomycin, or actinomycin D; [12] Decreased paclitaxel and etoposide resistance.
		No report for the following drug classes/indications		Antiepileptic Drugs, Antibiotics, Antidepressants, Cardiac Glycoside, Antiretroviral Therapy (HIV), Immunosuppressants, Opiate analgesic, Statins.	
8	**#ns22 E22/2677G>T/A (S893A/T)**	Antidepressants	Yes	[21] NS - Croatia (240).	[21] Olanzapine efficacy.
		Anti Cancer Agents	Yes	[22] *In vitro* cell lines; [23] NS - Australia (309).	[22] Increased vincristine transport with Ser893 and Thr893; [23] Taxane response.

			No	[24] Central European Caucasian (213).	[24] MTX efficacy.
		Cardiac Glycoside	Yes	[25] Caucasian (77).	[25] 2677T/A = digoxin stimulated saliva/serum ratio - assessed with other SNPs.
		Immunosuppressants	Yes	[26] NS - Japan (17); [27] NS - India (155); [28] Han Chinese (115).	[26] TRL induced neurotoxicity; [27] CsA dosage; [28] 2677T/A = TRL induced neurotoxicity;
		Statins	Yes	[29] Caucasian (1507).	[29] 2677T/A = Pravastatin Efficacy - assessed other SNPs
		Others/Mixed Substrates	Yes	[30] NS - Germany (55); [31] Caucasian (37), African American (23).	[30] TT/TA = elevated serum concentration-time curve values of talinolol; [31] Efflux of digoxin, Fexofenadine levels.
		No report for the following drug classes/indications		Antiepileptic Drugs, Antibiotics, Antiretroviral Therapy (HIV), Opiate analgesic.	
9	**E26/3151C>G (P1051A)**	Others/Mixed Substrates	Yes	[5] *In vitro* - cell lines.	[5] Influence resistance to valinomycin when in diplotype with E22/2677G>T/A (S893A/T).
		No report for the following drug classes/indications		Antiepileptic Drugs, Antibiotics, Antidepressants, Anti Cancer Agents, Cardiac Glycoside, Antiretroviral Therapy (HIV), Immunosuppressants, Opiate analgesic, Statins	
10	**E27/3322T>C (W1108R)**	Anti Cancer Agents	No	[5] *In vitro* - cell lines.	[5] Decrease resistance to daunorubicin, doxorubicin, valinomycin, or actinomycin D.
		No report for the following drug classes/indications		Antiepileptic Drugs, Antibiotics, Antidepressants, Cardiac Glycoside, Antiretroviral Therapy (HIV), Immunosuppressants, Opiate analgesic, Statins	
11	**E27/3421T>A (S1141T)**	Anti Cancer Agents	Yes	[5] *In vitro* - cell lines.	[5] Increased resistance to daunorubicin, doxorubicin, valinomycin, or actinomycin D.

		Immunosuppressants	Yes	[1] *In vitro* - cell lines.	[1] Reduced sensitivity to CsA inhibition of BODIPYL-FL-paclitaxel transport.
		No report for the following drug classes/indications		Antiepileptic Drugs, Antibiotics, Antidepressants, Cardiac Glycoside, Antiretroviral Therapy (HIV), Opiate analgesic, Statins	
12	**#s20 E27/3435C>T (I1145I)**	Antiepileptic Drugs	Yes	[32] NS - Turkey (96); [33] NS - France (2208).	[32] Pharmacokinetics of phenytoin; [33] TT = side effect from clopidogrel (cardiovascular event at 1yr).^^
			No	[34] NS - Turkey (189); [35] Meta analysis (3371); [36] NS-Turkey (104).	[34] Drug resistance; [35] Drug efficacy; [36] PK of valporic acid.
		Antibiotics	No	[37] Mixed - Caucasian (12), Asian (5), African American (1), NS (1); [38] Korean (210).	[37] Dicloxacillin pharmacokinetics. [38] Pantoprazole, amoxycillin and clarithromycin efficacy.
		Antidepressants	Yes	[39] "Predominantly Caucasian" (160); [40] NS - Sweden (116).	[39] 3435TT risk factor for nortriptyline-induced postural hypotension; [40] Olanzapine efficacy.
		Anti Cancer Agents	Yes	[41] Caucasian (191), Asian (5) African (2) Hindustani (3), Surinamese (3) and Israeli (1); [42] Caucasian (73); [43] NS - France (23); [44] NS - USA (324) [45] Northern Irish (184); [46] Caucasian (334).	[41] T allele assoc with adverse side effect after 6 months. No assoc with drug efficiacy. ^^ [42] TT = risk of side effects encephalopathy; [43] Irinotecan PK; [44] TT = adverse side effect; [45] CT = higher survival on vincristine, doxorubicin and dexamethasone; [46] lower response to epirubicin and doxorubicin.
			No	[47] Chinese (28), Malay (3), Indian (1); [48] Japanese (145), Japanese (197), NS -	[47] Docetaxel PK; [48] PK or adverse effects of a range of drugs. [49] Imatinib response.^^

			USA (184); [49] NS - Korea (229).	
		Cardiac Glycoside	Yes	[50] Caucasian (21); [51] German (461); [52] Japanese (15) [53] Japanese (11); [54] NS - France (12); [55] NS - France (32).	[50] digoxin plasma levels; [51] TT = increased intestinal uptake; [52] lower serum concentration; [53] suppression of duodenal absorption of drug; [54] TT = higher AUC @ 4 and 24hrs. [55] 3435T = volume of distribution, Higher conc. for T variants - assessed with other SNPs.
			No	[56] NS - France (12); [57] Meta analysis (183); [58] Caucasian (77).	[56, 57] Drug PK; [58] Serum conc.
		Antiretroviral Therapy (HIV)	Yes	[59] Caucasian (123); [60] NS - USA (71); [61] NS - Spain (74); [62] NS - France (32); [63] African (22), Caucasian (46) Other (8); [64] African (156).	[59] nelfinavir/efavirenz efficacy and plasma drug concentrations; [60] higher nelfinavir plasma levels; [61] Plasma levels of atazanavir and risk of hyperbilirubinemia; [62] absorption constant of indinavir; [63] Drug dependent changes in mean HDL-c levels; [64] T allele = protective effect from nevirapine induced hepatotoxicity.
			No	[65] NS - Finland (17).	[65] Saquinavir PK.
		Immunosuppressants	Yes	[66] Black (22), White (120), Middle Eastern (12), South Asian (26); [67] NS - USA (10); [68] NS - France (44); [69] NS - China (50); [70] NS - Iran (88); [71] Caucasian (75); [72] NS - Turkey (92); [73] Chinese (66); [74] NS - Italy (50); [75] Chinese (66), Malay (13), Indian (3); [76]	[66] CC genotype = minor effect on blood conc. of TRL; [67 higher CsA oral clearance; [68] CsA C:D ratio and dose requirement; [69] TRL dose requirement and dose-adjusted trough levels; [70] CsA PK and dose requirements; [71] CsA conc; [72] CC = lower dose-adjusted trough TRL conc; [73] TRL dosage requirements; [74] TT genotype = CsA induced gingival overgrowth; [75] CC genotype = higher TRL efflux/lower C:D ratio - due to higher protein expression^^; [76] TT = higher TRL C:D ratio @ 1-3days.

				Caucasian (42).	
			No	[77] Caucasian (142); [78] NS - USA (14); [79] NS - Japan (69); [80] Meta-analysis (1036); [81] Chinese - (155); [82] NS - Spain (53).	[77] CsA efficacy; [78] CsA PK; [79] TRL C:D ratio^^; [80] CsA PK; [81] sirolimus PK; [82] TRL dosage requirements.
		Opiate analgesic	Yes	[83] Caucasian - Italy (145).	[83] Morphine efficacy.^^
		Statins	Yes	[84] NS - Netherlands (1255).	[84] 3435T interacts with CYP3A4*1B and subsequently simvastatin and atorvastatin PK - assessed with other SNPs. ^^
		Others/Mixed Substrates	Yes	[85] NS - Croatia (60); [86] Caucasian (80); [87] Caucasian (31); [88] In Vitro - cell lines.	[85] Phenobarbital conc. in cerebrospinal fluid; [86] CC genotype = cannabis dependence; [87] Rhodamine 123 efflux; [88] Altered substrate specificity (range of compounds).
			No	[89] Caucasian (16); [90] NS - Turkey (58); [91] Japanese (12); [92] Japanese (80); [93] Caucasian (20).	[89] loperamide disposition or CNS effects; [90] Losartan disposition^^; [91] Telmisartan PK^^; [92] Phenolic MPA glucuronide PK; [93] fexofenadine PK.

13	**E29/3751G>A (V1251I)**	Immunosuppressants	Yes	[1] *In vitro* - cell lines.	[1] Modulate CsA effect on the intercellular accumulation of BODIPY-FL-paclitaxel transport.
		No report for the following drug classes/indications		Antiepileptic Drugs, Antibiotics, Antidepressants, Anti Cancer Agents, Cardiac Glycoside, Antiretroviral Therapy (HIV), Opiate analgesic, Statins	
14	**Haplotype consisting of two or all of SNPs: E13/1236C>T, E22/2677G>T/A and E27/3435C>T**	Antiepileptic Drugs	Yes	[94] *In vitro* - Cell Lines; [95] Caucasian (289).	[94] PK of various AEDs; [95] 3435TT and 2677TT = reduced drug resistance.
			No	[49] Korean (193); [96] North Indian (325); [97] Caucasian (463); [98] Indian (369).	[49, 96-98] Drug efficacy/response.^^
		Antibiotics	Yes	[99] Han Chinese (20).	[99] 2677TT and 3435TT = lower azithromycin plasma conc, higher Tmax, lower AUC.
		Antidepressants	Yes	[100] NS - Germany (15); [101] Japanese (68); [102] NS - Italy (60).	[100] 2677G>T influences citalopram plasma and CSF conc, 3435C>T does not; [101] Paroxetine efficacy; [102] 3435CC = higher clozapine dose required to equal plasma conc.
			No	[103] NS - Croatia (127).	[103] Paroxetine efficacy.
		Anti Cancer Agents	Yes	[104] Japanese (145); [105] NS - USA (73); [106] NS - Singapore (62); [1] *In vitro* - cell lines; [107] Han Chinese (69); [108] Korean (118); [109] Han Chinese (54); [110] NS - Germany (1047), Spain (49); [111] NS - Korea	[104] TTT = reduced clearance of irinotecan; [105] docataxel side effects; [106] doxorubicin PK; [1] paclitaxel transport; [107] Vinorelbine outcome; [108] 2677 genotype = hematological/gastrointestinal toxicities; [109] docetaxel-cisplatin efficacy; [110] lower mitoxantrone efflux.^^ [111] 3435CT = shorter OS, 2677GG = paclitaxel/doxorubicin resistance; [112] Adverse side effects, 3435CT and TT genotypes = mucositis, 2677T/A = diarrhea; [113] Imatinib efficacy.

				(121); [112] Korean (161); [113] NS - France (90).	
			No	[114] NS - France (42).	[114] Erlotinib PK.^^
		Cardiac Glycoside	Yes	[115] Caucasian (687); [116] NS - Japan (15); [117] Caucasian (25), African (6), Asian (1); [117] NS - Netherlands (195); [119] Han Chinese (20).	[115] 3435TT = higher Cmax and AUC, 2677T/3435T = higher AUC, 2677G/3435C = lower AUC; [116] 2677GG/3435CC = bioavailability; [117] 3435T and 2677T = higher conc. [118] TTT genotype = higher serum conc in the elderly; [119] PK differs between TTT-TTT and TGC-CGC carriers.
		Antiretroviral Therapy (HIV)	Yes	[120] Caucasian (118); [121] African American (13), Hispanic (4), Asian (1), Caucasian (13).	[120] 3435CT/2677TT = lower atazanavir levels; [121] CGC predicts slower oral clearance of atazanavir an ritonavir.
			No	[122] NS - USA (103), [123] NS - Belgium (53); [124] NS - Spain (115).	[122] atazanavir or lopinavir trough concentrations; [123] Lopinavir PK^^; [124] Tenofovir induced kidney tubular dysfunction^^.

		Immunosuppressants	Yes	[125] Chinese (106), Malay (92) and Indian (91); [126] Caucasian (73), African (7), Indian (1); [127] NS - USA. (65); [54] Caucasian (91); [128] NS - Canada (69); [129] NS - China (129); [130] NS - Netherlands (104); [131] Caucasian - Czech Republic (832); [132] Chinese (112); [133] Caucasian (38), Asian (10) and 3 Black Caribbean; [134] NS - Portugal (30).	[125] TTT = CsA PK - higher AUC and Cmax; [126] 2677T allele = TRL PK and dose requirements, 2677T/3434T = dose requirements; [127] 2677T/3434T = TRL blood conc. @ 6 and 12 months; [54]. TRL dosage requirements; [128] 2677GG/3435CC = CsA exposure @ 1 wk only; [129] CsA blood conc. [130] CsA oral bioavailibilty in children; [131] TRL/Cyclosporin efficacy - risk of acute rejection^^; [132] CGC, TGT and TTC genotypes assoc with CsA conc; [133] TTT = TRL induced nephrotoxicity, higher dose adjusted pre-dose conc.; [134] higher TRL conc. with 1236T and 2677T/A alleles.
			No	[135] Caucasian (82), African (2), South Asian (1); [136] Korean (29); [137] Caucasian (95); [138] NS - China (104); [139] NS - Norway (25); [140] NS - Switzerland (19);[141] NS - France (136); [142] NS - Belgium (29); [143] Japanese (63); [144] Korean (568); [145] Caucasian (32); [146] Caucasian (50); [147] Caucasian (192).	[135] Sirolimus and TRL PK and dose requirements^^; [136] TRL PK; 24. TRL PK; [137] TRL PK; [138] TRL PK^^; [139] CsA PK; [140] CsA PK in PBMCs; [141] TRL efficacy^^; [142] TRL/Fluconazole PK; [143] TRL PK^^; [144] TRL PK^^; [145] TRL dosage requirements^^; [146] TRL PK^^; [147] TRL and CsA PK, efficacy or nephrotoxicity.

	Opiate analgesic	Yes	[148] (138) Sephardic (Western Europe, Balkans and Morocco) (39%), Ashkenazi (Central and Eastern Europe) (22%), Oriental (Iraq, Iran, Yemen and Syria) (16%), mixed (13%) and unknown (10%); [149] Japanese (32).	[148] Methadone dose requirement; [149] 2677G and 3435C = reduced chance of vomiting.
	Statins	Yes	[150] Finnish (534); [151] Caucasian (85); 5. Korean (28).	[150] Simvastatin and Atorvastatin PK; [151] TTT and CGT = Simvastatin Efficacy; 5. 2677TT/3435TT = atorvastatin lactone, 2-hydroxyatorvastatin and 2-hydroxyatorvastatin lactone PK.
		No	[152] Caucasian (20).	[152] Fluvastatin, pravastatin, lovastatin, and rosuvastatin PK.
	Others/Mixed Substrates	Yes	[153] Caucasians (46); [154] Han Chinese (24); [155] *In vitro* - cell lines; [156] Korean (33).	[153] TTT haplotype = lower plasma concentration of the active metabolite of risperidone, 9-hydroxyrisperidone; [154] TT/TT and GT/CT = lower AUC, TT/TT = higher oral clearnce of Verapamil; [155] 2677G>T assoc. with Ibutilide resistance. 3435T associated with disrupted protein trafficking - reversed with fexofenadine; [156] 2677AA/3435CC = lower plasma conc. of fexofenadine.
		No	[157] Korean (104); [158] Korean (30); [159] Han Chinese (24); [160] Korean (10); [161] Japanese (95); [162] Japanese (65); [163] *In vitro* - cell lines.	[157] Cilostazol PK^^; [158] Rebamipide PK; [159] Valacyclovir absorption; [160] verapamil PK; [161] Prednisolone efficacy; [162] Prednisolone PK; [163] verapamil, digoxin, vinblastine and cyclosporin A PK.

Studies are grouped according to the substrates investigated. *The racial background of the cohort population is indicated. In the event that the study did not specify the racial background, an NS (Not Specified) is stated along with the country in which the study was conducted. ^^ Indicates that SNPs or SNP haplotypes located in other genes, including CYP3A4, CYP3A5 or other ABC proteins were associated with drug pharmacokinetics/response to an equal or greater extent than those of ABCB1.

Notes on drug classes:

Antiepileptic Drugs (AED)	Includes phenytoin, carbamazepine, lamotrigine, phenobarbital, valproic acid, levetiracetam, and gabapentin.
Antibiotics	Includes dicloxacillin, amoxicillin, clarithromycin and azithromycin
Antidepressants	Includes citalopram, clozapine, paroxetine, venlafaxine, mirtazapine.
Anti Cancer Agents (antimetabolite, general cytotoxics, tyrosine kinase inhibitors)	Includes methotrexate (MTX), docetaxel, irinotecan, doxorubicin, daunorubicin, vincristine, dexamethasone, epirubicin, imatinib, temozolamide, paclitaxel, cisplatin, mitoxantrone, and erlotinib.
Cardiac Glycoside	Includes digoxin
Antiretroviral Therapy (HIV)	Includes nelfinavir, efavirenz, atazanavir, indinavir, nevirapine, saquinavir, ritonavir, lopinavir, and tenofovir.
Immunosuppressants	Includes tacrolimus (TRL), cyclosporin A (CsA), and sirolimus.
Opiate analgesic	Includes Methadone and Morphine.
Statins	Includes simvastatin, atorvastatin, pravastatin, fluvastatin, pravastatin, lovastatin, and rosuvastatin.
Others/Mixed Substrates	Includes telmisartan, phenobarbital, cannabis, rhodamine 123, loperamide, losartan, telmisatan, phenolic MPA glucuronide, talinolol, digoxin, fexofenadine, verapamil, and ibutilide.

The reference numbers in Appendix 1 are in accordance to the reference list provided *Steven J. Wolf*, Maulana Bachtiar*, Jingbo Wang, Tiow Suan Sim, Samuel S Chong, and Caroline G.L. Lee,* ***Pharmacogenomics Journal****, 2011, 11: 315-325*

[Reference for Appendix 1]

[1] Gow, J.M., et al., *Substrate-dependent effects of human ABCB1 coding polymorphisms.* J Pharmacol Exp Ther, 2008. **325**(2): p. 435-42.

[2] Crettol, S., et al., *CYP3A7, CYP3A5, CYP3A4, and ABCB1 genetic polymorphisms, cyclosporine concentration, and dose requirement in transplant recipients.* Ther Drug Monit, 2008. **30**(6): p. 689-99.

[3] Crettol, S., et al., *ABCB1 and cytochrome P450 genotypes and phenotypes: influence on methadone plasma levels and response to treatment.* Clin Pharmacol Ther, 2006. **80**(6): p. 668-81.

[4] Kimchi-Sarfaty, C., J.J. Gribar, and M.M. Gottesman, *Functional characterization of coding polymorphisms in the human MDR1 gene using a vaccinia virus expression system.* Mol Pharmacol, 2002. **62**(1): p. 1-6.

[5] Jeong, H., et al., *Function-altering SNPs in the human multidrug transporter gene ABCB1 identified using a Saccharomyces-based assay.* PLoS Genet, 2007. **3**(3): p. e39.

[6] Choi, K.H., et al., *An altered pattern of cross-resistance in multidrug-resistant human cells results from spontaneous mutations in the mdr1 (P-glycoprotein) gene.* Cell, 1988. **53**(4): p. 519-29.

[7] Crouthamel, M.H., et al., *A novel MDR1 G1199T variant alters drug resistance and efflux transport activity of P-glycoprotein in recombinant Hek cells.* J Pharm Sci, 2006. **95**(12): p. 2767-77.

[8] Green, H., et al., *ABCB1 G1199A polymorphism and ovarian cancer response to paclitaxel.* J Pharm Sci, 2008. **97**(6): p. 2045-8.

[9] Tsai, S.M., et al., *Side effects after docetaxel treatment in Taiwanese breast cancer patients with CYP3A4, CYP3A5, and ABCB1 gene polymorphisms.* Clin Chim Acta, 2009. **404**(2): p. 160-5.

[10] Woodahl, E.L., et al., *MDR1 (ABCB1) G1199A (Ser400Asn) polymorphism alters transepithelial permeability and sensitivity to anticancer agents.* Cancer Chemother Pharmacol, 2009. **64**(1): p. 183-8.

[11] Innocenti, F., et al., *Comprehensive pharmacogenetic analysis of irinotecan neutropenia and pharmacokinetics.* J Clin Oncol, 2009. **27**(16): p. 2604-14.

[12] Liu, L., et al., *MDR1 C2005T polymorphism changes substrate specificity.* Cancer Chemother Pharmacol, 2010.

[13] Woodahl, E.L., et al., *Multidrug resistance gene G1199A polymorphism alters efflux transport activity of P-glycoprotein.* J Pharmacol Exp Ther, 2004. **310**(3): p. 1199-207.

[14] Woodahl, E.L., et al., *MDR1 G1199A polymorphism alters permeability of HIV protease inhibitors across P-glycoprotein-expressing epithelial cells.* AIDS, 2005. **19**(15): p. 1617-25.

[15] Fanta, S., et al., *Pharmacogenetics of cyclosporine in children suggests an age-dependent influence of ABCB1 polymorphisms.* Pharmacogenet Genomics, 2008. **18**(2): p. 77-90.

[16] Yin, O.Q., B. Tomlinson, and M.S. Chow, *Effect of multidrug resistance gene-1 (ABCB1) polymorphisms on the single-dose pharmacokinetics of cloxacillin in healthy adult Chinese men.* Clin Ther, 2009. **31**(5): p. 999-1006.

[17] Ranganathan, P., et al., *Methotrexate (MTX) pathway gene polymorphisms and their effects on MTX toxicity in Caucasian and African American patients with rheumatoid arthritis.* J Rheumatol, 2008. **35**(4): p. 572-9.

[18] Mathijssen, R.H., et al., *Irinotecan pathway genotype analysis to predict pharmacokinetics.* Clin Cancer Res, 2003. **9**(9): p. 3246-53.

[19] Schaich, M., et al., *A MDR1 (ABCB1) gene single nucleotide polymorphism predicts outcome of temozolomide treatment in glioblastoma patients.* Ann Oncol, 2009. **20**(1): p. 175-81.

[20] Qiu, X.Y., et al., *Association of MDR1, CYP3A4*18B, and CYP3A5*3 polymorphisms with cyclosporine pharmacokinetics in Chinese renal transplant recipients.* Eur J Clin Pharmacol, 2008. **64**(11): p. 1069-84.

[21] Bozina, N., et al., *Associations between MDR1 gene polymorphisms and schizophrenia and therapeutic response to olanzapine in female schizophrenic patients.* J Psychiatr Res, 2008. **42**(2): p. 89-97.

[22] Schaefer, M., I. Roots, and T. Gerloff, *In-vitro transport characteristics discriminate wild-type ABCB1 (MDR1) from ALA893SER and ALA893THR polymorphisms.* Pharmacogenet Genomics, 2006. **16**(12): p. 855-61.

[23] Johnatty, S.E., et al., *ABCB1 (MDR 1) polymorphisms and progression-free survival among women with ovarian cancer following paclitaxel/carboplatin chemotherapy.* Clin Cancer Res, 2008. **14**(17): p. 5594-601.

[24] Bohanec Grabar, P., et al., *Genetic determinants of methotrexate toxicity in rheumatoid arthritis patients: a study of polymorphisms affecting methotrexate transport and folate metabolism.* Eur J Clin Pharmacol, 2008. **64**(11): p. 1057-68.

[25] Bartnicka, L., et al., *Effect of ABCB1 (MDR1) 3435C >T and 2677G >A,T polymorphisms and P-glycoprotein inhibitors on salivary digoxin secretion in congestive heart failure patients.* Pharmacol Rep, 2007. **59**(3): p. 323-9.

[26] Yamauchi, A., et al., *Neurotoxicity induced by tacrolimus after liver transplantation: relation to genetic polymorphisms of the ABCB1 (MDR1) gene.* Transplantation, 2002. **74**(4): p. 571-2.

[27] Singh, R., et al., *ABCB1 G2677 allele is associated with high dose requirement of cyclosporin A to prevent renal allograft rejection in North India.* Arch Med Res, 2008. **39**(7): p. 695-701.

[28] Chen, B., et al., *Influence of the MDR1 haplotype and CYP3A5 genotypes on cyclosporine blood level in Chinese renal transplant recipients.* Xenobiotica, 2009. **39**(12): p. 931-8.

[29] Mega, J.L., et al., *Identification of genetic variants associated with response to statin therapy.* Arterioscler Thromb Vasc Biol, 2009. **29**(9): p. 1310-5.

[30] Siegmund, W., et al., *The effects of the human MDR1 genotype on the expression of duodenal P-glycoprotein and disposition of the probe drug talinolol.* Clin Pharmacol Ther, 2002. **72**(5): p. 572-83.

[31] Kim, R.B., et al., *Identification of functionally variant MDR1 alleles among European Americans and African Americans.* Clin Pharmacol Ther, 2001. **70**(2): p. 189-99.

[32] Kerb, R., et al., *The predictive value of MDR1, CYP2C9, and CYP2C19 polymorphisms for phenytoin plasma levels.* Pharmacogenomics J, 2001. **1**(3): p. 204-10.

[33] Simon, T., et al., *Genetic determinants of response to clopidogrel and cardiovascular events.* N Engl J Med, 2009. **360**(4): p. 363-75.

[34] Dericioglu, N., et al., *Multidrug resistance in patients undergoing resective epilepsy surgery is not associated with C3435T polymorphism in the ABCB1 (MDR1) gene.* Epilepsy Res, 2008. **80**(1): p. 42-6.

[35] Bournissen, F.G., et al., *Polymorphism of the MDR1/ABCB1 C3435T drug-transporter and resistance to anticonvulsant drugs: a meta-analysis.* Epilepsia, 2009. **50**(4): p. 898-903.

[36] Turgut, G., et al., *Association of MDR1 C3435T polymorphism with bipolar disorder in patients treated with valproic acid.* Mol Biol Rep, 2009. **36**(3): p. 495-9.

[37] Putnam, W.S., et al., *Effect of the MDR1 C3435T variant and P-glycoprotein induction on dicloxacillin pharmacokinetics.* J Clin Pharmacol, 2005. **45**(4): p. 411-21.

[38] Oh, J.H., et al., *Effects of CYP2C19 and MDR1 genotype on the eradication rate of Helicobacter pylori infection by triple therapy with pantoprazole, amoxycillin and clarithromycin.* J Gastroenterol Hepatol, 2009. **24**(2): p. 294-8.

[39] Roberts, R.L., et al., *A common P-glycoprotein polymorphism is associated with nortriptyline-induced postural hypotension in patients treated for major depression.* Pharmacogenomics J, 2002. **2**(3): p. 191-6.

[40] Alenius, M., et al., *Gene polymorphism influencing treatment response in psychotic patients in a naturalistic setting.* J Psychiatr Res, 2008. **42**(11): p. 884-93.

[41] Kooloos, W.M., et al., *Functional polymorphisms and methotrexate treatment outcome in recent-onset rheumatoid arthritis.* Pharmacogenomics, 2010. **11**(2): p. 163-75.

[42] Erdilyi, D.J., et al., *Synergistic interaction of ABCB1 and ABCG2 polymorphisms predicts the prevalence of toxic encephalopathy during anticancer chemotherapy.* Pharmacogenomics J, 2008. **8**(5): p. 321-7.

[43] Paule, B., et al., *MDR1 polymorphism role in patients treated with cetuximab and irinotecan in irinotecan refractory colorectal cancer.* Med Oncol, 2009.

[44] Lara, P.N., Jr., et al., *Phase III trial of irinotecan/cisplatin compared with etoposide/cisplatin in extensive-stage small-cell lung cancer: clinical and pharmacogenomic results from SWOG S0124.* J Clin Oncol, 2009. **27**(15): p. 2530-5.

[45] Drain, S., et al., *ABCB1 (MDR1) rs1045642 is associated with increased overall survival in plasma cell myeloma.* Leuk Lymphoma, 2009. **50**(4): p. 566-70.

[46] Cizmarikova, M., et al., *MDR1 (C3435T) polymorphism: relation to the risk of breast cancer and therapeutic outcome.* Pharmacogenomics J, 2010. **10**(1): p. 62-9.

[47] Goh, B.C., et al., *Explaining interindividual variability of docetaxel pharmacokinetics and pharmacodynamics in Asians through phenotyping and genotyping strategies.* J Clin Oncol, 2002. **20**(17): p. 3683-90.

[48] Gandara, D.R., et al., *Japanese-US common-arm analysis of paclitaxel plus carboplatin in advanced non-small-cell lung cancer: a model for assessing population-related pharmacogenomics.* J Clin Oncol, 2009. **27**(21): p. 3540-6.

[49] Kim, D.W., et al., *Lack of association between ABCB1, ABCG2, and ABCC2 genetic polymorphisms and multidrug resistance in partial epilepsy.* Epilepsy Res, 2009. **84**(1): p. 86-90.

[50] Hoffmeyer, S., et al., *Functional polymorphisms of the human multidrug-resistance gene: multiple sequence variations and correlation of one allele with P-glycoprotein expression and activity in vivo.* Proc Natl Acad Sci U S A, 2000. **97**(7): p. 3473-8.

[51] Cascorbi, I., et al., *Frequency of single nucleotide polymorphisms in the P-glycoprotein drug transporter MDR1 gene in white subjects.* Clin Pharmacol Ther, 2001. **69**(3): p. 169-74.

[52] Sakaeda, T., et al., *MDR1 genotype-related pharmacokinetics of digoxin after single oral administration in healthy Japanese subjects.* Pharm Res, 2001. **18**(10): p. 1400-4.

[53] Morita, Y., et al., *MDR1 genotype-related duodenal absorption rate of digoxin in healthy Japanese subjects.* Pharm Res, 2003. **20**(4): p. 552-6.

[54] Verstuyft, C., et al., *Dipyridamole enhances digoxin bioavailability via P-glycoprotein inhibition.* Clin Pharmacol Ther, 2003. **73**(1): p. 51-60.

[55] Comets, E., et al., *Modelling the influence of MDR1 polymorphism on digoxin pharmacokinetic parameters.* Eur J Clin Pharmacol, 2007. **63**(5): p. 437-49.

[56] Becquemont, L., et al., *Effect of grapefruit juice on digoxin pharmacokinetics in humans.* Clin Pharmacol Ther, 2001. **70**(4): p. 311-6.

[57] Chowbay, B., et al., *Meta-analysis of the influence of MDR1 C3435T polymorphism on digoxin pharmacokinetics and MDR1 gene expression.* Br J Clin Pharmacol, 2005. **60**(2): p. 159-71.

[58] Dragonas, C., et al., *The association of ABCB1 polymorphisms and elevated serum digitoxin concentrations in geriatric patients.* Eur J Clin Pharmacol, 2008. **64**(4): p. 367-72.

[59] Fellay, J., et al., *Response to antiretroviral treatment in HIV-1-infected individuals with allelic variants of the multidrug resistance transporter 1: a pharmacogenetics study.* Lancet, 2002. **359**(9300): p. 30-6.

[60] Saitoh, A., et al., *An MDR1-3435 variant is associated with higher plasma nelfinavir levels and more rapid virologic response in HIV-1 infected children.* AIDS, 2005. **19**(4): p. 371-80.

[61] Rodriguez Novoa, S., et al., *Plasma levels of atazanavir and the risk of hyperbilirubinemia are predicted by the 3435C-->T polymorphism at the multidrug resistance gene 1.* Clin Infect Dis, 2006. **42**(2): p. 291-5.

[62] Solas, C., et al., *Minimal effect of MDR1 and CYP3A5 genetic polymorphisms on the pharmacokinetics of indinavir in HIV-infected patients.* Br J Clin Pharmacol, 2007. **64**(3): p. 353-62.

[63] Mahungu, T.W., et al., *The relationships of ABCB1 3435C>T and CYP2B6 516G>T with high-density lipoprotein cholesterol in HIV-infected patients receiving Efavirenz.* Clin Pharmacol Ther, 2009. **86**(2): p. 204-11.

[64] Ciccacci, C., et al., *Nevirapine-induced hepatotoxicity and pharmacogenetics: a retrospective study in a population from Mozambique.* Pharmacogenomics, 2010. **11**(1): p. 23-31.

[65] Rahi, M., et al., *Influence of adenosine triphosphate and ABCB1 (MDR1) genotype on the P-glycoprotein-dependent transfer of saquinavir in the dually perfused human placenta.* Hum Exp Toxicol, 2008. **27**(1): p. 65-71.

[66] Macphee, I.A., et al., *Tacrolimus pharmacogenetics: polymorphisms associated with expression of cytochrome p4503A5 and P-glycoprotein correlate with dose requirement.* Transplantation, 2002. **74**(11): p. 1486-9.

[67] Yates, C.R., et al., *The effect of CYP3A5 and MDR1 polymorphic expression on cyclosporine oral disposition in renal transplant patients.* J Clin Pharmacol, 2003. **43**(6): p. 555-64.

[68] Bonhomme-Faivre, L., et al., *MDR-1 C3435T polymorphism influences cyclosporine a dose requirement in liver-transplant recipients.* Transplantation, 2004. **78**(1): p. 21-5.

[69] Jin, J., et al., *Impact of multidrug resistance 1 gene polymorphism on tacrolimus dose and concentration-to-dose ratio in Chinese liver transplantation recipients.* Zhonghua Yi Xue Yi Chuan Xue Za Zhi, 2005. **22**(6): p. 616-20.

[70] Azarpira, N., et al., *Association between cyclosporine concentration and genetic polymorphisms of CYP3A5 and MDR1 during the early stage after renal transplantation.* Exp Clin Transplant, 2006. **4**(1): p. 416-9.

[71] Foote, C.J., et al., *MDR1 C3435T polymorphisms correlate with cyclosporine levels in de novo renal recipients.* Transplant Proc, 2006. **38**(9): p. 2847-9.

[72] Akbas, S.H., et al., *The effect of MDR1 (ABCB1) polymorphism on the pharmacokinetic of tacrolimus in Turkish renal transplant recipients.* Transplant Proc, 2006. **38**(5): p. 1290-2.

[73] Li, D., et al., *Tacrolimus dosing in Chinese renal transplant patients is related to MDR1 gene C3435T polymorphisms.* Transplant Proc, 2006. **38**(9): p. 2850-2.

[74] De Iudicibus, S., et al., *Role of MDR1 gene polymorphisms in gingival overgrowth induced by cyclosporine in transplant patients.* J Periodontal Res, 2008. **43**(6): p. 665-72.

[75] Loh, P.T., et al., *Significant impact of gene polymorphisms on tacrolimus but not cyclosporine dosing in Asian renal transplant recipients.* Transplant Proc, 2008. **40**(5): p. 1690-5.

[76] Bonhomme-Faivre, L., et al., *Effect of the ABCB1 3435C>T polymorphism on tacrolimus concentrations and dosage requirements in liver transplant recipients.* Am J Health Syst Pharm, 2009. **66**(18): p. 1645-51.

[77] von Ahsen, N., et al., *No influence of the MDR-1 C3435T polymorphism or a CYP3A4 promoter polymorphism (CYP3A4-V allele) on dose-adjusted cyclosporin A trough concentrations or rejection incidence in stable renal transplant recipients.* Clin Chem, 2001. **47**(6): p. 1048-52.

[78] Min, D.I. and V.L. Ellingrod, *C3435T mutation in exon 26 of the human MDR1 gene and cyclosporine pharmacokinetics in healthy subjects.* Ther Drug Monit, 2002. **24**(3): p. 400-4.

[79] Goto, M., et al., *C3435T polymorphism in the MDR1 gene affects the enterocyte expression level of CYP3A4 rather than Pgp in recipients of living-donor liver transplantation.* Pharmacogenetics, 2002. **12**(6): p. 451-7.

[80] Jiang, Z.P., et al., *Meta-analysis of the effect of MDR1 C3435T polymorphism on cyclosporine pharmacokinetics.* Basic Clin Pharmacol Toxicol, 2008. **103**(5): p. 433-44.

[81] Miao, L.Y., et al., *Association study of ABCB1 and CYP3A5 gene polymorphisms with sirolimus trough concentration and dose requirements in Chinese renal transplant recipients.* Biopharm Drug Dispos, 2008. **29**(1): p. 1-5.

[82] Barrera-Pulido, L., et al., *Clinical relevance and prevalence of polymorphisms in CYP3A5 and MDR1 genes that encode tacrolimus biotransformation enzymes in liver transplant recipients.* Transplant Proc, 2008. **40**(9): p. 2949-51.

[83] Campa, D., et al., *Association of ABCB1/MDR1 and OPRM1 gene polymorphisms with morphine pain relief.* Clin Pharmacol Ther, 2008. **83**(4): p. 559-66.

[84] Becker, M.L., et al., *Influence of genetic variation in CYP3A4 and ABCB1 on dose decrease or switching during simvastatin and atorvastatin therapy.* Pharmacoepidemiol Drug Saf, 2010. **19**(1): p. 75-81.

[85] Basic, S., et al., *The influence of C3435T polymorphism of ABCB1 gene on penetration of phenobarbital across the blood-brain barrier in patients with generalized epilepsy.* Seizure, 2008. **17**(6): p. 524-30.

[86] Benyamina, A., et al., *Association between ABCB1 C3435T polymorphism and increased risk of cannabis dependence.* Prog Neuropsychopharmacol Biol Psychiatry, 2009. **33**(7): p. 1270-4.

[87] Hitzl, M., et al., *The C3435T mutation in the human MDR1 gene is associated with altered efflux of the P-glycoprotein substrate rhodamine 123 from CD56+ natural killer cells.* Pharmacogenetics, 2001. **11**(4): p. 293-8.

[88] Komar, A.A., *Silent SNPs: impact on gene function and phenotype.* Pharmacogenomics, 2007. **8**(8): p. 1075-80.

[89] Pauli-Magnus, C., et al., *No effect of MDR1 C3435T variant on loperamide disposition and central nervous system effects.* Clin Pharmacol Ther, 2003. **74**(5): p. 487-98.

[90] Yasar, U., M.O. Babaoglu, and A. Bozkurt, *Disposition of a CYP2C9 phenotyping agent, losartan, is not influenced by the common 3435C > T variation of the drug transporter gene ABCB1 (MDR1).* Basic Clin Pharmacol Toxicol, 2008. **103**(2): p. 176-9.

[91] Miura, M., et al., *Telmisartan pharmacokinetics in Japanese renal transplant recipients.* Clin Chim Acta, 2009. **399**(1-2): p. 83-7.

[92] Miura, M., et al., *Influence of drug transporters and UGT polymorphisms on pharmacokinetics of phenolic glucuronide metabolite of mycophenolic acid in Japanese renal transplant recipients.* Ther Drug Monit, 2008. **30**(5): p. 559-64.

[93] Drescher, S., et al., *MDR1 gene polymorphisms and disposition of the P-glycoprotein substrate fexofenadine.* Br J Clin Pharmacol, 2002. **53**(5): p. 526-34.

[94] Hung, C.C., et al., *Functional evaluation of polymorphisms in the human ABCB1 gene and the impact on clinical responses of antiepileptic drugs.* Pharmacogenet Genomics, 2008. **18**(5): p. 390-402.

[95] Sanchez, M.B., et al., *Genetic factors associated with drug-resistance of epilepsy: relevance of stratification by patient age and aetiology of epilepsy.* Seizure, 2010. **19**(2): p. 93-101.

[96] Lakhan, R., et al., *No association of ABCB1 polymorphisms with drug-refractory epilepsy in a north Indian population.* Epilepsy Behav, 2009. **14**(1): p. 78-82.

[97] Ufer, M., et al., *Non-response to antiepileptic pharmacotherapy is associated with the ABCC2 -24C>T polymorphism in young and adult patients with epilepsy.* Pharmacogenet Genomics, 2009. **19**(5): p. 353-62.

[98] Vahab, S.A., et al., *Analysis of genotype and haplotype effects of ABCB1 (MDR1) polymorphisms in the risk of medically refractory epilepsy in an Indian population.* Drug Metab Pharmacokinet, 2009. **24**(3): p. 255-60.

[99] He, X.J., et al., *Influence of ABCB1 gene polymorphisms on the pharmacokinetics of azithromycin among healthy Chinese Han ethnic subjects.* Pharmacol Rep, 2009. **61**(5): p. 843-50.

[100] Nikisch, G., C.B. Eap, and P. Baumann, *Citalopram enantiomers in plasma and cerebrospinal fluid of ABCB1 genotyped depressive patients and clinical response: a pilot study.* Pharmacol Res, 2008. **58**(5-6): p. 344-7.

[101] Kato, M., et al., *ABCB1 (MDR1) gene polymorphisms are associated with the clinical response to paroxetine in patients with major depressive disorder.* Prog Neuropsychopharmacol Biol Psychiatry, 2008. **32**(2): p. 398-404.

[102] Consoli, G., et al., *ABCB1 polymorphisms are associated with clozapine plasma levels in psychotic patients.* Pharmacogenomics, 2009. **10**(8): p. 1267-76.

[103] Mihaljevic Peles, A., et al., *MDR1 gene polymorphism: therapeutic response to paroxetine among patients with major depression.* Prog Neuropsychopharmacol Biol Psychiatry, 2008. **32**(6): p. 1439-44.

[104] Sai, K., et al., *Haplotype analysis of ABCB1/MDR1 blocks in a Japanese population reveals genotype-dependent renal clearance of irinotecan.* Pharmacogenetics, 2003. **13**(12): p. 741-57.

[105] Sissung, T.M., et al., *ABCB1 genetic variation influences the toxicity and clinical outcome of patients with androgen-independent prostate cancer treated with docetaxel.* Clin Cancer Res, 2008. **14**(14): p. 4543-9.

[106] Lal, S., et al., *Influence of ABCB1 and ABCG2 polymorphisms on doxorubicin disposition in Asian breast cancer patients.* Cancer Sci, 2008. **99**(4): p. 816-23.

[107] Pan, J.H., et al., *MDR1 single nucleotide polymorphisms predict response to vinorelbine-based chemotherapy in patients with non-small cell lung cancer.* Respiration, 2008. **75**(4): p. 380-5.

[108] Kim, H.S., et al., *Genetic polymorphisms affecting clinical outcomes in epithelial ovarian cancer patients treated with taxanes and platinum compounds: a Korean population-based study.* Gynecol Oncol, 2009. **113**(2): p. 264-9.

[109] Pan, J.H., et al., *MDR1 single nucleotide polymorphism G2677T/A and haplotype are correlated with response to docetaxel-cisplatin chemotherapy in patients with non-small-cell lung cancer.* Respiration, 2009. **78**(1): p. 49-55.

[110] Cotte, S., et al., *ABC-transporter gene-polymorphisms are potential pharmacogenetic markers for mitoxantrone response in multiple sclerosis.* Brain, 2009. **132**(Pt 9): p. 2517-30.

[111] Chang, H., et al., *Association of the ABCB1 gene polymorphisms 2677G>T/A and 3435C>T with clinical outcomes of paclitaxel monotherapy in metastatic breast cancer patients.* Ann Oncol, 2009. **20**(2): p. 272-7.

[112] Chang, H., et al., *Association of the ABCB1 3435C>T polymorphism and treatment outcomes in advanced gastric cancer patients treated with paclitaxel-based chemotherapy.* Oncol Rep, 2010. **23**(1): p. 271-8.

[113] Dulucq, S., et al., *Multidrug resistance gene (MDR1) polymorphisms are associated with major molecular responses to standard-dose imatinib in chronic myeloid leukemia.* Blood, 2008. **112**(5): p. 2024-7.

[114] Thomas, F., et al., *Population pharmacokinetics of erlotinib and its pharmacokinetic/pharmacodynamic relationships in head and neck squamous cell carcinoma.* Eur J Cancer, 2009. **45**(13): p. 2316-23.

[115] Johne, A., et al., *Modulation of steady-state kinetics of digoxin by haplotypes of the P-glycoprotein MDR1 gene.* Clin Pharmacol Ther, 2002. **72**(5): p. 584-94.

[116] Kurata, Y., et al., *Role of human MDR1 gene polymorphism in bioavailability and interaction of digoxin, a substrate of P-glycoprotein.* Clin Pharmacol Ther, 2002. **72**(2): p. 209-19.

[117] Verstuyft, C., et al., *Digoxin pharmacokinetics and MDR1 genetic polymorphisms.* Eur J Clin Pharmacol, 2003. **58**(12): p. 809-12.

[118] Aarnoudse, A.J., et al., *Common ATP-binding cassette B1 variants are associated with increased digoxin serum concentration.* Pharmacogenet Genomics, 2008. **18**(4): p. 299-305.

[119] Xu, P., et al., *Impact of MDR1 haplotypes derived from C1236T, G2677T/A and C3435T on the pharmacokinetics of single-dose oral digoxin in healthy Chinese volunteers.* Pharmacology, 2008. **82**(3): p. 221-7.

[120] Rodriguez-Novoa, S., et al., *Genetic factors influencing atazanavir plasma concentrations and the risk of severe hyperbilirubinemia.* AIDS, 2007. **21**(1): p. 41-6.

[121] Anderson, P.L., et al., *Atazanavir pharmacokinetics in genetically determined CYP3A5 expressors versus non-expressors.* J Antimicrob Chemother, 2009. **64**(5): p. 1071-9.

[122] Ma, Q., et al., *Multidrug resistance 1 polymorphisms and trough concentrations of atazanavir and lopinavir in patients with HIV.* Pharmacogenomics, 2007. **8**(3): p. 227-35.

[123] Elens, L., et al., *Association between ABCC2 polymorphism and lopinavir accumulation in peripheral blood mononuclear cells of HIV-infected patients.* Pharmacogenomics, 2009. **10**(10): p. 1589-97.

[124] Rodriguez-Novoa, S., et al., *Predictors of kidney tubular dysfunction in HIV-infected patients treated with tenofovir: a pharmacogenetic study.* Clin Infect Dis, 2009. **48**(11): p. e108-16.

[125] Chowbay, B., et al., *Genetic polymorphisms in MDR1 and CYP3A4 genes in Asians and the influence of MDR1 haplotypes on cyclosporin disposition in heart transplant recipients.* Pharmacogenetics, 2003. **13**(2): p. 89-95.

[126] Anglicheau, D., et al., *Association of the multidrug resistance-1 gene single-nucleotide polymorphisms with the tacrolimus dose requirements in renal transplant recipients.* J Am Soc Nephrol, 2003. **14**(7): p. 1889-96.

[127] Zheng, H., et al., *Tacrolimus dosing in pediatric heart transplant patients is related to CYP3A5 and MDR1 gene polymorphisms.* Am J Transplant, 2003. **3**(4): p. 477-83.

[128] Foote, C.J., et al., *Polymorphisms of multidrug resistance gene (MDR1) and cyclosporine absorption in de novo renal transplant patients.* Transplantation, 2007. **83**(10): p. 1380-4.

[129] Zhang, Y.T., et al., *ABCB1 polymorphisms may have a minor effect on ciclosporin blood concentrations in myasthenia gravis patients.* Br J Clin Pharmacol, 2008. **66**(2): p. 240-6.

[130] Hesselink, D.A., et al., *A drug transporter for all ages? ABCB1 and the developmental pharmacogenetics of cyclosporine.* Pharmacogenomics, 2008. **9**(6): p. 783-9.

[131] Bandur, S., et al., *Haplotypic structure of ABCB1/MDR1 gene modifies the risk of the acute allograft rejection in renal transplant recipients.* Transplantation, 2008. **86**(9): p. 1206-13.

[132] Wang, Y., et al., *Effect of genetic polymorphisms of CYP3A5 and MDR1 on cyclosporine concentration during the early stage after renal transplantation in Chinese patients co-treated with diltiazem.* Eur J Clin Pharmacol, 2009. **65**(3): p. 239-47.

[133] Hawwa, A.F., et al., *Influence of ABCB1 polymorphisms and haplotypes on tacrolimus nephrotoxicity and dosage requirements in children with liver transplant.* Br J Clin Pharmacol, 2009. **68**(3): p. 413-21.

[134] Mendes, J., et al., *Genetic polymorphisms in CYP3A5 and MDR1 genes and their correlations with plasma levels of tacrolimus and cyclosporine in renal transplant recipients.* Transplant Proc, 2009. **41**(3): p. 840-2.

[135] Mourad, M., et al., *Sirolimus and tacrolimus trough concentrations and dose requirements after kidney transplantation in relation to CYP3A5 and MDR1 polymorphisms and steroids.* Transplantation, 2005. **80**(7): p. 977-84.

[136] Choi, J.H., et al., *Influence of the CYP3A5 and MDR1 genetic polymorphisms on the pharmacokinetics of tacrolimus in healthy Korean subjects.* Br J Clin Pharmacol, 2007. **64**(2): p. 185-91.

[137] Kuypers, D.R., et al., *CYP3A5 and CYP3A4 but not MDR1 single-nucleotide polymorphisms determine long-term tacrolimus disposition and drug-related nephrotoxicity in renal recipients.* Clin Pharmacol Ther, 2007. **82**(6): p. 711-25.

[138] Li, D., et al., *Population pharmacokinetics of tacrolimus and CYP3A5, MDR1 and IL-10 polymorphisms in adult liver transplant patients.* J Clin Pharm Ther, 2007. **32**(5): p. 505-15.

[139] Falck, P., et al., *Reduced elimination of cyclosporine A in elderly (>65 years) kidney transplant recipients.* Transplantation, 2008. **86**(10): p. 1379-83.

[140] Ansermot, N., et al., *Influence of ABCB1 gene polymorphisms and P-glycoprotein activity on cyclosporine pharmacokinetics in peripheral blood mononuclear cells in healthy volunteers.* Drug Metab Lett, 2008. **2**(2): p. 76-82.

[141] Quteineh, L., et al., *Influence of CYP3A5 genetic polymorphism on tacrolimus daily dose requirements and acute rejection in renal graft recipients.* Basic Clin Pharmacol Toxicol, 2008. **103**(6): p. 546-52.

[142] Kuypers, D.R., et al., *Effects of CYP3A5 and MDR1 single nucleotide polymorphisms on drug interactions between tacrolimus and fluconazole in renal allograft recipients.* Pharmacogenet Genomics, 2008. **18**(10): p. 861-8.

[143] Hosohata, K., et al., *MDR1 haplotypes conferring an increased expression of intestinal CYP3A4 rather than MDR1 in female living-donor liver transplant patients.* Pharm Res, 2009. **26**(7): p. 1590-5.

[144] Jun, K.R., et al., *Tacrolimus concentrations in relation to CYP3A and ABCB1 polymorphisms among solid organ transplant recipients in Korea.* Transplantation, 2009. **87**(8): p. 1225-31.

[145] Provenzani, A., et al., *The effect of CYP3A5 and ABCB1 single nucleotide polymorphisms on tacrolimus dose requirements in Caucasian liver transplant patients.* Ann Transplant, 2009. **14**(1): p. 23-31.

[146] Zhao, W., et al., *Population pharmacokinetics and pharmacogenetics of tacrolimus in de novo pediatric kidney transplant recipients.* Clin Pharmacol Ther, 2009. **86**(6): p. 609-18.

[147] Taegtmeyer, A.B., et al., *ATP-binding cassette subfamily B member 1 polymorphisms do not determine cyclosporin exposure, acute rejection or nephrotoxicity after heart transplantation.* Transplantation, 2010. **89**(1): p. 75-82.

[148] Levran, O., et al., *ABCB1 (MDR1) genetic variants are associated with methadone doses required for effective treatment of heroin dependence.* Hum Mol Genet, 2008. **17**(14): p. 2219-27.

[149] Fujita, K.I., et al., *Association of UGT2B7 and ABCB1 genotypes with morphine-induced adverse drug reactions in Japanese patients with cancer.* Cancer Chemother Pharmacol, 2009.

[150] Keskitalo, J.E., et al., *ABCB1 haplotypes differentially affect the pharmacokinetics of the acid and lactone forms of simvastatin and atorvastatin.* Clin Pharmacol Ther, 2008. **84**(4): p. 457-61.

[151] Becker, M.L., et al., *Common genetic variation in the ABCB1 gene is associated with the cholesterol-lowering effect of simvastatin in males.* Pharmacogenomics, 2009. **10**(11): p. 1743-51.

[152] Keskitalo, J.E., et al., *No significant effect of ABCB1 haplotypes on the pharmacokinetics of fluvastatin, pravastatin, lovastatin, and rosuvastatin.* Br J Clin Pharmacol, 2009. **68**(2): p. 207-13.

[153] Gunes, A., et al., *ABCB1 polymorphisms influence steady-state plasma levels of 9-hydroxyrisperidone and risperidone active moiety.* Ther Drug Monit, 2008. **30**(5): p. 628-33.

[154] Zhao, L.M., et al., *Influence of ABCB1 gene polymorphisms on the pharmacokinetics of verapamil among healthy Chinese Han ethnic subjects.* Br J Clin Pharmacol, 2009. **68**(3): p. 395-401.

[155] McBride, B.F., T. Yang, and D.M. Roden, *Influence of the G2677T/C3435T haplotype of MDR1 on P-glycoprotein trafficking and ibutilide-induced block of HERG.* Pharmacogenomics J, 2009. **9**(3): p. 194-201.

[156] Yi, S.Y., et al., *A variant 2677A allele of the MDR1 gene affects fexofenadine disposition.* Clin Pharmacol Ther, 2004. **76**(5): p. 418-27.

[157] Yoo, H.D., H.Y. Cho, and Y.B. Lee, *Population pharmacokinetic analysis of cilostazol in healthy subjects with genetic polymorphisms of CYP3A5, CYP2C19 and ABCB1.* Br J Clin Pharmacol, 2010. **69**(1): p. 27-37.

[158] Cho, H.Y., et al., *Pharmacokinetics and bioequivalence of two formulations of rebamipide 100-mg tablets: a randomized, single-dose, two-period, two-sequence crossover study in healthy Korean male volunteers.* Clin Ther, 2009. **31**(11): p. 2712-21.

[159] Zhang, Y., et al., *MDR1 genotypes do not influence the absorption of a single oral dose of 600 mg valacyclovir in healthy Chinese Han ethnic males.* Br J Clin Pharmacol, 2008. **66**(2): p. 247-54.

[160] Pan, W., et al., *Dietary salt does not influence the disposition of verapamil enantiomers in relation to efflux transporter ABCB1 genetic polymorphism in healthy Korean subjects.* Xenobiotica, 2008. **38**(4): p. 422-34.

[161] Miura, M., et al., *Influence of CYP3A5, ABCB1 and NR1I2 polymorphisms on prednisolone pharmacokinetics in renal transplant recipients.* Steroids, 2008. **73**(11): p. 1052-9.

[162] Miura, M., et al., *Inter-individual difference determinant of prednisolone pharmacokinetics for Japanese renal transplant recipients in the maintenance stage.* Xenobiotica, 2009. **39**(12): p. 939-45.

[163] Morita, N., T. Yasumori, and K. Nakayama, *Human MDR1 polymorphism: G2677T/A and C3435T have no effect on MDR1 transport activities.* Biochem Pharmacol, 2003. **65**(11): p. 1843-52.

Appendix 2. Drug pathways curated by the PharmGKB database

ID	Drug-response Pathway	Total Genes	PharmGKB Pathway	No. Genes	Associated Tissue(s) of Expression	Associated Drug(s)
1	ACE inhibitor	19	ACE inhibitor pathway	19	Non-tissue specific	Ace Inhibitors, Plain
2	Anti diabetic	33	Anti diabetic drug pathway (Nateglinide PK)	3	Liver	Nateglinide
			Anti diabetic drug pathway (Potassium channel inhibitors PD)	29	Pancreatic β cells	Chlorpropamide, gliclazide, glimepiride, glipizide, nateglinide, repaglinide, tolbutamide
			Anti diabetic drug pathway (Repaglinide PK)	3	Liver	Repaglinide
3	Anti estrogen	19	Anti estrogen pathway (Aromatase inhibitor)	5	Adrenals, ovary, peripheral tissues, liver & circulation, breast	Anti estrogen drugs inhibiting aromatase-mediated synthesis
			Anti estrogen pathway (Estrogen metabolism)	11	Liver and peripheral tissues	Anti estrogen
			Anti estrogen pathway (Summary)	4	Breast tissues	Anastrozole, estradiol, estrogens, estrone, exemestane, letrozole, raloxifene, tamoxifen, toremifene
			Anti estrogen pathway (Tamoxifen PK)	6	Liver and breast tissues	Raloxifene, tamoxifen, toremifene
4	Antiarrhythmic	59	Antiarrhythmic Drug Pathways	59	Cardiomyocyte	Amiodarone, antiarrhythmics, class i and iii, arsenic trioxide, cisapride, disopyramide, dofetilide, droperidol, flecainide, halofantrine, haloperidol, ibutilide, lidocaine, mesoridazine,

						methadone, mexiletine, pentamidine, pimozide, procainamide, propafenone, quinidine, sotalol, sparfloxacin, thioridazine, tocainide
5	Antiplatelet	9	Antiplatelet Drug Clopidogrel Pathway (PK)	9	Liver, intestine, platelet	Clopidogrel
6	Benzodiazepine	25	Benzodiazepine pathway (PD)	14	Presynaptic/postsynaptic neurons	Benzodiazepine
			Benzodiazepine pathway (PK)	19		Alprazolam, bromazepam, clonazepam, diazepam, flunitrazepam, flurazepam, lorazepam, midazolam, oxazepam, temazepam, triazolam
7	Beta-agonist and beta-blocker	71	Beta-agonist and beta-blocker Pathway	71	Airways	Albuterol, metaproterenol, fenoterol, salmeterol, formoterol
8	Bisphosphonate	23	Bisphosphonate Pathway	23	Osteoclasts	Alendronate, clodronate, etidronic acid, ibandronate, pamidronate, risedronate, tiludronate, zoledronate
9	Celecoxib	27	Celecoxib (Celebrex) Pathway	27		Acetaminophen, aspirin, bromfenac, celecoxib, diclofenac, etodolac, fenoprofen, flurbiprofen, ibuprofen, indomethacin, ketoprofen, ketorolac, meclofenamic acid, mefenamic acid, meloxicam, nabumetone, naproxen, oxaprozin,

						phenylbutazone, piroxicam, rofecoxib, valdecoxib
10	Citalopram	4	Citalopram (PK)	4	Liver, intestine	Citalopram
11	Codeine and morphine	9	Codeine and Morphine Pathway (PK)	9	Liver	Codeine, morphine
12	Cyclophosphamide	24	Cyclophosphamide Pathway	17	Non-tissue specific cancer cells	Cyclophosphamide
			Cyclophosphamide Pathway (PK)	7	Liver	
13	Doxorubicin	74	Doxorubicin (Cancer PD)	18	Cancer cells	Doxorubicin
			Doxorubicin (Cardio PD)	15	Cardiomyocyte	Dexrazoxane, doxorubicin
			Doxorubicin Pathway	56		Doxorubicin
			Doxorubicin (PK)	22		
14	EGFR inhibitors	66	EGFR Inhibitor Pathway (PD)	66	Non-tissue specific cancer cells	Cetuximab, erlotinib, gefitinib, lapatinib
15	Erlotinib	8	Erlotinib Pathway (PK)	8		Erlotinib
16	Etoposide	16	Etoposide Pathway	16		Antineoplastic agents, dexamethasone, etoposide, rifampin
17	Fluoropyrimidine	33	Fluoropyrimidine (PD)	10	Non-tissue specific cancer cells	Capecitabine, fluorouracil, tegafur
			Fluoropyrimidine (PK)	24		
18	Gefitinib	8	Gefitinib Pathway (PK)	8		Gefitinib
19	Gemcitabine	10	Gemcitabine Pathway	10	Non-tissue specific cancer cells	Antineoplastic agents, gemcitabine
20	Glucocorticoid and inflammatory	16	Glucocorticoid and Inflammatory genes Pathway (HPA axis)	7	Hypothalamic pituitary adrenal axis	Corticotropin, dexamethasone, glucocorticoids

			Glucocorticoid and Inflammatory genes Pathway (PD)	9	CNS and peripheral tissues (heart, lung, vasculature, and gut)	Budesonide, cortisone acetate, dexamethasone, glucocorticoids, mifepristone, prednisone
			Glucocorticoid and Inflammatory genes Pathway (Regulation)	0[#]		
21	Ifosfamide	19	Ifosfamide Pathway	13	Non-tissue specific cancer cells	Antineoplastic agents, ifosfamide
			Ifosfamide Pathway (PK)	6	Liver	Ifosfamide
22	Imatinib	12	Imatinib	12	Blood, intestine, liver	Dasatinib, imatinib
23	Imipramine desipramine	4	Imipramine Desipramine Pathway (PK)	4	Liver	Desipramine, imipramine
24	Irinotecan	23	Irinotecan pathway	14		Antineoplastic agents, irinotecan
			Irinotecan Pathway (Cancer)	19	Non-tissue specific cancer cells	Irinotecan
25	Losartan	5	Losartan (PK)	5	Liver	Losartan
26	Methotrexate	30	Methotrexate Pathway	30		Antineoplastic agents, cyanocobalamin, folic acid, leucovorin, methotrexate, pyridoxine
27	Nicotine	34	Nicotine Pathway	10	Liver	Nicotine
			Nicotine PD Pathway (Chromaffin Cell)	4	chromaffin cells	
			Nicotine PD Pathway (Dopaminergic Neuron)	21	dopaminergic neurons	
28	Phenytoin	10	Phenytoin PK Pathway	10	Liver	Phenytoin
29	Platelet aggregation	56	Platelet Aggregation Pathway (PD)	56	Platelet	Abciximab, aspirin, clopidogrel, eptifibatide, ticlopidine, tirofiban
30	Platinum	45	Platinum Pathway	45		Antineoplastic agents, cisplatin, oxaliplatin, platinum
31	Proton pump inhibitor	53	Proton Pump Inhibitor (PD)	50		Esomeprazole, lansoprazole, omeprazole, pantoprazole, rabeprazole

			Proton Pump Inhibitor (PK)	3		Omeprazole
32	RAAS	20	Renin-Angiotensin-Aldosterone-System-acting Drug Pathway	20		Ace Inhibitors, Plain, Aldosterone antagonists, aliskiren, Angiotensin II Antagonists, benazepril, candesartan, captopril, cilazapril, enalapril, eplerenone, eprosartan, fosinopril, irbesartan, lisinopril, loratadine, losartan, moexipril, olmesartan, perindopril, quinapril, ramipril, spirapril, spironolactone, tasosartan, telmisartan, trandolapril, valsartan
33	SSRI	33	Selective Serotonin Reuptake Inhibitors (SSRI) Pathway	28	Presynaptic/postsynaptic neurons	Citalopram, fluoxetine, paroxetine, sertraline
			SSRI Fluoxetine Pathway (PK)	5	Liver	Fluoxetine
34	Statin	46	Statin Pathway (Atorvastatin Lovastatin and Simvastatin PK)	14	Liver, intestine	Atorvastatin, lovastatin, simvastatin
			Statin Pathway (Cholesterol and Lipoprotein Transport PD)	27	Liver, enterocyte, peripheral tissue	Atorvastatin, fluvastatin, hmg coa reductase inhibitors, lovastatin, pravastatin, rosuvastatin, simvastatin
			Statin Pathway (Fluvastatin PK)	13	Liver, intestine	Fluvastatin
			Statin Pathway (PK)	19	Liver, intestine, kidney	Atorvastatin, fluvastatin, hmg coa reductase inhibitors, lovastatin, pravastatin, rosuvastatin, simvastatin
			Statin Pathway (Pravastatin PK)	9		Pravastatin
			Statin Pathway (Rosuvastatin PK)	4	Liver, intestine	Rosuvastatin

35	Sympathetic nerve	31	Sympathetic Nerve Pathway (Neuroeffector Junction)	22	Neuroeffector junction	Acebutolol, atenolol, Beta Blocking Agents, brimonidine, clonidine, desipramine, disulfiram, dobutamine, dopamine, epinephrine, esmolol, formoterol, guanfacine, isoproterenol, l-methyldopa, methoxamine, metoprolol, nadolol, norepinephrine, orciprenaline, phenoxybenzamine, phentolamine, phenylephrine, prazosin, propranolol, ritodrine, salbutamol, salmeterol, selective beta-2-adrenoreceptor agonists, terazosin, terbutaline, timolol, tolazoline, yohimbine
			Sympathetic Nerve Pathway (Pre- and Post-Ganglionic Junction)	13	Sympathetic neuronal junction	Acetylcholine, carbachol, carbidopa, metyrosine, nicotine, trimethaphan
36	Taxane	16	Taxane Pathway	16		Docetaxel, paclitaxel
37	Tenofovir adefovir	7	Tenofovir Adefovir pathway	7	Kidney	Adefovir dipivoxil, tenofovir
38	Thiopurine	33	Thiopurine Pathway	33		Antineoplastic agents, azathioprine, folic acid, mercaptopurine, thioguanine
39	VEGF	65	VEGF Pathway	65	Endothelial cell	Antineoplastic agents, bevacizumab, semaxanib, sorafenib, vatalanib

40	Vinca alkaloids	10	Vinca Alkaloids (PK)	10	Liver	Vincristine
41	Warfarin	24	Warfarin Pathway (PD)	15		Warfarin
			Warfarin Pathway (PK)	9		

#The Glucocorticoid and Inflammatory genes Pathway (Regulation) of the PharmGKB did not have gene index that could be mapped into the NCBI Gene database

Appendix 3. SNP density of the drug-response pathways

ID	Drug Response Pathway	SNP Density / Kbp					
		Promoter	Intron	5' UTR	3' UTR	Non-synonymous	Synonymous
1	ACE inhibitor	8.81	10.13	9.40	8.06	6.46	4.53
2	Anti diabetic	8.42	8.50	8.45	12.09	6.76	3.71
3	Anti estrogen	11.44	12.46	12.72	11.53	14.45	7.08
4	Antiarrhythmic	8.61	7.87	10.85	11.26	6.31	4.11
5	Antiplatelet	10.79	12.32	25.83	18.76	17.18	5.52
6	Benzodiazepine	10.60	9.62	9.14	9.07	9.55	4.47
7	Beta-agonist and beta-blocker	7.83	7.78	5.65	11.77	3.88	3.68
8	Bisphosphonate	8.74	9.90	9.13	11.18	8.51	4.13
9	Celecoxib	10.24	11.55	12.43	15.10	10.39	5.35
10	Citalopram	14.00	16.32	17.55	11.92	28.01	7.91
11	Codeine and morphine	11.98	13.65	10.57	16.16	18.59	6.24
12	Cyclophosphamide	10.27	11.01	11.82	14.27	12.05	5.69
13	Doxorubicin	10.04	10.30	12.66	16.81	11.52	5.44
14	EGFR inhibitors	7.77	8.23	6.11	9.84	3.98	3.34
15	Erlotinib	12.32	10.87	18.22	16.54	15.72	3.40
16	Etoposide	10.50	10.45	15.86	16.59	10.57	4.78
17	Fluoropyrimidine	9.72	11.38	18.39	14.18	6.78	4.55
18	Gefitinib	14.39	13.33	19.68	14.71	21.51	5.92
19	Gemcitabine	10.46	9.59	32.46	12.18	6.71	3.60
20	Glucocorticoid and inflammatory	8.02	8.69	9.54	9.55	4.08	3.23
21	Ifosfamide	10.28	11.20	13.26	16.70	12.43	5.73
22	Imatinib	12.00	12.45	16.21	16.37	16.40	5.93
23	Imipramine desipramine	13.77	18.44	12.77	7.53	29.79	8.19
24	Irinotecan	10.43	10.59	19.33	19.64	13.94	4.27
25	Losartan	12.51	11.66	7.16	21.98	16.58	5.89
26	Methotrexate	9.62	10.07	12.73	10.42	5.64	4.03
27	Nicotine	9.17	10.51	4.07	15.31	8.86	6.13
28	Phenytoin	10.98	11.20	12.67	13.96	18.82	4.16
29	Platelet aggregation	8.61	10.18	7.01	13.22	9.08	4.62
30	Platinum	8.96	8.96	10.69	14.86	10.65	3.64
31	Proton pump inhibitor	8.46	8.49	6.45	11.71	4.73	3.18
32	RAAS	8.66	10.07	9.36	7.61	6.21	4.49

33	SSRI	8.20	8.62	9.56	8.01	7.56	4.52
34	Statin	10.92	11.36	14.67	12.99	12.65	4.65
35	Sympathetic nerve	9.77	7.50	9.04	14.08	7.00	5.06
36	Taxane	10.57	12.55	16.88	15.73	8.97	4.38
37	Tenofovir adefovir	8.21	8.18	17.72	9.89	7.21	4.34
38	Thiopurine	9.27	9.55	11.44	11.47	6.48	4.28
39	VEGF	8.53	9.40	7.56	11.00	4.39	3.90
40	Vinca alkaloids	8.67	9.79	15.91	13.83	8.21	3.39
41	Warfarin	9.91	10.22	9.67	12.75	10.38	4.55

Appendix 4. Common DRP genes housing one or more expression-associated functional SNPs

ID	Pathway	Expression-associated potentially functional SNPs in Common Genes								
		rSNPs		srSNPs				cSNPs		
		TF Binding Sites	miRNA Binding Sites	3' UTR Conserved	Splicing Regulatory Sites	Nonsense-mediated Decay	Codon Usage Difn.	Protein Deleterious	Post Translation Modif. Sites	Protein Domains
1	ACE inhibitor				RYR2 (rs2275288)					
2	Anti diabetic									
3	Anti estrogen	UGT1A1 (rs10929302)								
4	Antiarrhythmic									
5	Antiplatelet	ABCB1 (rs3747802)								
6	Benzodiazepine									

7	Beta-agonist and beta-blocker				GNAS (rs919196), PLCB2 (rs2229690, rs12439272)						
8	Bisphosphonate										
9	Celecoxib	PTGS1 (rs10985624)	PTGS1 (rs10306189, rs10306190)		AKR1C3 (rs34376012), PTGS1 (rs10306173, rs3842803)						
10	Citalopram	ABCB1 (rs3747802)									
11	Codeine and morphine	ABCB1 (rs3747802), UGT1A1 (rs10929302)									
12	Cyclophosphamide	AKR1A1 (rs11211139, rs9147), ERCC1 (rs3212935)			AKR1A1 (rs2088102), ERCC1 (rs3212955, rs3212935, rs2298881, rs3212978)						
13	Doxorubicin	ABCB1 (rs3747802), AKR1A1 (rs11211139, rs9147), GSTM1 (rs4147563, rs4147563), GSTP1 (rs7927381, rs1871041)			RYR2 (rs2275288), TOP2B (rs11709485, rs10510570, rs11712723, rs17016894)						

14	EGFR inhibitors	MAPK8 (rs10857561), SOS1 (rs10165968)			MAP2K2 (rs2289859), MAPK8 (rs2289805), SOS1 (rs963730, rs10190377, rs7565814)						
15	Erlotinib	ABCB1 (rs3747802), UGT1A1 (rs10929302)									
16	Etoposide	ABCB1 (rs3747802), GSTM1 (rs4147563, rs4147563), GSTP1 (rs7927381, rs1871041), PTGS1 (rs10985624), UGT1A1 (rs10929302)	PTGS1 (rs10306189, rs10306190)		PTGS1 (rs10306173, rs3842803), TOP2B (rs11709485, rs10510570, rs11712723, rs17016894)						
17	Fluoropyrimidine	SLC29A1 (rs3734701)	TYMS (rs699517)		SLC29A1 (rs1128930), TYMS (rs1059394, rs2612098, rs2853532, rs2853534, rs2853537, rs2853740)						
18	Gefitinib	ABCB1 (rs3747802)									
19	Gemcitabine	SLC29A1 (rs3734701)			SLC29A1 (rs1128930)						

20	Glucocorticoid and inflammatory									
21	Ifosfamide	AKR1A1 (rs11211139, rs9147)			AKR1A1 (rs2088102)					
22	Imatinib	ABCB1 (rs3747802)								
23	Imipramine desipramine									
24	Irinotecan	ABCB1 (rs3747802), UGT1A1 (rs10929302)			XRCC1 (rs762507)		XRCC1 (rs3547)			
25	Losartan	UGT1A1 (rs10929302)								
26	Methotrexate	ABCB1 (rs3747802), MTRR (rs162028, rs162029, rs162030, rs16879248, rs16879259, rs2966952, rs326118, rs3733781), SHMT1(rs643333, rs2688052)	SHMT1 (rs12952556), TYMS (rs699517)	MTRR (rs9332), SHMT1 (rs12952556)	MTRR (rs12347, rs161869, rs161870, rs162037, rs2303081, rs3776454), SHMT1 (rs672356), TYMS (rs1059394, rs2612098, rs2853532, rs2853534, rs2853537, rs2853740)		MTRR (rs12347)	MTRR (rs10380)		MTRR (rs10380)
27	Nicotine									
28	Phenytoin	UGT1A1 (rs10929302)								
29	Platelet aggregation	PTGS1 (rs10985624)	PTGS1 (rs10306189, rs10306190)		PTGS1 (rs10306173, rs3842803)					

30	Platinum	ERCC1 (rs3212935), GSTM1 (rs4147563, rs4147563), GSTP1 (rs7927381, rs1871041)			ERCC1 (rs3212955, rs3212935, rs2298881, rs3212978), XRCC1 (rs762507)		XRCC1 (rs3547)			
31	Proton pump inhibitor	ABCB1 (rs3747802)			PLCB2 (rs2229690, rs12439272)					
32	RAAS									
33	SSRI				GNAS (rs919196), PLCB2 (rs2229690, rs12439272)					
34	Statin	ABCB1 (rs3747802), UGT1A1 (rs10929302)			SLCO1B3 (rs17680137, rs7973653)			SLCO1B3 (rs60140950)		
35	Sympathetic nerve									
36	Taxane	ABCB1 (rs3747802)			SLCO1B3 (rs17680137, rs7973653)			SLCO1B3 (rs60140950)		
37	Tenofovir adefovir									
38	Thiopurine	GSTM1 (rs4147563, rs4147563), MTRR (rs162028, rs162029, rs162030, rs16879248, rs16879259, rs2966952,	SHMT1 (rs12952556), TYMS (rs699517)	MTRR (rs9332), SHMT1 (rs12952556)	MTRR (rs12347, rs161869, rs161870, rs162037, rs2303081, rs3776454), SHMT1 (rs672356), SLC29A1		MTRR (rs12347)	MTRR (rs10380)		MTRR (rs10380)

		rs326118, rs3733781), SHMT1(rs643333, rs2688052), SLC29A1 (rs3734701)			(rs1128930), TYMS (rs1059394, rs2612098, rs2853532, rs2853534, rs2853537, rs2853740)					
39	VEGF	MAPK8 (rs10857561), SOS1 (rs10165968)			MAP2K2 (rs2289859), MAPK8 (rs2289805), SOS1 (rs963730, rs10190377, rs7565814)					
40	Vinca alkaloids	ABCB1 (rs3747802)								
41	Warfarin	ABCB1 (rs3747802)								

Appendix 5. Common DRP genes housing one or more highly-differentiated functional SNPs

ID	Pathway	Highly-differentiated Potentially Functional SNPs in Common Genes								
		rSNPs		srSNPs				cSNPs		
		TF Binding Sites	miRNA Binding Sites	3' UTR Conserved	Splicing Regulatory Sites	Nonsense-mediated Decay	Codon Usage Difn.	Protein Deleterious	Post Translation Modif. Sites	Protein Domains
1	ACE inhibitor									
2	Anti diabetic	AKT1 (rs2494752)			ABCC8 (rs2077654, rs12293228, rs2077655), CACNA1D (rs6766988)					
3	Anti estrogen	CYP1B1 (rs162556)		CYP1B1 (rs2855658)	CYP1B1 (rs1056837), CYP3A5 (rs776746)					

4	Antiarrhythmic				ABCC8 (rs2077654, rs12293228, rs2077655), CACNA1D (rs6766988)					
5	Antiplatelet				CYP3A5 (rs776746)					
6	Benzodiazepine				CYP3A5 (rs776746)					
7	Beta-agonist and beta-blocker	ADCY6 (rs9668292), ADCY9(rs2532001, rs11076808), PRKCG (rs2545046)			ADCY2 (rs326145), ADCY6 (rs3729980), ADCY9 (rs2072346), CACNA1B (rs6559261, rs7874239, rs7865267), CACNA1D (rs6766988), PLCB4 (rs7267438, rs1028338), PRKCH (rs2296274)			ADCY8 (rs13278912)		ADCY8 (rs13278912)
8	Bisphosphonate	VDR (rs11568820)			FDFT1 (rs4841600, rs2409836, rs10903343)					
9	Celecoxib	AKT1 (rs2494752), PTGS2 (rs2143416)			CYP3A5 (rs776746), PTGS1 (rs5788, rs10306150)					
10	Citalopram									

11	Codeine and morphine									
12	Cyclophosphamide	ADH4 (rs3762894, rs1800759), ADH5 (rs1312200), ALDH2 (rs886205)			ADH1A (rs3819197), CYP3A5 (rs776746)					
13	Doxorubicin	MLH1 (rs9311149), RALBP1 (rs7245045)			ABCG2 (rs2231164, rs2725267), CYP3A5 (rs776746), TOP1 (rs2235362)					
14	EGFR inhibitors	AKT1 (rs2494752), GRB2 (rs930296, rs4789193, rs930297), MAP2K2 (rs7258783), PRKCG (rs2545046)			PLCG2 (rs4611451, rs4258608)					
15	Erlotinib				ABCG2 (rs2231164, rs2725267), CYP3A5 (rs776746)					
16	Etoposide	PTGS2 (rs2143416), VDR (rs11568820)			CYP3A5 (rs776746), PTGS1 (rs5788, rs10306150)					
17	Fluoropyrimidine	RRM1 (rs1561876, rs3794050,			ABCG2 (rs2231164, rs2725267), DHFR					

		rs1465952, rs1662162)			(rs10072026)					
18	Gefitinib				ABCG2 (rs2231164, rs2725267), CYP3A5 (rs776746)					
19	Gemcitabine	RRM1 (rs1561876, rs3794050, rs1465952, rs1662162)								
20	Glucocorticoid and inflammatory									
21	Ifosfamide	ADH4 (rs3762894, rs1800759), ADH5 (rs1312200), ALDH2 (rs886205)			ADH1A (rs3819197), CYP3A5 (rs776746)					
22	Imatinib				ABCG2 (rs2231164, rs2725267), CYP3A5 (rs776746)					
23	Imipramine desipramine									
24	Irinotecan				ABCG2 (rs2231164, rs2725267), CYP3A5 (rs776746), TOP1 (rs2235362)					
25	Losartan									

26	Methotrexate				ABCG2 (rs2231164, rs2725267), DHFR (rs10072026), SLC22A8 (rs4149181)					
27	Nicotine				ADCY2 (rs326145), SLC18A2 (rs363333)					
28	Phenytoin				CYP3A5 (rs776746)					
29	Platelet aggregation				ADCY2 (rs326145), PLCB4 (rs7267438, rs1028338), PLCG2 (rs4611451, rs4258608), PTGS1 (rs5788, rs10306150)			ADCY8 (rs1327 8912)		ADCY8 (rs13278912)
30	Platinum	MLH1 (rs9311149)			ABCG2 (rs2231164, rs2725267)					
31	Proton pump inhibitor	ADCY6 (rs9668292), ADCY9 (rs2532001, rs11076808), PRKCG (rs2545046)			ADCY2 (rs326145), ADCY6 (rs3729980), ADCY9 (rs2072346), PLCB4 (rs7267438, rs1028338), PLCG2 (rs4611451, rs4258608),			ADCY8 (rs1327 8912)		ADCY8 (rs13278912)

					PRKCH (rs2296274)						
32	RAAS										
33	SSRI				ADCY2 (rs326145), CYP3A5 (rs776746), PLCB4 (rs7267438, rs1028338), SLC18A2 (rs363333)						
34	Statin				ABCG2 (rs2231164, rs2725267), CYP3A5 (rs776746), FDFT1 (rs4841600, rs2409836, rs10903343), SLC22A8 (rs4149181), SLCO1B3 (rs3764006)						
35	Sympathetic nerve				CACNA1B (rs6559261, rs7874239, rs7865267), SLC18A2 (rs363333)						

36	Taxane	CYP1B1 (rs162556)		CYP1B1 (rs2855658)	ABCG2 (rs2231164, rs2725267), CYP1B1 (rs1056837), CYP3A5 (rs776746), SLCO1B3 (rs3764006)					
37	Tenofovir adefovir				SLC22A8 (rs4149181)					
38	Thiopurine				DHFR (rs10072026)					
39	VEGF	AKT1 (rs2494752), GRB2 (rs930296, rs4789193, rs930297), MAP2K2 (rs7258783), PRKCG (rs2545046)			PLCG2 (rs4611451, rs4258608), PRKCH (rs2296274)					
40	Vinca alkaloids	RALBP1 (rs7245045)			CYP3A5 (rs776746)					
41	Warfarin									

Appendix 6. Studies describing drug-response variation

Drug-response Pathway	PharmGKB Pathway	Drug	Population Information					Reference		
			Populations involved	Same Response	Responders	Non-responders	Adverse drug reaction (ADR) or Toxicity	Remark	Author(s)	Journal Information
ACE inhibitor	ACE inhibitor pathway	ACE inhibitors	ASW_CEU		CEU	ASW		Whites had greater BP reduction with enalapril than blacks	Exner et al	N Engl J Med. 2001 May 3;344(18):1351-7.
	ACE inhibitor pathway	ACE inhibitors	ASW_CEU		CEU	ASW		Lesser Response to Angiotensin-Converting–Enzyme Inhibitor Therapy in Black as Compared with White Patients with Left Ventricular Dysfunction	Exner et al	N Engl J Med 2001; 344:1351-1357May 3, 2001
	ACE inhibitor pathway	Enalapril	ASW_CEU		CEU	ASW			Exner et al	N Engl J Med. 2001 May 3;344(18):1351-7.
	ACE inhibitor pathway	Angiotensin I	ASW_CEU		CEU	ASW		Black subjects required twice as much angiotensin I as Caucasian subjects to achieve similar DBP elevations	Joubert	Eur J Clin Pharmacol. 1990;39(2):183-5.
	ACE inhibitor pathway	ACE inhibitors	CHB_CEU				CHB	The high incidence and greater odds ratio of	Woo et al	

						cough among Chinese patients taking ACE inhibitors raises the possibility of an enhanced susceptibility to this adverse effect among the Chinese.		
	ACE inhibitor pathway	Captopril	ASW_CEU	CEU	ASW	Antihypertensive effect greater in Whites		Br J Clin Pharmacol. 1982;14 Suppl 2:97S-101S.
	ACE inhibitor pathway	Enalapril	CHS_MAS			Ethnic distribution of ADR reports showed statistically significant difference with the general population		
	ACE inhibitor pathway	Enalapril	CHS_INS			Ethnic distribution of ADR reports showed statistically significant difference with the general population		
	ACE inhibitor pathway	Enalapril	MAS_INS			Ethnic distribution of ADR reports showed statistically significant difference with the general population		
Anti diabetic	Anti diabetic drug pathway (Potassium channel inhibitors PD)	Insulin	CEU_GIH	GIH	CEU		Becker et al	Clin Pharmacokinet. 2008;47(1):7-20.
	Anti diabetic drug pathway (Potassium channel inhibitors PD)	Insulin	CHB_GIH	GIH	CHB	The time of onset of insulin effect was approximately 10 minutes earlier in the Indian subjects	Becker et al	Clin Pharmacokinet. 2008;47(1):7-20.
Anti estrogen	Anti estrogen pathway (Aromatase inhibitor)	Ethinyl estradiol	Nigeria, Singapore, Sri Lanka, and United States			Sex hormone concentrations were reported as being highest in women from Southeast Asia	Bennink HJ	Eur J Contracept Reprod Health Care. 2000 Sep;5 Suppl 2:12-20.
	Anti estrogen	Estrogen	CHB_CEU	CEU	CHB	Caucasians more	Massart F	Gynecol

pathway (Aromatase inhibitor)					estrogen-sensitive than other human races		Endocrinol. 2005 Jan;20(1):36-44.
Anti estrogen pathway (Aromatase inhibitor)	Estrogen	CEU_MEX	CEU	MEX	Caucasians more estrogen-sensitive than other human races	Massart F	Gynecol Endocrinol. 2005 Jan;20(1):36-44.
Anti estrogen pathway (Aromatase inhibitor)	Estrogen	CEU_YRI	CEU	YRI	Caucasians more estrogen-sensitive than other human races	Massart F	Gynecol Endocrinol. 2005 Jan;20(1):36-44.
Anti estrogen pathway (Estrogen metabolism)	Ethinyl estradiol	Nigeria, Singapore, Sri Lanka, and United States			Sex hormone concentrations were reported as being highest in women from Southeast Asia	Bennink HJ	Eur J Contracept Reprod Health Care. 2000 Sep;5 Suppl 2:12-20.
Anti estrogen pathway (Estrogen metabolism)	Estrogen	CHB_CEU	CEU	CHB	Caucasians more estrogen-sensitive than other human races	Massart F	Gynecol Endocrinol. 2005 Jan;20(1):36-44.
Anti estrogen pathway (Estrogen metabolism)	Estrogen	CEU_MEX	CEU	MEX	Caucasians more estrogen-sensitive than other human races	Massart F	Gynecol Endocrinol. 2005 Jan;20(1):36-44.
Anti estrogen pathway (Estrogen metabolism)	Estrogen	CEU_YRI	CEU	YRI	Caucasians more estrogen-sensitive than other human races	Massart F	Gynecol Endocrinol. 2005 Jan;20(1):36-44.
Anti estrogen pathway (Summary)	Ethinyl estradiol	Nigeria, Singapore, Sri Lanka, and United States			Sex hormone concentrations were reported as being highest in women from Southeast Asia	Bennink HJ	Eur J Contracept Reprod Health Care. 2000 Sep;5 Suppl 2:12-20.
Anti estrogen pathway (Summary)	Estrogen	CHB_CEU	CEU	CHB	Caucasians more estrogen-sensitive than other human races	Massart F	Gynecol Endocrinol. 2005 Jan;20(1):36-44.
Anti estrogen	Estrogen	CEU_MEX	CEU	MEX	Caucasians more	Massart F	Gynecol

	pathway (Summary)					estrogen-sensitive than other human races		Endocrinol. 2005 Jan;20(1):36-44.
	Anti estrogen pathway (Summary)	Estrogen	CEU_YRI	CEU	YRI	Caucasians more estrogen-sensitive than other human races	Massart F	Gynecol Endocrinol. 2005 Jan;20(1):36-44.
	Anti-arrhythmic Drug Pathways	Haloperidol	CHB_CEU/ASW	CHB	CEU/ ASW	These results suggest that the metabolism and disposition of haloperidol and reduced haloperidol could differ among ethnic populations	Jann et al	Psychiatry Res. 1989 Oct;30(1):45-52.
	Anti-arrhythmic Drug Pathways	Haloperidol	CHB_"Other populations"	CHB	"Other populations"	The Chinese group differed from the other ethnic populations	Jann et al	Prog Neuropsychopharmacol Biol Psychiatry. 1992 Mar;16(2):193-202.
Anti-arrhythmic	Anti-arrhythmic Drug Pathways	Haloperidol	ASIAN_CEU	CHB	CEU	The results were similar between the two Asian groups but significantly different between Caucasians and Asians.	Lin et al	J Clin Psychopharmacol. 1988 Jun;8(3):195-201.
	Anti-arrhythmic Drug Pathways	Haloperidol	CHB_CEU	CHB	CEU	Chinese schizophrenic patients (in the People's Republic of China) had 52% higher plasma haloperidol concentrations than U.S. non-Asian schizophrenic patients.	Potkin et al	Psychiatry Res. 1984 Jun;12(2):167-72.
	Anti-arrhythmic Drug Pathways	Haloperidol	CHB_ASW	CHB	ASW	Chinese patients were shown to produce 40–50% higher plasma haloperidol concentrations compared to non-	Poolsup et al	J Clin Pharm Ther. 2000 Jun;25(3):197-220.

	Anti-arrhythmic Drug Pathways	Haloperidol	CHB_CEU	CHB	CEU	Chinese patients (Caucasians and blacks) Chinese patients were shown to produce 40–50% higher plasma haloperidol concentrations compared to non-Chinese patients (Caucasians and blacks)	Poolsup et al	J Clin Pharm Ther. 2000 Jun;25(3):197-220.
	Anti-arrhythmic Drug Pathways	Haloperidol	CHB_CEU	CHB	CEU	Chinese patients were shown to produce 40–50% higher plasma haloperidol concentrations compared to non-Chinese patients (Caucasians and blacks)	Poolsup et al	J Clin Pharm Ther. 2000 Jun;25(3):197-220.
	Anti-arrhythmic Drug Pathways	Haloperidol	CHB_CEU	CHB	CEU	The higher haloperidol levels noted in Chinese may be explained by the lower clearance in this population (6·17 ml/min/kg) (120) compared with Caucasians reported in the literature (10·8 ml/min/kg) (121).	Poolsup et al	J Clin Pharm Ther. 2000 Jun;25(3):197-220.
Antiplatelet	Antiplatelet Drug Clopidogrel Pathway (PK)	Calcium channel blockers	ASW_CEU	ASW	CEU	Different reduction of blood pressure	Brewster et al	Ann Intern Med. 2004 Oct 19;141(8):614-27.
Benzo-diazepine	Benzo-diazepine pathway (PD)	Adinazolam	CHB_ASW	CHB	ASW	Greater exposure of adinazolam in Asians	Ajir et al	Psychopharmacology (Berl). 1997 Feb;129(3):265-70.

Benzoiazepine pathway (PD)	Adinazolam	CHB_CEU	CHB	CEU	Greater exposure of adinazolam in Asians	Ajir et al	Psychopharmacology (Berl). 1997 Feb;129(3):265-70.
Benzoiazepine pathway (PD)	Adinazolam	JPT_ASW	JPT	ASW	Greater exposure of adinazolam in Asians	Ajir et al	Psychopharmacology (Berl). 1997 Feb;129(3):265-70.
Benzoiazepine pathway (PD)	Adinazolam	JPT_CEU	JPT	CEU	Greater exposure of adinazolam in Asians	Ajir et al	Psychopharmacology (Berl). 1997 Feb;129(3):265-70.
Benzoiazepine pathway (PD)	Benzodiazepines, Alprazolam	CHB_CEU	CHB	CEU		Ghoneim et al	Clin Pharmacol Ther. 1981 Jun;29(6):749-56.
Benzoiazepine pathway (PD)	Triazolam	CEU_GIH	GIH	CEU	Higher Cmax was observed in Asians	Kinirons et al	Br J Clin Pharmacol. 1996 Jan;41(1):69-72.
Benzoiazepine pathway (PD)	Diazepam	CHB_CEU	CHB	CEU	Greater volume of distribution and clearance in Caucasians than Chinese	Kumana et al	Eur J Clin Pharmacol. 1987;32(2):211-5.
Benzoiazepine pathway (PD)	Alprazolam	CHB_CEU	CHB	CEU	Lower CLo, CLs in Asians than Caucasians	Lin et al	Psychopharmacology (Berl). 1988;96(3):365-9.
Benzoiazepine pathway (PD)	Alprazolam	CHB_CEU	CHB	CEU	Ethnic pharmacokinetic differences between Caucasians and Asians (30% greater AUC after oral dosing and 25% after IV in Asians)	Lin et al	Psychopharmacology (Berl). 1988;96(3):365-9.
Benzoiazepine pathway (PD)	Midazolam	JPT_CEU	JPT	CEU	25% lower systemic clearance of midazolam in Japanese as compared to Caucasians	Tateishi et al	Clin Pharmacol Ther. 2001 May;69(5):333-9.

Benzoiazepine pathway (PD)	Midazolam	ASW_CEU	ASW_ CEU			Similar CL, CL/F, F in African- and Caucasian Americans	Wandel et al	Clin Pharmacol Ther. 2000 Jul;68(1):82-91.
Benzoiazepine pathway (PD)	Diazepam	CHB_CEU		CHB	CEU	Lower clearance for diazepam in Chinese of the EM phenotype for S-mephenytoin versus historical controls	Zhang et al	Clin Pharmacol Ther. 1990 Nov;48(5):496-502.
Benzoiazepine pathway (PD)	Triazolam	CEU_GIH				Similar CLo, CLm in Asian Indians and Caucasians	Kinirons et al	Br J Clin Pharmacol. 1996 Jan;41(1):69-72.
Benzo-diazepine pathway (PK)	Benzo-diazepines, Alprazolam	CHB_CEU		CHB	CEU		Ghoneim et al	Clin Pharmacol Ther. 1981 Jun;29(6):749-56.
Benzo-diazepine pathway (PK)	Triazolam	CEU_GIH		GIH	CEU	Higher Cmax was observed in Asians	Kinirons et al	Br J Clin Pharmacol. 1996 Jan;41(1):69-72.
Benzo-diazepine pathway (PK)	Diazepam	CHB_CEU		CHB	CEU	Greater volume of distribution and clearance in Caucasians than Chinese	Kumana et al	Eur J Clin Pharmacol. 1987;32(2):211-5.
Benzo-diazepine pathway (PK)	Alprazolam	CHB_CEU		CHB	CEU	Lower CLo, CLs in Asians than Caucasians	Lin et al	Psychopharmacology (Berl). 1988;96(3):365-9.
Benzo-diazepine pathway (PK)	Alprazolam	CHB_CEU		CHB	CEU	Ethnic pharmacokinetic differences between Caucasians and Asians (30% greater AUC after oral dosing and 25% after IV in Asians)	Lin et al	Psychopharmacology (Berl). 1988;96(3):365-9.
Benzo-diazepine pathway (PK)	Midazolam	JPT_CEU		JPT	CEU	25% lower systemic clearance of midazolam in Japanese as compared to Caucasians	Tateishi et al	Clin Pharmacol Ther. 2001 May;69(5):333-9.
Benzo-	Midazolam	ASW_CEU	ASW_			Similar CL, CL/F, F in	Wandel et al	Clin Pharmacol

	diazepine pathway (PK)		CEU			African- and Caucasian Americans		Ther. 2000 Jul;68(1):82-91.
	Benzo-diazepine pathway (PK)	Diazepam	CHB_CEU		CHB	CEU	Lower clearance for diazepam in Chinese of the EM phenotype for S-mephenytoin versus historical controls	Zhang et al	Clin Pharmacol Ther. 1990 Nov;48(5):496-502.
	Benzo-diazepine pathway (PK)	Triazolam	CEU_GIH				Similar CLo, CLm in Asian Indians and Caucasians	Kinirons et al	Br J Clin Pharmacol. 1996 Jan;41(1):69-72.
	Beta-agonist and beta-blocker Pathway	Propranolol	MAS_CHS		MAS	CHS	Malays to have significantly greater propranolol responses compared to Chinese healthy male subjects	Rasool et al	Int J Clin Pharmacol Ther. 2000 May;38(5):260-9.
	Beta-agonist and beta-blocker Pathway	propranolol	CHB_CEU		CHB	CEU	Altered sensitivity to and clearance of propranolol in men of Chinese descent as compared with American whites	Zhou et al	N Engl J Med. 1989 Mar 2;320(9):565-70.
Beta-agonist and beta-blocker	Beta-agonist and beta-blocker Pathway	propranolol	CHB_CEU		CHB	CEU	Chinese men had a 2- to 3-fold greater sensitivity to propranolol's effect on heart rate and a 10-fold greater sensitivity to the BP effect than Caucasian men.	Zhou et al	N Engl J Med. 1989 Mar 2;320(9):565-70.
	Beta-agonist and beta-blocker Pathway	Propranolol	ASW_CEU		CEU	ASW	Antihypertensive effect greater in Whites		JAMA. 1982 Oct 22;248(16):1996-2003.
	Beta-agonist and beta-blocker Pathway	Propranolol	CHB_CEU		CHB	CEU	Chinese twice as sensitive to effects on blood pressure and heart rate		JAMA. 1982 Oct 22;248(16):1996-2003.

	Beta-agonist and beta-blocker Pathway	Propranolol	ASW_CEU	CEU	ASW		As evidenced by these results, propranolol is as efficacious as HCTZ in whites, but HCTZ is more effective than propranolol in African Americans [24]		JAMA. 1982 Oct 22;248(16):1996-2003.
	Celecoxib (Celebrex) Pathway	Acetaminophen / paracetamol	CEU_YRI	CEU	YRI		Recovery of mercapturic acid and cysteine conjugates of acetaminophen was found to be 9.3% in Caucasians compared With only 5.2% and 4.4% in Ghanaians and Kenyans, respectively	Critchley et al	Br J Clin Pharmacol. 1986 Dec;22(6):649-57.
	Celecoxib (Celebrex) Pathway	COX-2 Inhibitors (Celecoxib)	ASW_CEU	ASW	CEU		Approximately 40% higher AUC and 11% lower apparent clearance in Blacks compared to Whites	Davies et al	Clin Pharmacokinet. 2000 Mar;38(3):225-42.
Celecoxib	Celecoxib (Celebrex) Pathway	Caffeine	CHB_CEU				The 7-demethylation and 8-hydroxylation activities were found to be different between Asians and Caucasians	Grant et al	Clin Pharmacol Ther. 1983 May;33(5):591-602.
	Celecoxib (Celebrex) Pathway	NSAIDs (Ibuprofen)	ASW_CEU			ASW	Suggested that Black patients should be monitored for an increase in adverse effects of NSAIDs in the presence of histamine H_2 receptor antagonists	Small et al	Clin Pharm. 1989 Jul;8(7):471-2.
	Celecoxib (Celebrex) Pathway	Diclofenac	CHS_MAS				Ethnic distribution of ADR reports showed statistically significant difference with the general population		
	Celecoxib (Celebrex) Pathway	Diclofenac	CHS_INS				Ethnic distribution of ADR reports showed statistically significant difference with the general population		

	Celecoxib (Celebrex) Pathway	Diclofenac	MAS_INS			Ethnic distribution of ADR reports showed statistically significant difference with the general population			
	Celecoxib (Celebrex) Pathway	Naproxen	CHS_MAS			Ethnic distribution of ADR reports showed statistically significant difference with the general population			
	Celecoxib (Celebrex) Pathway	Naproxen	CHS_INS			Ethnic distribution of ADR reports showed statistically significant difference with the general population			
	Celecoxib (Celebrex) Pathway	Naproxen	MAS_INS			Ethnic distribution of ADR reports showed statistically significant difference with the general population			
	Codeine and Morphine Pathway (PK)	Opioid (codeine, morphine and pethidine)	CHB_CEU		CHB	Some differences in metabolism between Chinese and Caucasians.	Lee et al	Anaesth Intensive Care. 1997 Dec;25(6):665-70.	
	Codeine and Morphine Pathway (PK)	Codeine	CHB_CEU	CEU	CHB	Chinese were less able to metabolize codeine particularly by glucuronidation, compared with Caucasians	Yue et al	Br J Clin Pharmacol. 1991 Jun;31(6):643-7.	
Codeine and morphine	Codeine and Morphine Pathway (PK)	Codeine	CHB_JPT	JPT	CHB	Chinese metabolized codeine less effectively than Japanese and Koreans	Yue et al	Pharmacogenetics. 1995 Jun;5(3):173-7.	
	Codeine and Morphine Pathway (PK)	Codeine	CHB_JPT	JPT	CHB	Chinese metabolized codeine less effectively than Japanese and Koreans	Yue et al	Pharmacogenetics. 1995 Jun;5(3):173-7.	
	Codeine and Morphine Pathway (PK)	Morphine	CHB_CEU	CEU	CHB	Caucasians more sensitive to the cardiovascular and rerespiratory effects of morphine than Chinese. Chinese less sensitive	Zhou et al	Clin Pharmacol Ther. 1993 Nov;54(5):507-13.	

							to its gastrointestinal side effects		
	Codeine and Morphine Pathway (PK)	Morphine	CHB_CEU	CHB	CEU		Apparent clearance of morphine significantly higher in Chinese than in Caucasians	Zhou et al	Clin Pharmacol Ther. 1993 Nov;54(5):507-13.
	Cyclo-phosphamide Pathway	Cyclophosphamide	ASW_CEU	CEU	ASW		Renal survival was significantly worse in blacks compared with white patients.	Dooley et al	Kidney Int. 1997 Apr;51(4):1188-95.
Cyclophosphamide	Cyclo-phosphamide Pathway	Cyclophosphamide	JPT_CEU			JPT		Ma B	Radiother Oncol. 2002 Feb;62(2):185-9.
	Cyclo-phosphamide Pathway (PK)	Cyclophosphamide	ASW_CEU	CEU	ASW		Renal survival was significantly worse in blacks compared with white patients.	Dooley et al	Kidney Int. 1997 Apr;51(4):1188-95.
	Cyclo-phosphamide Pathway (PK)	cyclophosphamide	JPT_CEU			JPT		Ma B	Radiother Oncol. 2002 Feb;62(2):185-9.
	Doxorubicin Pathway	Doxorubicin	ASW_CEU			ASW	Increased cardiotoxicity in AA	Hershman et al	J Clin Oncol. 2005 Sep 20;23(27):6639-46.
Doxoru-bicin									
	Doxorubicin Pathway	Doxorubicin	ASW_CEU			ASW	Increased cardiotoxicity in AA	Hershman et al	J Clin Oncol. 2005 Sep 20;23(27):6639-46.
	EGFR Inhibitor Pathway (PD)	Erlotinib	JPT_CEU	JPT	CEU		Higher response to gefitinib and erlotinib in patients of Asian origin	Calvo et al	J Clin Oncol. 2006 May 10;24(14):2158-63.
EGFR inhibit	EGFR Inhibitor Pathway (PD)	Gefitinib	JPT_CEU	JPT	CEU		Higher response to gefitinib and erlotinib in patients of Asian origin	Calvo et al	J Clin Oncol. 2006 May 10;24(14):2158-63.
	EGFR	EGFR	JPT_CEU	JPT	CEU		The response rates to	Fukuoka et	J Clin Oncol.

	Inhibitor Pathway (PD)	antagonists (NSCLC, erlotinib)					treatment with these agents are higher in Asians than Caucasians, but highest in Asian females with tumours of adenocarcinoma histology who have never smoked.	al; Kris et al	2003 Jun 15;21(12):2237-46. Epub 2003 May 14; JAMA. 2003 Oct 22;290(16):2149-58
	EGFR Inhibitor Pathway (PD)	EGFR inhibitors	JPT_CEU	JPT	CEU			Thatcher et al	Lancet. 2005 Oct 29-Nov 4;366(9496):1527-37.
Erlotinib	Erlotinib Pathway (PK)	EGFR antagonists (NSCLC, erlotinib)	JPT_CEU				The response rates to treatment with these agents are higher in Asians than Caucasians, but highest in Asian females with tumours of adenocarcinoma histology who have never smoked.	Fukuoka et al; Kris et al	J Clin Oncol. 2003 Jun 15;21(12):2237-46. Epub 2003 May 14; JAMA. 2003 Oct 22;290(16):2149-58
Etoposide	Etoposide Pathway	Cisplatin and irinotecan/etoposide	JPT_CEU			JPT	Much greater toxicity is experienced by Japanese patients treated with cisplatin and either irinotecan or etoposide as Compared to similarly treated Caucasian patients	Phan et al	Expert Opin Drug Metab Toxicol. 2009 Mar;5(3):243-57.
Fluoropyrimidine	Fluoro-pyrimidine (PK)	5-fluorouracil	ASW_CEU			ASW	Hematologic toxicities more common in AA > CAU;	McCollum et al	J Natl Cancer Inst. 2002 Aug 7;94(15):1160-7.
	Fluoro-pyrimidine (PK)	5-fluorouracil	ASW_CEU			CEU	Overall incidence of any toxicity was actually lower in African Americans compared with	McCollum et al	J Natl Cancer Inst. 2002 Aug 7;94(15):1160-7.

							Caucasians ($P = 0.005$)		
	Fluoro-pyrimidine (PK)	irinotecan/fluorouracil, fluorouracil/oxaliplatin, or irinotecan/oxaliplatin	ASW_CEU			ASW	Response rate and adverse events vary considerably by race	Sanoff et al	J Clin Oncol. 2009 Sep 1;27(25):4109-15. Epub 2009 Jul 27.
	Fluoro-pyrimidine (PK)	irinotecan/fluorouracil, fluorouracil/oxaliplatin, or irinotecan/oxaliplatin	ASW_CEU			ASW	Response rate and adverse events vary considerably by race	Sanoff et al	J Clin Oncol. 2009 Sep 1;27(25):4109-15. Epub 2009 Jul 27.
	Fluoro-pyrimidine (PK)	irinotecan/fluorouracil, fluorouracil/oxaliplatin, or irinotecan/oxaliplatin	ASW_CEU			ASW	Response rate and adverse events vary considerably by race	Sanoff et al	J Clin Oncol. 2009 Sep 1;27(25):4109-15. Epub 2009 Jul 27.
	Fluoro-pyrimidine (PK)	irinotecan/fluorouracil, fluorouracil/oxaliplatin, or irinotecan/oxaliplatin	ASW_CEU			ASW	Response rate and adverse events vary considerably by race	Sanoff et al	J Clin Oncol. 2009 Sep 1;27(25):4109-15. Epub 2009 Jul 27.
Glucocorticoid and inflammatory	Glucocorticoid and Inflammatory genes Pathway (HPA axis)	Glucocorticoids	ASW_CEU			ASW	Race-dependent clinically significant adverse effects (steroid-associated diabetes)	Tornatore et al	Transplantation. 1995 Mar 15;59(5):729-36.
	Glucocorticoid and Inflammatory genes Pathway (HPA axis)	Glucocorticoids (methylprednisolone)	ASW_CEU	ASW	CEU		Almost 50% lower clearance in the Black population	Tornatore et al	Pharmacotherapy. 1993 Sep-Oct;13(5):481-6.
	Glucocorticoid and	Nicotine	ASW_MEX	ASW	CEU		Blacks have a higher exposure to nicotine	Caraballo et al	JAMA. 1998 Jul 8;280(2):135-9.

Inflammatory genes Pathway (PD)						and cotinine than Caucasians and Mexican Americans		
Glucocorticoid and Inflammatory genes Pathway (PD)	Nicotine	ASW_CEU	ASW	MEX		Blacks have a higher exposure to nicotine and cotinine than Caucasians and Mexican Americans	Caraballo et al	JAMA. 1998 Jul 8;280(2):135-9.
Glucocorticoid and Inflammatory genes Pathway (PD)	Glucocorticoids	ASW_CEU			ASW	Race-dependent clinically significant adverse effects (steroid-associated diabetes)	Tornatore et al	Transplantation. 1995 Mar 15;59(5):729-36.
Glucocorticoid and Inflammatory genes Pathway (PD)	Glucocorticoids (methylprednisolone)	ASW_CEU	ASW	CEU		Almost 50% lower clearance in the Black population	Tornatore et al	Pharmacotherapy. 1993 Sep-Oct;13(5):481-6.
Glucocorticoid and Inflammatory genes Pathway (Regulation)	Nicotine	ASW_MEX	ASW	CEU		Blacks have a higher exposure to nicotine and cotinine than Caucasians and Mexican Americans	Caraballo et al	JAMA. 1998 Jul 8;280(2):135-9.
Glucocorticoid and Inflammatory genes Pathway (Regulation)	Nicotine	ASW_CEU	ASW	MEX		Blacks have a higher exposure to nicotine and cotinine than Caucasians and Mexican Americans	Caraballo et al	JAMA. 1998 Jul 8;280(2):135-9.
Glucocorticoid and Inflammatory genes Pathway (Regulation)	Glucocorticoids	ASW_CEU			ASW	Race-dependent clinically significant adverse effects (steroid-associated diabetes)	Tornatore et al	Transplantation. 1995 Mar 15;59(5):729-36.
Glucocorticoid and Inflammatory genes Pathway	Glucocorticoids (methylprednisolone)	ASW_CEU	ASW	CEU		Almost 50% lower clearance in the Black population	Tornatore et al	Pharmacotherapy. 1993 Sep-Oct;13(5):481-6.

	(Regulation)						
Ifosfamide	Ifosfamide Pathway	Ifosfamide	JPT_CEU		CYP2C19 is involved in the metabolism of cyclophosphamide, ifosfamide, S-mephenytoin, R-warfarin and antidepressants, all of which are commonly used in cancer patients.	Phan et al	Expert Opin Drug Metab Toxicol. 2009 Mar;5(3):243-57.
	Ifosfamide Pathway (PK)	Ifosfamide	JPT_CEU		CYP2C19 is involved in the metabolism of cyclophosphamide, ifosfamide, S-mephenytoin, R-warfarin and antidepressants, all of which are commonly used in cancer patients.	Phan et al	Expert Opin Drug Metab Toxicol. 2009 Mar;5(3):243-57.
Imatinib	Imatinib pathway	Desatinib	Asians_CEU	Asian-CEU	Results from 55 Asian and 615 non-Asian patients demonstrated that the efficacy and safety of dasatinib was comparable. Dasatinib was well tolerated, with no observed toxicities exclusive to Asian patients. A higher incidence of adverse events and lower rate of response observed among Asian patients with myeloid blast phase CML reflected the aggressive nature of the disease.	Kim et al	

	Imipramine Desipramine Pathway (PK)	Clomipramine, doxepin, imipramine	Barbados and Saudi Arabia		Higher plasma levels for tricyclic antidepresantidepressants in patients from Barbados (clomipramine) or Saudi Arabia (doxepin or imipramine) in patients from Barbados (clomipramine)102 or	El Yazigi et al	Pharm Res 1987;4: S87
Imipramine desi-pramine	Imipramine Desipramine Pathway (PK)	Clomipramine, doxepin, imipramine	Barbados and Saudi Arabia		Higher plasma levels for tricyclic antidepresantidepressants in patients from Barbados (clomipramine) or Saudi Arabia (doxepin or imipramine) in patients from Barbados (clomipramine)102 or	Mahy GE	West Indian Med J. 1978 Jun;27(2):75-80.
	Imipramine Desipramine Pathway (PK)	Desipramine	CHB_CEU		Mean total clearance of DMI (CLDMI) from plasma was significantly (P less than 0.05) higher in the Caucasians (123 +/- 57 l/h) than in the Chinese (73.5 +/- 38.8 l/h).	Rudorfer et al	Br J Clin Pharmacol. 1984 April; 17(4): 433–440.
Irinotecan	Irinotecan Pathway	Cisplatin and irinotecan/etoposide	JPT_CEU	JPT	Much greater toxicity is experienced by Japanese patients treated with cisplatin and either irinotecan or etoposide as compared to similarly treated	Phan et al	Expert Opin Drug Metab Toxicol. 2009 Mar;5(3):243-57.

				Caucasian patients		
Irinotecan Pathway	Irinotecan/fluorouracil, fluorouracil/oxaliplatin, or irinotecan/oxaliplatin	ASW_CEU	ASW	Response rate and adverse events vary considerably by race	Sanoff et al	J Clin Oncol. 2009 Sep 1;27(25):4109-15. Epub 2009 Jul 27.
Irinotecan Pathway	Irinotecan/fluorouracil, fluorouracil/oxaliplatin, or irinotecan/oxaliplatin	ASW_CEU	ASW	Response rate and adverse events vary considerably by race	Sanoff et al	J Clin Oncol. 2009 Sep 1;27(25):4109-15. Epub 2009 Jul 27.
Irinotecan Pathway	Irinotecan/fluorouracil, fluorouracil/oxaliplatin, or irinotecan/oxaliplatin	ASW_CEU	ASW	Response rate and adverse events vary considerably by race	Sanoff et al	J Clin Oncol. 2009 Sep 1;27(25):4109-15. Epub 2009 Jul 27.
Irinotecan Pathway	Irinotecan/fluorouracil, fluorouracil/oxaliplatin, or irinotecan/oxaliplatin	ASW_CEU	ASW	Response rate and adverse events vary considerably by race	Sanoff et al	J Clin Oncol. 2009 Sep 1;27(25):4109-15. Epub 2009 Jul 27.
Irinotecan Pathway	Carboplatin, paclitaxel, cisplatin, irinotecan, etopopside	JPT_CEU	JPT	Neutropenia in patients receiving a combination of platinum and antimicrotubule agents may be more severe in Japanese than in Europeans and Americans	Sekine et al	Br J Cancer. 2008 Dec 2;99(11):1757-62. Epub 2008 Nov 4.
Irinotecan Pathway (Cancer)	Cisplatin and irinotecan/etoposide	JPT_CEU	JPT	much greater toxicity is experienced by Japanese patients treated with cisplatin	Phan et al	Expert Opin Drug Metab Toxicol. 2009 Mar;5(3):243-57.

				and either irinotecan or etoposide as compared to similarly treated Caucasian patients		
Irinotecan Pathway (Cancer)	Irinotecan/fluorouracil, fluorouracil/oxaliplatin, or irinotecan/oxaliplatin	ASW_CEU	ASW	Response rate and adverse events vary considerably by race	Sanoff et al	J Clin Oncol. 2009 Sep 1;27(25):4109-15. Epub 2009 Jul 27.
Irinotecan Pathway (Cancer)	Irinotecan/fluorouracil, fluorouracil/oxaliplatin, or irinotecan/oxaliplatin	ASW_CEU	ASW	Response rate and adverse events vary considerably by race	Sanoff et al	J Clin Oncol. 2009 Sep 1;27(25):4109-15. Epub 2009 Jul 27.
Irinotecan Pathway (Cancer)	Irinotecan/fluorouracil, fluorouracil/oxaliplatin, or irinotecan/oxaliplatin	ASW_CEU	ASW	Response rate and adverse events vary considerably by race	Sanoff et al	J Clin Oncol. 2009 Sep 1;27(25):4109-15. Epub 2009 Jul 27.
Irinotecan Pathway (Cancer)	Irinotecan/fluorouracil, fluorouracil/oxaliplatin, or irinotecan/oxaliplatin	ASW_CEU	ASW	Response rate and adverse events vary considerably by race	Sanoff et al	J Clin Oncol. 2009 Sep 1;27(25):4109-15. Epub 2009 Jul 27.
Irinotecan Pathway (Cancer)	Carboplatin, paclitaxel, cisplatin, irinotecan, etopopside	JPT_CEU	JPT	Neutropenia in patients receiving a combination of platinum and antimicrotubule agents may be more severe in Japanese than in Europeans and Americans	Sekine et al	Br J Cancer. 2008 Dec 2;99(11):1757-62. Epub 2008 Nov 4.
Irinotecan	Irinotecan	JPT_CEU	JPT	Different grade of	Sekine et al	Br J Cancer.

	Pathway (Cancer)						toxicity		2008 Dec 2;99(11):1757-62. Epub 2008 Nov 4.
Metho-trexate	Methotrexate	Methotrexate	ASW_CEU			ASW	The presence of the C allele at the exon 7 rs4846051 SNP was associated with a higher mean toxicity score among African-Americans than Caucasians (0.371 v 0.078, p=0.050). The presence of at least one copy of haplotype 4 (which contains the rs4846051 C allele) was also associated with a higher toxicity score among the African-Americans (p=0.03).	Hughes et al	Ann Rheum Dis. 2006 Sep;65(9):1213-8. Epub 2006 Jan 26.
	Methotrexate	Methotrexate	ASW_CEU				Methotrexate (MTX) pathway gene polymorphisms and their effects on MTX toxicity in Caucasian and African American patients with rheumatoid arthritis.	Ranganathan et al	J Rheumatol. 2008 Apr;35(4):572-9. Epub 2008 Mar 15.
Nicotine	Nicotine Pathway	Nicotine	ASW_MEX	ASW	CEU		Blacks have a higher exposure to nicotine and cotinine than Caucasians and Mexican Americans	Caraballo et al	JAMA. 1998 Jul 8;280(2):135-9.
	Nicotine Pathway	Nicotine	ASW_CEU	ASW	MEX		Blacks have a higher exposure to nicotine	Caraballo et al	JAMA. 1998 Jul 8;280(2):135-9.

					and cotinine than Caucasians and Mexican Americans		
Nicotine Pathway	Nicotine	ASW_CEU	ASW	CEU	Total clearance of cotinine, the fractional conversion of nicotine to cotinine, and the metabolic clearance of nicotine to cotinine were all significantly lower in Blacks than in Caucasians	Ethnic differences in N-glucuronidation of nicotine and cotinine.	J Pharmacol Exp Ther. 1999 Dec;291(3):1196-203.
Nicotine Pathway	Nicotine	JPT_ASW	ASW	JPT	Decrease nicotine-to-cotinine metabolism in Japanese compared to other pops	Nakajima et al	Clin Pharmacol Ther. 2006 Sep;80(3):282-97
Nicotine Pathway	Nicotine	JPT_CEU	CEU	JPT	Decrease nicotine-to-cotinine metabolism in Japanese compared to other pops	Nakajima et al	Clin Pharmacol Ther. 2006 Sep;80(3):282-97.
Nicotine Pathway	Nicotine	JPT_KOREAN	KOREAN	JPT	Decrease nicotine-to-cotinine metabolism in Japanese compared to other pops	Nakajima et al	Clin Pharmacol Ther. 2006 Sep;80(3):282-97
Nicotine PD Pathway (Chromaffin Cell)	Nicotine	ASW_MEX	ASW	CEU	Blacks have a higher exposure to nicotine and cotinine than Caucasians and Mexican Americans	Caraballo et al	JAMA. 1998 Jul 8;280(2):135-9.
Nicotine PD Pathway (Chromaffin Cell)	Nicotine	ASW_CEU	ASW	MEX	Blacks have a higher exposure to nicotine and cotinine than Caucasians and Mexican Americans	Caraballo et al	JAMA. 1998 Jul 8;280(2):135-9.
Nicotine PD Pathway (Chromaffin	Nicotine	ASW_CEU	ASW	CEU	Total clearance of cotinine, the fractional conversion of nicotine	Ethnic differences in N-	J Pharmacol Exp Ther. 1999 Dec;291(3):1196

Cell)					to cotinine, and the metabolic clearance of nicotine to cotinine were all significantly lower in Blacks than in Caucasians	glucuronidation of nicotine and cotinine.	-203.
Nicotine PD Pathway (Chromaffin Cell)	Nicotine	JPT_ASW	ASW	JPT	Decrease nicotine-to-cotinine metabolism in Japanese compared to other pops	Nakajima et al	Clin Pharmacol Ther. 2006 Sep;80(3):282-97
Nicotine PD Pathway (Chromaffin Cell)	Nicotine	JPT_CEU	CEU	JPT	Decrease nicotine-to-cotinine metabolism in Japanese compared to other pops	Nakajima et al	Clin Pharmacol Ther. 2006 Sep;80(3):282-97
Nicotine PD Pathway (Chromaffin Cell)	Nicotine	JPT_KOREAN	KOREAN	JPT	Decrease nicotine-to-cotinine metabolism in Japanese compared to other pops	Nakajima et al	Clin Pharmacol Ther. 2006 Sep;80(3):282-97
Nicotine PD Pathway (Dopaminergic Neuron)	Nicotine	ASW_MEX	ASW	CEU	Blacks have a higher exposure to nicotine and cotinine than Caucasians and Mexican Americans	Caraballo et al	JAMA. 1998 Jul 8;280(2):135-9.
Nicotine PD Pathway (Dopaminergic Neuron)	Nicotine	ASW_CEU	ASW	MEX	Blacks have a higher exposure to nicotine and cotinine than Caucasians and Mexican Americans	Caraballo et al	JAMA. 1998 Jul 8;280(2):135-9.
Nicotine PD Pathway (Dopaminergic Neuron)	Nicotine	ASW_CEU	ASW	CEU	Total clearance of cotinine, the fractional conversion of nicotine to cotinine, and the metabolic clearance of nicotine to cotinine were all significantly lower in Blacks than in Caucasians	Ethnic differences in N-glucuronidation of nicotine and cotinine.	J Pharmacol Exp Ther. 1999 Dec;291(3):1196-203.

	Nicotine PD Pathway (Dopaminergic Neuron)	Nicotine	JPT_ASW	ASW	JPT		Decrease nicotine-to-cotinine metabolism in Japanese compared to other pops	Nakajima et al	Clin Pharmacol Ther. 2006 Sep;80(3):282-97
	Nicotine PD Pathway (Dopaminergic Neuron)	Nicotine	JPT_CEU	CEU	JPT		Decrease nicotine-to-cotinine metabolism in Japanese compared to other pops	Nakajima et al	Clin Pharmacol Ther. 2006 Sep;80(3):282-97
	Nicotine PD Pathway (Dopaminergic Neuron)	Nicotine	JPT_KOREAN	KOREAN	JPT		Decrease nicotine-to-cotinine metabolism in Japanese compared to other pops	Nakajima et al	Clin Pharmacol Ther. 2006 Sep;80(3):282-97
	Phenytoin PK Pathway	Phenytoin	JPT_CEU	JPT	CEU		Approximately four- to fivefold more Asians are "slow" phenytoin metabolizers	Bertilsson et al	Clin Pharmacol Ther. 1993 May;53(5):608-10.
Phenytoin	Phenytoin PK Pathway	Phenytoin	ASW_CEU	ASW	CEU		Phenytoin metabolism is slowed in Black as compared to Caucasian individuals	Edeki et al	Drug Metab Rev. 1995;27(3):449-69.
Platelet aggregation	Platelet Aggregation Pathway (PD)	Calcium channel blockers	ASW_CEU	ASW	CEU		Different reduction of blood pressure	Brewster et al	Ann Intern Med. 2004 Oct 19;141(8):614-27.
	Platinum Pathway	Cisplatin and irinotecan/etoposide	JPT_CEU			JPT		Phan et al	Expert Opin Drug Metab Toxicol. 2009 Mar;5(3):243-57.
Platinum	Platinum Pathway	Carboplatin and paclitaxel	JPT_CEU			JPT	To support this approach, they have published data comparing toxicity and response for similar chemotherapy regimens used in Asian and Caucasian cohorts	Sekine et al	Lung Cancer. 2006 Aug;53(2):157-64. Epub 2006 Jun 15.
	Platinum	Carboplatin,	JPT_CEU			JPT	Neutropenia in patients	Sekine et al	Br J Cancer.

	Pathway	paclitaxel, cisplatin, irinotecan, etopopside				receiving a combination of platinum and antimicrotubule agents may be more severe in Japanese than in Europeans and Americans		2008 Dec 2;99(11):1757-62. Epub 2008 Nov 4.
	Platinum Pathway	Paclitaxel	JPT_CEU	JPT	CEU	Paclitaxel sensitivity in cancer cells	Kwon et al	Cancer Lett. 2009 May 18;277(2):155-63. Epub 2009 Jan 12.
Proton pump inhibitor	Proton Pump Inhibitor (PD)	Omeprazole	CHB_CEU	CHB	CEU	The AUC values noted in the Chinese group were significantly higher than those in the Caucasian group	Caraco et al	Clin Pharmacol Ther. 1996 Aug;60(2):157-67.
	Proton Pump Inhibitor (PK)	Omeprazole	CHB_CEU	CHB	CEU	The AUC values noted in the Chinese group were significantly higher than those in the Caucasian group	Caraco et al	Clin Pharmacol Ther. 1996 Aug;60(2):157-67.
RAAS	Renin-Angiotensin-Aldosterone-System-acting Drug Pathway	Enalapril	ASW_CEU	CEU	ASW	Hypertension and hospitalization for heart failure reduced in Whites but not in Blacks with left ventricular dysfunction	Exner et al	N Engl J Med. 2001 May 3;344(18):1351-7.
	Renin-Angiotensin-Aldosterone-System-acting Drug Pathway	Captopril/Candesartan	ASW_CEU	CEU	ASW	White individuals demonstrated a strong, significant correlation between responses to these drugs ($r = 0.78$, $P = 0.008$) and a significantly greater increase in the renal plasma flow in response	Forman et al	

						to candesartan compared with captopril		
	Renin-Angiotensin-Aldosterone-System-acting Drug Pathway	Enalapril	CHS_MAS			Ethnic distribution of ADR reports showed statistically significant difference with the general population		
	Renin-Angiotensin-Aldosterone-System-acting Drug Pathway	Enalapril	CHS_INS			Ethnic distribution of ADR reports showed statistically significant difference with the general population		
	Renin-Angiotensin-Aldosterone-System-acting Drug Pathway	Enalapril	MAS_INS			Ethnic distribution of ADR reports showed statistically significant difference with the general population		
SSRI	Selective Serotonin Reuptake Inhibitors (SSRI) Pathway	Fluoxetine	Latinos-ASW-CEU	CEU	ASW	Attrition was greater among Latinos than either blacks or whites. Black patients were more likely than whites to be nonresponders to fluoxetine. Latinos were more likely to respond to placebo compared with blacks and whites.	Wagner et al	Psychiatr Serv 49:239-240, February 1998
Statin	Statin Pathway (Atorvastatin Lovastatin and Simvastatin PK)	Statins	CHB_CEU	CHB	CEU	Higher plasma levels of statins in Asians compared with Caucasians	Liao JK	Am J Cardiol. 2007 Feb 1;99(3):410-4. Epub 2006 Dec 15.
	Statin Pathway (Atorvastatin Lovastatin and	Statins	JPT_CEU	JPT	CEU	Higher plasma levels of statins in Asians compared with	Liao JK	Am J Cardiol. 2007 Feb 1;99(3):410-4.

Simvastatin PK)					Caucasians		Epub 2006 Dec 15.
Statin Pathway (Cholesterol and Lipoprotein Transport PD)	Statins	CHB_CEU	CHB	CEU	Higher plasma levels of statins in Asians compared with Caucasians	Liao JK	Am J Cardiol. 2007 Feb 1;99(3):410-4. Epub 2006 Dec 15.
Statin Pathway (Cholesterol and Lipoprotein Transport PD)	Statins	JPT_CEU	JPT	CEU	Higher plasma levels of statins in Asians compared with Caucasians	Liao JK	Am J Cardiol. 2007 Feb 1;99(3):410-4. Epub 2006 Dec 15.
Statin Pathway (Fluvastatin PK)	Statins	CHB_CEU	CHB	CEU	Higher plasma levels of statins in Asians compared with Caucasians	Liao JK	Am J Cardiol. 2007 Feb 1;99(3):410-4. Epub 2006 Dec 15.
Statin Pathway (Fluvastatin PK)	Statins	JPT_CEU	JPT	CEU	Higher plasma levels of statins in Asians compared with Caucasians	Liao JK	Am J Cardiol. 2007 Feb 1;99(3):410-4. Epub 2006 Dec 15.
Statin Pathway (PK)	Statins	CHB_CEU	CHB	CEU	Higher plasma levels of statins in Asians compared with Caucasians	Liao JK	Am J Cardiol. 2007 Feb 1;99(3):410-4. Epub 2006 Dec 15.
Statin Pathway (PK)	Statins	JPT_CEU	JPT	CEU	Higher plasma levels of statins in Asians compared with Caucasians	Liao JK	Am J Cardiol. 2007 Feb 1;99(3):410-4. Epub 2006 Dec 15.
Statin Pathway (Pravastatin PK)	Statins	CHB_CEU	CHB	CEU	Higher plasma levels of statins in Asians compared with Caucasians	Liao JK	Am J Cardiol. 2007 Feb 1;99(3):410-4. Epub 2006 Dec 15.

	Statin Pathway (Pravastatin PK)	Statins	JPT_CEU	JPT	CEU	Higher plasma levels of statins in Asians compared with Caucasians	Liao JK	Am J Cardiol. 2007 Feb 1;99(3):410-4. Epub 2006 Dec 15.
	Statin Pathway (Rosuvastatin PK)	Statins	CHB_CEU	CHB	CEU	Higher plasma levels of statins in Asians compared with Caucasians	Liao JK	Am J Cardiol. 2007 Feb 1;99(3):410-4. Epub 2006 Dec 15.
	Statin Pathway (Rosuvastatin PK)	Statins	JPT_CEU	JPT	CEU	Higher plasma levels of statins in Asians compared with Caucasians	Liao JK	Am J Cardiol. 2007 Feb 1;99(3):410-4. Epub 2006 Dec 15.
	Sympathetic Nerve Pathway (Neuroeffector Junction)	Isoproterenol	ASW_CEU	CEU	ASW	Vasodilation response to isoproterenol markedly lower in Blacks	Johnson et al	J Cardiovasc Pharmacol. 1995 Jan;25(1):90-6.
	Sympathetic Nerve Pathway (Neuroeffector Junction)	Propranolol	MAS_CHS	MAS	CHS	Malays to have significantly greater propranolol responses compared to Chinese healthy male subjects	Rasool et al	Int J Clin Pharmacol Ther. 2000 May;38(5):260-9.
Sympa-thetic nerve	Sympathetic Nerve Pathway (Neuroeffector Junction)	Nitroglycerin	ASW_CEU	CEU	ASW	Lower transdermal availability of nitroglycerin in 4 Black subjects as opposed to 12 Caucasian and Asian subjects	Williams et al	Pharm Res. 1991 Jun;8(6):744-9.
	Sympathetic Nerve Pathway (Neuroeffector Junction)	Nitroglycerin	ASW_CEU	CHB	ASW	Lower transdermal availability of nitroglycerin in 4 Black subjects as opposed to 12 Caucasian and Asian subjects	Williams et al	Pharm Res. 1991 Jun;8(6):744-9.

Sympathetic Nerve Pathway (Neuroeffector Junction)	Nitroglycerin	ASW_CEU	JPT	ASW	Lower transdermal availability of nitroglycerin in 4 Black subjects as opposed to 12 Caucasian and Asian subjects	Williams et al	Pharm Res. 1991 Jun;8(6):744-9.
Sympathetic Nerve Pathway (Neuroeffector Junction)	Propranolol	CHB_CEU	CHB	CEU	Chinese men had a 2- to 3-fold greater sensitivity to propranolol's effect on heart rate and a 10-fold greater sensitivity to the BP effect than Caucasian men.	Zhou et al	N Engl J Med. 1989 Mar 2;320(9):565-70.
Sympathetic Nerve Pathway (Neuroeffector Junction)	Beta-blockers	ASW_CEU	ASW	CEU	Different reduction of blood pressure		JAMA. 1982 Oct 22;248(16):2004 -11.
Sympathetic Nerve Pathway (Neuroeffector Junction)	Beta-blockers	ASW_CEU	ASW	CEU	Different reduction of blood pressure		JAMA. 1982 Oct 22;248(16):2004 -11.
Sympathetic Nerve Pathway (Neuroeffector Junction)	Propranolol	ASW_CEU	CEU	ASW	Antihypertensive effect greater in Whites		JAMA. 1982 Oct 22;248(16):1996 -2003.
Sympathetic Nerve Pathway (Neuroeffector Junction)	Propranolol	CHB_CEU	CHB	CEU	Chinese twice as sensitive to effects on blood pressure and heart rate		JAMA. 1982 Oct 22;248(16):1996 -2003.
Sympathetic Nerve Pathway (Pre-	Nicotine	ASW_MEX	ASW	CEU	Blacks have a higher exposure to nicotine and cotinine than	Caraballo et al	JAMA. 1998 Jul 8;280(2):135-9.

	and Post-Ganglionic Junction)					Caucasians and Mexican Americans		
	Sympathetic Nerve Pathway (Pre- and Post-Ganglionic Junction)	Nicotine	ASW_CEU	ASW	MEX	Blacks have a higher exposure to nicotine and cotinine than Caucasians and Mexican Americans	Caraballo et al	JAMA. 1998 Jul 8;280(2):135-9.
	Sympathetic Nerve Pathway (Pre- and Post-Ganglionic Junction)	Nicotine	ASW_CEU	ASW	CEU	Total clearance of cotinine, the fractional conversion of nicotine to cotinine, and the metabolic clearance of nicotine to cotinine were all significantly lower in Blacks than in Caucasians	Ethnic differences in N-glucuronidation of nicotine and cotinine.	J Pharmacol Exp Ther. 1999 Dec;291(3):1196-203.
	Sympathetic Nerve Pathway (Pre- and Post-Ganglionic Junction)	Nicotine	JPT_ASW	ASW	JPT	Decrease nicotine-to-cotinine metabolism in Japanese compared to other pops	Nakajima et al	Clin Pharmacol Ther. 2006 Sep;80(3):282-97
	Sympathetic Nerve Pathway (Pre- and Post-Ganglionic Junction)	Nicotine	JPT_CEU	CEU	JPT	Decrease nicotine-to-cotinine metabolism in Japanese compared to other pops	Nakajima et al	Clin Pharmacol Ther. 2006 Sep;80(3):282-97
	Sympathetic Nerve Pathway (Pre- and Post-Ganglionic Junction)	Nicotine	JPT_KOREAN	KOREAN	JPT	Decrease nicotine-to-cotinine metabolism in Japanese compared to other pops	Nakajima et al	Clin Pharmacol Ther. 2006 Sep;80(3):282-97
Taxane	Taxane	Docetaxel	ASW_CEU	CEU-		Docetaxel clearance	Lewis et al	Clin Cancer Res

	Pathway			ASW			and its associated myelosuppression were similar in African-American and Caucasian cancer patients.		June 1, 2007 13; 3302
	Taxane Pathway	Carboplatin and paclitaxel	JPT_CEU			JPT	To support this approach, they have published data comparing toxicity and response for similar chemotherapy regimens used in Asian and Caucasian cohorts	Sekine et al	Lung Cancer. 2006 Aug;53(2):157-64. Epub 2006 Jun 15.
	Taxane Pathway	Carboplatin, paclitaxel, cisplatin, irinotecan, etopopside	JPT_CEU			JPT	Neutropenia in patients receiving a combination of platinum and antimicrotubule agents may be more severe in Japanese than in Europeans and Americans	Sekine et al	Br J Cancer. 2008 Dec 2;99(11):1757-62. Epub 2008 Nov 4.
	Taxane Pathway	Paclitaxel	JPT_CEU		JPT	CEU	paclitaxel sensitivity in cancer cells	Kwon et al	Cancer Lett. 2009 May 18;277(2):155-63. Epub 2009 Jan 12.
Tenofovir adefovir	Tenofovir Adefovir pathway	Adefovir dipivoxil	CHB_CEU	CHB-CEU			There were no significant differences in treatment response between Asians and Caucasians. Adefovir dipivoxil was well tolerated and no resistance developed up to week 48 in both racial groups	Lim et al	

Thiopurine	Thiopurine Pathway	Azathioprine	ASW_CEU	ASW	CEU		Racial differences are likely to be a minor determinant in concentrations; however, a greater percentage of Japanese and African-Americans display increased TPMT activity	Chocair et al	Q J Med. 1993 Jun;86(6):359-63.
	Thiopurine Pathway	Azathioprine	JPT_CEU	JPT	CEU		Racial differences are likely to be a minor determinant in concentrations; however, a greater percentage of Japanese and African-Americans display increased TPMT activity	Chocair et al	Q J Med. 1993 Jun;86(6):359-63.
Vinca alkaloids	Vinca Alkaloids (PK)	Vincristine	ASW_CEU			CEU	Neurotoxicity more common in CAU compared with AA	Renbarger et al	Pediatr Blood Cancer. 2008 Apr;50(4):769-71.
Warfarin	Warfarin Pathway (PD)	Warfarin	Afro_Caribbeans, non_Far East Asians and Caucasians				Greater daily warfarin dose requirements in Afro-Caribbeans compared with non–Far East Asians and Caucasians	Blann et al	Br J Haematol. 1999 Oct;107(1):207-9.
	Warfarin Pathway (PD)	Warfarin	ASW_CEU	CEU	ASW		Differences in dose required (Asian<Hispanic<White<African American)	Dang et al	Ann Pharmacother. 2005 Jun;39(6):1008-12. Epub 2005 Apr 26.
	Warfarin Pathway (PD)	Warfarin	ASW_CEU	CEU	ASW		Differences in dose required (Asian<Hispanic<Whit	Dang et al	Ann Pharmacother. 2005

						e<African American)		Jun;39(6):1008-12. Epub 2005 Apr 26.
	Warfarin Pathway (PD)	Warfarin	CHB_MEX	CHB	MEX	Differences in dose required (Asian<Hispanic<White<African American)	Dang et al	Ann Pharmacother. 2005 Jun;39(6):1008-12. Epub 2005 Apr 26.
	Warfarin Pathway (PD)	Warfarin	CHB_MEX	CHB	MEX	Differences in dose required (Asian<Hispanic<White<African American)	Dang et al	Ann Pharmacother. 2005 Jun;39(6):1008-12. Epub 2005 Apr 26.
	Warfarin Pathway (PD)	Warfarin	CEU_MEX	MEX	CEU	Differences in dose required (Asian<Hispanic<White<African American)	Dang et al	Ann Pharmacother. 2005 Jun;39(6):1008-12. Epub 2005 Apr 26.
	Warfarin Pathway (PD)	Warfarin	CEU_MEX	MEX	CEU	Differences in dose required (Asian<Hispanic<White<African American)	Dang et al	Ann Pharmacother. 2005 Jun;39(6):1008-12. Epub 2005 Apr 26.
	Warfarin Pathway (PD)	Warfarin	CHB_CEU	CHB	CEU	Asians might require a lower INR for protection from thromboembolism and might be at increased risk of bleeding at lower INRs	Johnson JA.	Circulation. 2008 Sep 23;118(13):1383-93.
	Warfarin Pathway (PD)	R-warfarin	JPT_CEU	JPT	CEU	CYP2C19 is involved in the metabolism of	Phan et al	Expert Opin Drug Metab

						cyclophosphamide, ifosfamide, S-mephenytoin, R-warfarin and antidepressants, all of which are commonly used in cancer patients.		Toxicol. 2009 Mar;5(3):243-57.
	Warfarin Pathway (PD)	Warfarin	CHB_CEU	CHB	CEU	Chinese patients have been reported to have lower daily warfarin dose requirements than what has been reported for similarly anticoagulated patients in the United States and the United Kingdom	Yu et al	Q J Med 1996;89: 127-135
	Warfarin Pathway (PK)	Warfarin	Afro_Caribbeans, non_Far East Asians and Caucasians			Greater daily warfarin dose requirements in Afro-Caribbeans compared with non–Far East Asians and Caucasians	Blann et al	Br J Haematol. 1999 Oct;107(1):207-9.
	Warfarin Pathway (PK)	Warfarin	ASW_CEU	CEU	ASW	Differences in dose required (Asian<Hispanic<White<African American)	Dang et al	Ann Pharmacother. 2005 Jun;39(6):1008-12. Epub 2005 Apr 26.
	Warfarin Pathway (PK)	Warfarin	ASW_CEU	CEU	ASW	Differences in dose required (Asian<Hispanic<White<African American)	Dang et al	Ann Pharmacother. 2005 Jun;39(6):1008-12. Epub 2005 Apr 26.
	Warfarin Pathway (PK)	Warfarin	CHB_MEX	CHB	MEX	Differences in dose required (Asian<Hispanic<Whit	Dang et al	Ann Pharmacother. 2005

					e<African American)		Jun;39(6):1008-12. Epub 2005 Apr 26.
	Warfarin Pathway (PK)	Warfarin	CHB_MEX	CHB	MEX	Differences in dose required (Asian<Hispanic<White<African American)	Dang et al	Ann Pharmacother. 2005 Jun;39(6):1008-12. Epub 2005 Apr 26.
	Warfarin Pathway (PK)	Warfarin	CEU_MEX	MEX	CEU	Differences in dose required (Asian<Hispanic<White<African American)	Dang et al	Ann Pharmacother. 2005 Jun;39(6):1008-12. Epub 2005 Apr 26.
	Warfarin Pathway (PK)	Warfarin	CEU_MEX	MEX	CEU	Differences in dose required (Asian<Hispanic<White<African American)	Dang et al	Ann Pharmacother. 2005 Jun;39(6):1008-12. Epub 2005 Apr 26.
	Warfarin Pathway (PK)	Warfarin	CHB_CEU	CHB	CEU	Asians might require a lower INR for protection from thromboembolism and might be at increased risk of bleeding at lower INRs	Johnson JA.	Circulation. 2008 Sep 23;118(13):1383-93.
	Warfarin Pathway (PK)	R-warfarin	JPT_CEU	JPT	CEU	CYP2C19 is involved in the metabolism of cyclophosphamide, ifosfamide, S-mephenytoin, R-warfarin and antidepressants, all of which are commonly used in cancer patients.	Phan et al	Expert Opin Drug Metab Toxicol. 2009 Mar;5(3):243-57.

	Warfarin Pathway (PK)	Warfarin	CHB_CEU	CHB	CEU	Chinese patients have been reported to have lower daily warfarin dose requirements than what has been reported for similarly anticoagulated patients in the United States and the United Kingdom	Yu et al	Q J Med 1996;89: 127-135

Appendix 7. Functional gene region distribution of chromosome 6 tcd SNPs in the CEU-GBR population pair

Gene	Promoter	5' UTR	3' UTR	3' DR	Exon										Intron													NA	Grand Total
					1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10	11	13	14	19		
ATAT1				**1**																**3**									**4**
NM_001031722.2																				1									1
NM_001254952.1																				1									1
NM_014046.3				1																									1
NM_024909.2																				1									1
C6orf136	**3**	**1**		**11**												**2**													**17**
NM_001031722.2				3																									3
NM_001109938.2	1															1													2
NM_001161376.1	1															1													2
NM_001164239.1				3																									3
NM_001254952.1				1																									1
NM_003587.4				3																									3
NM_024909.2				1																									1
NM_145029.3	1	1																											2
C6orf15	**8**			**9**	**3**																								**20**
NM_001264.4				8																									8
NM_014068.2	8																												8
NM_014070.2				1	3																								4
CCHCR1																							**3**						**3**
NM_001105563.1																							1						1

NM_001105564.1														1				1
NM_019052.3														1				1
CDSN	**37**	**12**	**18**	**1**		**10**	**30**											**108**
NM_001264.4		12	18	1		10												41
NM_014068.2	2						30											32
NM_014070.2	35																	35
DDR1	**18**							**3**	**1**	**1**	**1**			**2**	**2**	**2**		**30**
NM_001202521.1	2							1						1				4
NM_001202522.1	2							1						1				4
NM_001202523.1	4								1						1			6
NM_001954.4	4										1					1		6
NM_013993.2	4									1						1		6
NM_013994.2	2							1							1			4
DHX16	**4**			**9**								**2**					**2**	**17**
NM_001109938.2				1														1
NM_001134870.1				3														3
NM_001161376.1				1														1
NM_001164239.1	2											1					1	4
NM_003587.4	2											1					1	4
NM_133471.3				3														3
NM_145029.3				1														1
DPCR1	**3**	**2**	**1**		**2**		**4**											**12**
NM_080870.3	3	2	1		2		4											12
FLOT1									**5**				**15**					**20**
NM_005803.2									5				15					20
GTF2H4	**1**			**6**														**7**
NM_001202521.1				1														1
NM_001202522.1				1														1
NM_001202523.1				1														1

NM_001517.4	1																												1
NM_001954.4				1																									1
NM_013993.2				1																									1
NM_013994.2				1																									1
HCG18	**27**																												**27**
NM_001199119.1	3																												3
NM_021253.3	12																												12
NM_172016.2	12																												12
HLA-C	**16**		**2**	**6**			**6**	**10**	**4**		**4**					**9**	**12**	**14**	**27**	**6**	**6**							**1**	**123**
NM_001243042.1	8		1	2			3	5	2		2					4	6	7	13	3	3								59
NM_002117.5	8		1	4			3	5	2		2					5	6	7	14	3	3							1	64
MDC1	**2**			**1**											**1**											**1**			**5**
NM_014641.2				1											1											1			3
NM_178014.2	2																												2
MRPS18B	**1**																	**1**											**2**
NM_002714.3	1																												1
NM_014046.3																		1											1
MUC21	**1**		**3**			**1**																							**5**
NM_001010909.2			3			1																							4
NM_001198815.1	1																												1
MUC22															**1**	**6**													**7**
NM_001198815.1															1	6													7
POU5F1				**4**											**1**	**2**													**7**
NM_001077511.1				2																									2
NM_001173531.1																1													1
NM_002701.4																1													1
NM_007109.2				2																									2
NM_203289.4															1														1
PPP1R10																**1**						**1**							**2**

NM_002714.3															1						1								2
PPP1R11	**2**			**14**																									**16**
NM_001278785.1				2																									2
NM_001278786.1				2																									2
NM_014596.5				2																									2
NM_021959.2	2																												2
NM_025236.3				3																									3
NM_170769.2				3																									3
NM_170783.3				2																									2
PPP1R18				**15**	**3**									**3**															**21**
NM_001134870.1														3															3
NM_001270707.1				3																									3
NM_001270708.1				3																									3
NM_001270709.1				3																									3
NM_001270710.1				3																									3
NM_007243.2				3																									3
NM_133471.3					3																								3
PRR3	**1**														**1**	**1**													**3**
NM_001077497.2															1														1
NM_005275.3	1																												1
NM_025263.3																1													1
PSORS1C1	**1**		**1**	**18**												**4**													**24**
NM_001105563.1				5																									5
NM_001105564.1				5																									5
NM_014068.2			1													4													5
NM_014069.2	1			3																									4
NM_019052.3				5																									5
PTMAP1																			**9**										**9**
NM_001031722.2																			3										3

NM_001254952.1																				3										3
NM_024909.2																				3										3
RNF39				**2**											**2**															**4**
NM_021959.2				2																										2
NM_025236.3															1															1
NM_170769.2															1															1
RPP21	**6**		**7**	**40**					**14**			**1**		**5**			**6**	**3**				**2**	**1**							**85**
NM_001199119.1			1	1								1		5								2	1							11
NM_001199120.1	2		1	1					7								2	1												14
NM_001199121.1	2		4	1													2	1												10
NM_021253.3				18																										18
NM_024839.2	2		1	1					7								2	1												14
NM_172016.2				18																										18
SFTA2	**5**																													**5**
NM_205854.2	5																													5
TRIM10	**6**	**2**	**26**	**21**																										**55**
NM_006778.3		1	13	1																										15
NM_033229.2	6																													6
NM_052828.2		1	13	1																										15
NM_138700.3				19																										19
TRIM15	**12**														**1**				**1**											**14**
NM_006778.3	6																													6
NM_033229.2															1				1											2
NM_052828.2	6																													6
TRIM26		**6**	**2**	**2**				**4**							**2**	**6**	**4**	**1**	**11**	**10**										**48**
NM_001242783.1		6	1	1											1	4		1	10											24
NM_003449.4			1	1				4							1	2	4		1	10										24
TRIM31																			**3**	**7**										**10**
NM_007028.3																			3	7										10

TRIM39	**31**	**4**		**6**						**1**		**1**	**1**				**3**		**6**	**4**										**57**
NM_001199119.1	8									1							3			4										16
NM_001199120.1	5																													5
NM_001199121.1	5																													5
NM_021253.3	4	2		3									1						3											13
NM_024839.2	5																													5
NM_172016.2	4	2		3								1							3											13
TRIM39-RPP21	**12**		**3**	**16**					**14**	**1**		**2**	**1**	**5**																**54**
NM_001199119.1										1		1		5																7
NM_001199120.1	4								7																					11
NM_001199121.1	4		3																											7
NM_021253.3				8									1																	9
NM_024839.2	4								7																					11
NM_172016.2				8								1																		9
TRIM40	**6**	**29**	**1**	**22**					**1**						**8**	**2**		**1**												**70**
NM_006778.3				10																										10
NM_052828.2				10																										10
NM_138700.3	6	29	1	2					1						8	2		1												50
VARS2		**2**		**2**		**4**																				**1**	**2**			**11**
NM_001167733.1		2																								1				3
NM_001167734.1						2																					1			3
NM_001517.4				2																										2
NM_020442.4						2																					1			3
ZNRD1	**2**																**1**	**3**												**6**
NM_001278785.1																		1												1
NM_001278786.1																		1												1
NM_014596.5																		1												1
NM_021959.2	2																													2
NM_170783.3																	1													1

License Agreement from Springer

SPRINGER LICENSE
TERMS AND CONDITIONS

Jul 31, 2014

This is a License Agreement between Maulana Bachtiar ("You") and Springer ("Springer") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Springer, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3439401209977
License date	Jul 31, 2014
Licensed content publisher	Springer
Licensed content publication	Current Genetic Medicine Reports
Licensed content title	Genetics of Population Differences in Drug Response
Licensed content author	Maulana Bachtiar
Licensed content date	Jan 1, 2013
Volume number	1
Issue number	3
Type of Use	Thesis/Dissertation
Portion	Full text
Number of copies	10
Author of this Springer article	Yes and you are a contributor of the new work
Order reference number	None
Title of your thesis / dissertation	IDENTIFICATION OF 'PHARMA-SNPS' FOR PREDICTING RESPONSE TO DRUG THERAPIES
Expected completion date	Aug 2014
Estimated size(pages)	300
Total	0.00 USD
Terms and Conditions	

Introduction

The publisher for this copyrighted material is Springer Science + Business Media. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

Limited License

With reference to your request to reprint in your thesis material on which Springer Science and Business Media control the copyright, permission is granted, free of charge, for the use indicated in your enquiry.

Licenses are for one-time use only with a maximum distribution equal to the number that you identified in the licensing process.

This License includes use in an electronic form, provided its password protected or on the university's intranet or repository, including UMI (according to the definition at the Sherpa website: http://www.sherpa.ac.uk/romeo/). For any other electronic use, please contact Springer at (permissions.dordrecht@springer.com or permissions.heidelberg@springer.com).

The material can only be used for the purpose of defending your thesis limited to university-use only. If the thesis is going to be published, permission needs to be re-obtained (selecting "book/textbook" as the type of use).

Although Springer holds copyright to the material and is entitled to negotiate on rights, this license is only valid, subject to a courtesy information to the author (address is given with the article/chapter) and provided it concerns original material which does not carry references to other sources (if material in question appears with credit to another source, authorization from that source is required as well).

Permission free of charge on this occasion does not prejudice any rights we might have to charge for reproduction of our copyrighted material in the future.

Altering/Modifying Material: Not Permitted
You may not alter or modify the material in any manner. Abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of the author(s) and/or Springer Science + Business Media. (Please contact Springer at (permissions.dordrecht@springer.com or permissions.heidelberg@springer.com)

Reservation of Rights
Springer Science + Business Media reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

Copyright Notice:Disclaimer
You must include the following copyright and permission notice in connection with any reproduction of the licensed material: "Springer and the original publisher /journal title, volume, year of publication, page, chapter/article title, name(s) of author(s), figure number(s), original copyright notice) is given to the publication in which the material was originally published, by adding; with kind permission from Springer Science and Business Media"

Warranties: None

Example 1: Springer Science + Business Media makes no representations or warranties with respect to the licensed material.

Example 2: Springer Science + Business Media makes no representations or warranties with respect to the licensed material and adopts on its own behalf the limitations and disclaimers established by CCC on its behalf in its Billing and Payment terms and conditions for this licensing transaction.

Indemnity
You hereby indemnify and agree to hold harmless Springer Science + Business Media and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

No Transfer of License
This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without Springer Science + Business Media's written permission.

No Amendment Except in Writing
This license may not be amended except in a writing signed by both parties (or, in the case of Springer Science + Business Media, by CCC on Springer Science + Business Media's behalf).

Objection to Contrary Terms
Springer Science + Business Media hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and Springer Science + Business Media (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

Jurisdiction
All disputes that may arise in connection with this present License, or the breach thereof, shall be settled exclusively by arbitration, to be held in The Netherlands, in accordance with Dutch law, and to be conducted under the Rules of the 'Netherlands Arbitrage Instituut' (Netherlands Institute of Arbitration).*OR:*

All disputes that may arise in connection with this present License, or the breach thereof, shall be settled exclusively by arbitration, to be held in the Federal Republic of Germany, in accordance with German law.

Other terms and conditions:

v1.3

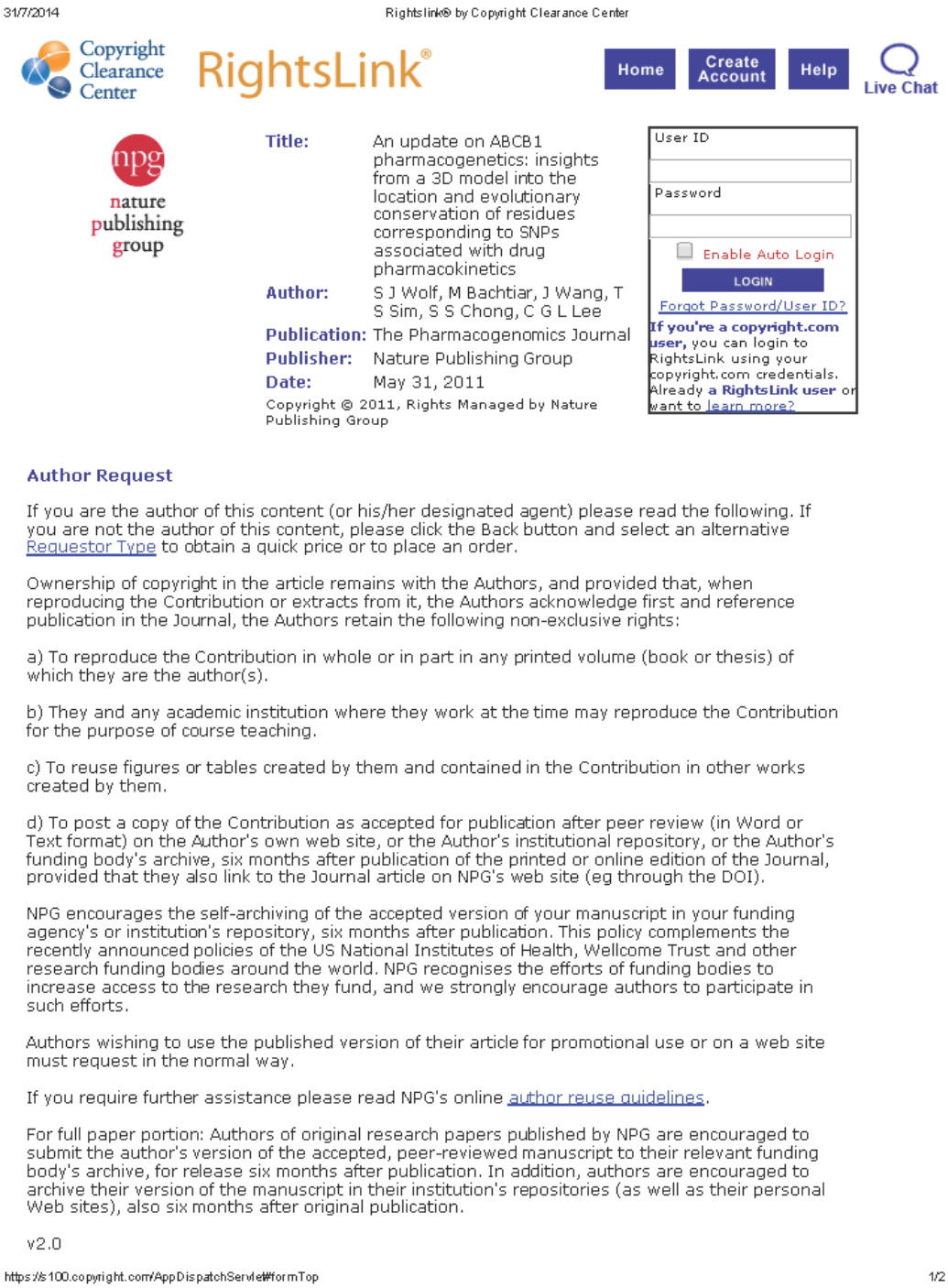
You will be invoiced within 48 hours of this transaction date. You may pay your invoice by credit card upon receipt of the invoice for this transaction. Please follow instructions provided at that time.
To pay for this transaction now; please remit a copy of this document along with your payment. Payment should be in the form of a check or money order referencing your

account number and this invoice number RLNK501366223.
Make payments to "COPYRIGHT CLEARANCE CENTER" and send to:

Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006
Please disregard electronic and mailed copies if you remit payment in advance.
Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

Explanation of Copyright Ownership from Nature Publishing Group



31/7/2014 Rightslink® by Copyright Clearance Center

Copyright Clearance Center RightsLink®

Home | Create Account | Help | Live Chat

nature publishing group

Title: An update on ABCB1 pharmacogenetics: insights from a 3D model into the location and evolutionary conservation of residues corresponding to SNPs associated with drug pharmacokinetics

Author: S J Wolf, M Bachtiar, J Wang, T S Sim, S S Chong, C G L Lee

Publication: The Pharmacogenomics Journal

Publisher: Nature Publishing Group

Date: May 31, 2011

Copyright © 2011, Rights Managed by Nature Publishing Group

User ID

Password

Enable Auto Login

LOGIN

Forgot Password/User ID?

If you're a copyright.com user, you can login to RightsLink using your copyright.com credentials. Already **a RightsLink user** or want to learn more?

Author Request

If you are the author of this content (or his/her designated agent) please read the following. If you are not the author of this content, please click the Back button and select an alternative Requestor Type to obtain a quick price or to place an order.

Ownership of copyright in the article remains with the Authors, and provided that, when reproducing the Contribution or extracts from it, the Authors acknowledge first and reference publication in the Journal, the Authors retain the following non-exclusive rights:

a) To reproduce the Contribution in whole or in part in any printed volume (book or thesis) of which they are the author(s).

b) They and any academic institution where they work at the time may reproduce the Contribution for the purpose of course teaching.

c) To reuse figures or tables created by them and contained in the Contribution in other works created by them.

d) To post a copy of the Contribution as accepted for publication after peer review (in Word or Text format) on the Author's own web site, or the Author's institutional repository, or the Author's funding body's archive, six months after publication of the printed or online edition of the Journal, provided that they also link to the Journal article on NPG's web site (eg through the DOI).

NPG encourages the self-archiving of the accepted version of your manuscript in your funding agency's or institution's repository, six months after publication. This policy complements the recently announced policies of the US National Institutes of Health, Wellcome Trust and other research funding bodies around the world. NPG recognises the efforts of funding bodies to increase access to the research they fund, and we strongly encourage authors to participate in such efforts.

Authors wishing to use the published version of their article for promotional use or on a web site must request in the normal way.

If you require further assistance please read NPG's online author reuse guidelines.

For full paper portion: Authors of original research papers published by NPG are encouraged to submit the author's version of the accepted, peer-reviewed manuscript to their relevant funding body's archive, for release six months after publication. In addition, authors are encouraged to archive their version of the manuscript in their institution's repositories (as well as their personal Web sites), also six months after original publication.

v2.0

https://s100.copyright.com/AppDispatchServlet#formTop 1/2

BACK CLOSE WINDOW