

Thesis

Open Access

Discovery of Novel DNA Variants in Jordanians Population by Re-Genotyping Affymetrix DMET Arrays Data Using DNA Sequencing

Marzooq Ammar AL*

Faculty of Graduate Studies, Jordan University of Science and Technology, Manama, Bahrain

Abstract

The Affymetrix DMETT^{MM} plus platform (Affymetrix, Santa Clara, CA, USA) is a GeneChip where 1936 SNPs can be genotyped in any given sample at once. The 1936 SNPs were distributed cross 225 genes in the genome. Thirtynine genes on the chip belong to phase I–II drug metabolism, disposition and drug transport gene family. These genes are functional in metabolizing the most widely prescribed anticancer drugs in the world including aromatase inhibitors, tamoxifen and thiopurines groups. In a high-throughput GeneChip array which is based on hybridization with allelespecific probes; genotyping errors are very common which limits the technologies applications; in addition missing calls for many SNPs on the chip immerge as a bigger and more serious problem in high- throughput genotyping methods. This study focuses on re-genotyping these No-call genotypes to maintain sample sizes that are already genotyped by Affymetrix DMETT^{MM} plus platform. Sixty six different variations were identified, 39 of them had No-call genotypes; our re-genotyping resulted in increasing calling rate from 89.08% to 95.56%. Furthermore, Among the variations that were identified, 8 were non-reported: C.-209T > G in UGT2B7 and -698°C > A in CYP1A2 as promoter variants, c.252G > T in SLC22A6 and c.1356T > C in SLC15A1 as silent mutation, c.1277 + 69°C > T and c.1277 + 82°C > T in SLC22A1 as intronic variants associated with each other and the most important variations are two missense (non- synonymous) in NAT2 gene: D20N c.58G > A and G11S c.31G > A. Affymetrix DMETT^{MM} plus platform.

Abbreviations: PHBC: Princess Haya Biotechnology Center; PCR: Polymerase Chain Reaction; RBC: Red Blood Cells; SDS: Sodium Dodecyl Sulphate; SE: Sodium EDTA; SNP: Single Nucleotide Polymorphism; TE: Tris-EDTA; DNA: Deoxyribonucleic Acid; EDTA: Ethylene Diamine Tetra Acetic Acid; TBE: Tris-base Borate EDTA; dNTPs: Deoxy Nucleoside Triphosphate; bp-Base Pair; COX2: Cyclooxygenase Isoenzymese 2; DMET: Drug Metabolizing Enzymes and Transporter; ADME: Absorption, Distribution, Metabolism, and Excretion; ABCB1: ATP-Binding Cassette Sub-Family B Member 1; ABCG2-ATP: Binding Cassette Sub-Family G Member 2; VKORC1: Vitamin K Epoxide Reductase Complex Subunit 1; FMO2: Dimethylaniline Monooxygenase [N-Oxide-Forming] 2; SLC22A1: Solute Carrier Family 22 Member 1; SLC15A1: Solute Carrier Family 15 Member 1; SLC15A2: Solute Carrier Family 15 Member 2; SLC22A6: Solute Carrier Family 22 Member 6; CYP1A1: Cytochrome P450, Family 1, Subfamily A, Polypeptide 1; CYP1A2: ytochrome P450, Family 1, Subfamily A, Polypeptide 2; TPMT: Thiopurine Methyltransferase; NAT2: N-Acetyltransferase 2; CDA: Cytidine Deaminase; UGT1A1-UDP-Glucuronosyltransferase 1-1; UGT2B7: UDP-Glucuronosyltransferase-2B7; µl: Micro liter; Nm: Nano liter; pmol: Pico mol; ml: Mili liter; g: gram; c: CELIOUS; 3`UTR-3' Untranslated Region; D or Asp: Aspartic Acid; S or Ser: Serine; G or Gly: Glysine; N or Asn: Asparagine

Introduction

The therapeutic efficacy of most drugs is influenced by a number of different factors that in part include age, weight and concurrent drug use [1]. These factors may vary between patients [2]. In addition, fixed parameters such as gender and human genome sequence variation can contribute. This genetic variation underlies every individual's response to drugs [3]. The vast majority of the enzymes involved in drug metabolism are highly polymorphic [4] and allele frequencies of low-activity variants often differ by population [5]. Consequently, their activity may differ depending upon an individual's genotype(s). For example, drugs may be metabolized more slowly in individuals who are carriers of a genetic polymorphism that results in a decreased or null activity of a given enzyme. These individuals are at particular risk for adverse drug reactions (ADRs) or therapeutic failure [6]. Conversely, drug therapy could be ineffective if the drug is metabolized too rapidly. Genetically determined variation particularly impacts drugs with narrow therapeutic indices, increasing the risk for the development of ADRs [7].

In the common complex diseases, multiple genes contribute to the phenotype, each with a small effect. To identify these genetic factors, pharmacogenomics approaches now include microarrays, highthroughput automated DNA sequencing. This new approach are more efficient and sensitive allowing for multiplexed analysis of mutations/ variants in several genes and thus are more predictive for complex diseases wherein multiple genetic factors are involved [8].

The Affymetrix' DMET[™] Plus platform (DMET stands for Drug Metabolizing Enzymes and Transporters) is such of this techniques, which enables highly multiplexed genotyping of known polymorphisms in Absorption, Distribution, Metabolism, and Elimination (ADME)-related genes on a single array [9].

Recently at PHBC they have been using the Affymetrix^{*} DMET^{TMM} plus platform (Affymetrix, Santa Clara, CA, USA) to genotyped 1936 SNPs variants (1931 single nucleotide polymorphisms "SNPs" and 5 copy number variations "CNVs"). The 1936 SNPs were generated from a total of 225 ADME-related genes [1]. These genes are functional in metabolizing the most widely prescribed anticancer drug in the world including aromatase inhibitors, tamoxifen and thiopurines groups [10,11]. The array uses molecular inversion probes (MIPs) that amplify and hybridize independently of genomic sequences as well as universal primers and tag sequences [12]. The platform uses a single-sample genotype calling method that compares each marker to an expected

*Corresponding author: Marzooq Ammar AL, M.Sc., Faculty of Graduate Studies, Jordan University of Science and Technology, Manama, Bahrain, Tel: 0097338898925; E-Mail: am2004mar@gmail.com

Received March 13, 2015; Accepted June 24, 2015; Published June 30, 2015

Citation: Marzooq Ammar AL (2015) Discovery of Novel DNA Variants in Jordanians Population by Re- Genotyping Affymetrix DMET Arrays Data Using DNA Sequencing. Mol Biol 4: 126. doi:10.4172/2168-9547.1000126

Copyright: © 2015 Marzooq Ammar AL, This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

signal distribution defined by large training sets at Affymetrix^{*} [10]. The genotypes for the 1,936 single nucleotide polymorphism (SNP) on the chip were essentially performed by applying the laboratory protocol followed the standard procedures described in the targeted genotyping (TG) System user guide developed by Affymetrix ^{*} Company (<u>http://www.affymetrix.com</u>).

In a high-throughput Gene Chip array which based on hybridization with allele-specific probes; genotyping errors is very common which limited the technology application; in addition missing No-call s for many SNPs on the chip immerge as bigger and more serious problem in high-throughput genotyping methods [10]. Several factors contribute to a No-call SNPs on the chip including; microarray manufacturing, probe hybridization stringent conditions, poor quality of DNA samples. These factors affected widely the final analysis of the genotyping result and frequently resulted in discarded the genotype for many SNPs [10]. Eliminations of many genotyping data because of the No-call genotypes SNPs leading to unexpected high percentages of heterozygosis for many SNPs on the chip and departure from Hardy-Weinberg equilibrium [13], since that, is important to achieve this study to:

- Confirmation by DNA sequencing the No-call alleles generated from Affymetrix DMETT^{MM} microarray data using DNA samples from the same personnel previously their DNA used on Affymetrix^{*} DMETTM.
- To investigate the extended effect of the No-call alleles on the general population allele frequency for these SNPs and reevaluate our cut-off and QC standard in filtering our Affymetrix DMETT^{MM} microarray data.
- To discover new DNA variants could be unique to Jordanian population and evaluate their role on drug efficacy and adverse reactions.

Literature Review

Pharmacogenetics and pharmacogenomics overview

The study of the relationship between genetics and therapeutic drugs is usually called Pharmacogenetics or pharmacogenomics [14]. Pharmacogenetics investigated the adverse reaction when it is happen at the genetics level where when investigations started at the genetics level such as GWAS studies to explain certain drug adverse issue in the patients [15,16].

Pharmacogenetics as a terminology covers the mechanism of drug actions as well as the genetic predisposition for adverse reactions [10,17].

Pharmacogenomics studies are directly linked to the population genetics studies; since the population genetics is investigated quantitatively the genetic variation in populations and how these variation maintained and inherited from generation to generation and fixed in given population. All these information is very helpful in pharmacogenomics studies and aids in highlighted genetic mutation and polymorphism in association with adverse reaction for drugs [1,2,18].

Population genetics research contributed enormously to molecular genetics testing by providing disease allele frequency for many disease in specific populations and resulted in defining targets genes and variation to be screen for in certain country or among specific ethnicity [19]. Population genetics; also established allele frequencies for many variants and these data been used in setting risk calculations and genetic susceptibility for complex diseases [1], and resulted in large scale testing for many peoples and improve the medical services provided for them [2]. Drugs response and adverse reaction variation cross different population and within the same communities suggested genetic variations also; and that's evident with a slow metabolizer phenotype for given drugs in certain population where the same drug metabolize the same drugs faster [18,19].

Page 2 of 49

Phenotypes showing poor and extra extensive metabolism have been documented at the genetic levels where genetic variation in gene encoding enzymes involved in metabolism (Drug Metabolism Enzymes and Transporters, DMET) [20] is characterized and reproducible [21].

Drug response is complex where several genes and several mutation and polymorphism shape up the outcome of the drug efficacy [22]. Add to this complexity is environmental factors; diet and others factors [8,23].

Racial and race-based genetic susceptibility in correlation with drug therapy also been documented where White Americans and African Americans with congestive heart failure are responding differently to the enalapril drug an angiotensin-converting enzyme inhibitor drugs [24-26]. In conclusion; personalized medicine is becoming a reality because of these finding of genetic variation and the association with drug phonotypic response and adverse reactions [2].

Genetic variants and drugs metabolism: Genetic variation in the human genome can be at single nucleotides or at the chromosome level. Cytogenetic rearrangements such as large deletions, inversions, and duplications occurred in the genome and been associated with diseases or it could be just normal variants. Single nucleotide polymorphism (SNP) also is one of the most common genome variation and occur in coding region of genes; regulatory regions; introns and at the splice junctions and there is and there is 0.1% between ant two genome at the SNPs level [27,28].

Most of the enzymes involved in drug metabolism are highly polymorphic, their allele's frequency also varied in different populations and their activities cross individuals is also different [4-7,29,30].

Drugs with narrow therapeutic indices (ED50/TD50) (Figure 2.1) are the most genes categories receive geneticist research. Today molecular diagnostic is very helpful in choosing the right drug and the optimal dosage for patients [31].







Figure 2.2: Number of drugs that were approved with pharmacogenomic information in their drug labels during each 10-year period from 1945-2005. During the 60 years covered by this analysis, 121 drugs were approved that have genomic biomarker information in current product labeling (Frueh et al., 2008).

The use of genetic profiles to individualize drug therapy is the aim of personalized medicine [32]. A recent study done by the Food and Drug Administration [33] shows at least 121 drugs were approved for molecular markers testing (Figure 2.2). Appling molecular testing in Pharmacogenetics contributes to reduce the death rate, improved quality of life, and very good positive economic impact [28].

Response rate for cancer drug and analgesic COX2 inhibitors drugs range between 25% for cancer treatments to 80% respectively and this due to genetic and environmental heterogeneity [31].

Haplotypes: All variation on single each chromosome is called a haplotype [34] and each set of these polymorphisms which are in phase passed on from generation to generation [29].

In population genetics, haplotype frequency and variation were successfully used in several genetics application including; pharmacogenomics, migration and immigration rates, genetic demography and human evolutionary history [35-37]. More importantly, haplotyping studies contributed into discovery for many genetics disease by mapping the causative genes [38]; and becoming more powerful than those based on single markers [35,39,40]. Haplotypes can be generated either by experimental base or from a family base studies [41-44].

Haplotype block structure for hundreds of genomic regions which scattered and mapped to all chromosomes were developed with their tagging SNPs and some of them been documented and link with drugs phenotypes including UM, EM, IM, PM metabolizing status [46] [Figure 2.3].

Historical overview of drugs metabolizing, eliminating and transporters dmet genes and phaemacogenetics

The effect of genetic factors on drug response was observed way back to the fifty's and serious studies started in seventies when Robert L Smith and colleagues studied the metabolism and PKs of debrisoquine (an antihypertensive drug) by giving the drug to themselves and watching its metabolism [47]. Population's studies shows at least 6–10% of Caucasians were poor metabolizers of debrisoquine a drug linked phenotypically to genetic variation in CYP2D6 gene (Figure 2.4) [48]. Genotype–phenotype studies of CYP2D6 variants are now performed using debrisoquine or dextromethorphanas surrogate substrates as probe drug [49].

Important allelic variants to drug treatment results have been

found in the genes encoding enzymes and transporters used in drug pharmacokinetics: absorption, distribution, metabolism and excretion (ADME) and equally important in Phase I and II enzymes [8,12].

Most of the phase I reactions are catalyzed by the cytochrome P450 (CYP) enzymes [52,53], However, CYP1, CYP2 and CYP3, catalyze most phase I reactions of drugs; where over 75% of prescribed drugs are metabolized in mainly by three subfamilies: CYP3A, CYP2D6 and CYP2C [54]. N- Acetyl Transferases 1 and 2 (NAT1 and



Figure 2.3: Any two copies of the human genome differ from one another by approximately 0.1% of nucleotide sites. In this example, most of the DNA sequence is identical in these chromosomes, but there are three nucleotides where variation occurs. Apattern of DNA sequence variation defines a haplotype (Catanzaro and Labb, 2009).



Figure 2.4: Pharmacogenetics of debrisoquine and nortriptyline. The activity of the *CYP2D6* enzyme is measured by the metabolic ratio (MR), which is the ratio of amounts of a substrate drug, debrisoquine and its metabolic product in urine after a standard dose of the drug. High ratios show poor conversion due to low enzyme activity (Wadelius and Piromohamed, 2007). The graph shows observed ratios for a range of patients. The same enzyme is largely responsible for phase I catabolism of the antidepressant drug nortriptyline. Depending on the *CPY2D6* phenotype as measured by the metabolic rario, patients require different doses of nortriptyline (Strachan and Read, 2011).

NAT2), Thiopurine S- Methyl Transferase (TPMT), and the Uridine Diphosphate Glucuronosyl Transferase (UGT) are major enzymes in Phase II reaction [55].

CYP3A family catalyzed over 37% of the drugs, where 17% ; 15%, 10%, 9% 6%, 4% and 2% are catalyzed by CYP2C9, CYP2D6, CYP2C19, CYP1A2, CYP2C8, CYP2A6 and CYP2E1, respectively [56]. CYP2C9, CYP2C19, and CYP2D6 genes polymorphisms are among the best established in correlation to drug therapy [52]. Currently the FDA approves molecular testing for many genes and approves also labeling of many drugs with strong recommendation for genotyping patients before given certain drugs such as atomoxetine, thioridazine, voriconazole and irinotecan [16].

Microarrays

Several technologies have been used in identifying singe gene mutation with Pharmacogenetics application such as RFLP and DNA sequencing [57]. Recently, microarrays, high-throughput automated DNA sequencing and genotyping, informatics, and mass spectrometry are widely used in high-throughput experiment in pharmacogenetics studies [58].

The first commercial SNP array was released in 1996 by Affymetrix (Santa Clara, CA) and targeted about 1,500 human SNPs [59] out of millions of SNPs characterized in the human genome [60] followed by Illumina (SanDiego, CA) and Nimblegen (Madison, WI) arrays which all are commercially available. Several statistical and computational tools have been developed to increase the efficiency of the Chip data [61,62].

Affymetrix[®] DMET[™]Plus platform

Affymetrix[®] DMET^{TMM} plus platform (Affymetrix[®], Santa Clara, CA, USA) Offers a complete profile of 1,936 markers in 225 genes focused on drug metabolism. Genes including on the chip are belong to different transcription regulators, selected drug targets and several classes of drug transporters [13,17]. The polymorphisms and mutations described on DMET Plus Arrays were chosen because of their Pharmacogenetics application and their association with clinical outcome in patients [11,63]; and these polymorphism and their clinical application have been discussed and approved by pharmaceutical industry and academia (ADME Consortium; https://pharmaadme.org/). The DMET plus Assay Panel has been tested across a minimum of 1,200 individuals from multiple populations including 715 DNA samples from Caucasian, African, and Japanese and Chinese populations from the International HapMap Consortium [20]. Key genes are included on the chip including VKORC1, cytochrome P450 [CYP] 2C9, CYP4F2 [64], CYP3A4, CYP2C9, CYP2C19 (Figure 2.5) [65].

Molecular inversion probe (MIP): A molecular inversion probe (MIP) is the technology which used in the Affymetrix^{*} DMETT^{MM} plus platform hybridization. MIP is a single oligonucleotide that recognizes and hybridizes to a specific SNP target and the probe design consist of two inverted recognition complementary regions flanks the target SNP need to be genotype [66]. The mechanism of the MIP is shown and explains in [figure 2.6] [67]. Another advantage for the MIP is an additional barcode sequence unique to each oligonucleotide [68]. The barcode has advantage in increasing the specificity of the SNP genotype regardless of the neighboring genomic sequence [69,70].

Affymetrix^{*} DMETT^{MM} plus platform data analysis: DMET[™] Console software is widely used in analyzing the Affymetrix^{*} DMETT^{MM} plus platform data. The soft wear translates the genotype results into star nomenclature [71]. The star nomenclature style is always used

YP2C19 Genotype Frequency

Page 4 of 49

Genotype *1/*1	Caucasian 77%	Chinese 37%	Japanese 31%	African 61%
*1/*2	18%	42%	27%	27%
*1/*3	0%	8%	17%	0%
*1/*9	0%	0%	0%	3%
*2/*2	0%	9%	12%	4%
*1/*3	5%	4%	10%	0%
*D/*D	0%	0%	3%	1%

*D/*D=*3/*3 for Japanese, *2/*9 for African

Figure 2.5: Allelic frequencies of *CYP2C19* variants across populations. *1/*3 variant (noted in red) is found in much greater frequency in Chinese and Japanese populations. This information can help inform recruitment for trials examining this variant. If you have an indication that *1/*3 is resulting in a PK issue, you would need a very huge trial if you recruited patients from anywhere other than Chinese/Japanese pops (Brandt et al., 2007).



in research studies of pharmacogenomics. DMET[™] Console software equipped with preset and analytically validated cluster for each SNPs on the Chip [20] which facilitated the transformation of the genotype into clinical application [9].

DMET Case Control Data Analysis workflow [13] summarized in the following steps, as shown in [Figure 2.7].

- Biological samples collection in preparing for microarray experiments [13].
- Generation raw microarray data (CEL data) [13].
- DMET data preprocessing: DMET Console software produces a table summarizing the detected SNP and linked them to the sample.
- SNPs detection and find there significant [13].

No-call values

Studying millions of single nucleotide polymorphisms (SNPs) and genotyping them is facilitated because of the advance in high-throughput microarrays technologies [72,73]. Several methods have



Figure 2.7: Workflow of a clinical bioinformatics experiment from the sample collection to the data analysis.Workflow of data in a typical DMET Case-Control Data analysis.

been developed to detect genotyping errors or removing its effects on analyses [74-76].

Human's high density SNP microarray chips are successfully implemented more than in other species Chip and that due to the sequence of the human genome; and the genotyping errors in the range from 0.05% to 5% [55]. Where other species chip such as cattle microarray the genotyping errors could reach up to 20% (Su, 2005).

Most of the time when one SNP genotype score is wrong it affected building accurate haplotype which leads to ignore all the data generated for one sample and resulted in the loss of a huge data portions (Yu et al., 2009). Genotyping errors and missing genotyping SNPs demonstrated in the finding of the data of the Phase I and Phase II Hap Map, where less than 20% data that failed to pass QC was due to genotyping, while more than 65% of the markers show missing data in over 20% individuals [10,77]. Various methods of estimating missing values in a dataset exist and been developed for deletion, insertion of sample mean, and linear regression [78]. Each approach presents problems built in the method itself or in the nature of the pattern of missing data [79]. Correction for these data is too expensive and can be done in two different ways including repeat the genotyping or alter the data analysis tools to make room for the missing data [80-83].

Large studies on the accuracy of haplotype inference methods have been carried out [38,84], but until recently there has been little investigation of the effect of missing data on these techniques [85].

There are two major causes for No-call genotypes. One is due to poor quality of DNA where this resulted in not enough strong intensity of fluorescence signals over the background. The other comes when an observation, i.e., a read out of fluorescence signals, cannot be put clearly to any of the clusters of genotype, therefore, is subject to 'Nocall' procedure [10] (Figure 2.8).

Also, there are many different types of genetic variations that are not seen well by genotyping technologies, like rare SNPs, and short deletions and insertions. Because of this another genotype technique should take a place which No-call could be as a signal for Novel SNPs or other variations types [86].

Materials and Methods

Sampling frame

Jordanian individuals donated blood samples to be used as a probe for Affymetrix^{*} DMETT^{MM} plus platform (Affymetrix^{*}, Santa Clara, CA, USA); with the aims of determine the pharmacogenomics of ADME genes profile in the Jordanian population. The goal is defining any pathological haplotype frequency with the hope of setting molecular diagnostic for personalized medicine in Jordan.

Inclusion and exclusion criteria

Some of the samples out-put No-call genotypes data in specific SNPs after using the Affymetrix^{*} DMET^{TMM} plus platform (Affymetrix^{*}, Santa Clara, CA, USA) in PHBC, this included important genes like: CDA, FMO2, SLC22A6, UGT1A1, UGT2B7, CYP1A1, SLC22A1, NAT2, ABCB1, ABCG2, SLC15A2, VKORC1, CYP1A2, SLC15A1, TPMT. Copy number variants and pseudo genes validation is excluded from this study.

Sampling procedure

All the samples that fit our criteria were included in our study as possible.

Sample size calculation

Sample size needed to satisfy the objectives was calculated using the following equation:

 $n = ((SD^* t)/\Delta) 2)$

where SD is the Standard Deviation which is assumed to be 17.12 based on the study made in University of Jordan; t is the value of t-test at α (type one error) assumed to be 0.05; Δ is The meaningful difference needed to be observed among different genotypes. The sample size equal 44 but we increase the sample size to 101 to account for potential missing data.



Figure 2.8: The clustering results based on a one-marker-at-a-time method. Values on the X-axis and Y-axis are normalized signal intensities of two alternative alleles (A andC). Estimated genotypes "AA", "AC", and "CC" are indicated by symbols "0", "1", and "2", respectively. Question marks represent missing values (i.e. No-call s) (Yu et al., 2009).

Page 6 of 49

Hypothesis to be treated

Jordanian populations have their privet genetic variations in ADME genes that influence the sufficient therapeutic dosage for a numerous number of drugs. No-call genotype for several SNPs (Affymetrix^{*} DMETT^{MM} plus platform) is limited the technology. Novel SNPs at the DMET probe positions might be the cause of the No-call result, for this reason we carried massive genotyping for the No-call using DNA sequence technology.

Institutional research board (IRB) approval

The study has been approved by the Institutional Research Board (IRB) (Appendix D).

Subjects

A total of 101 control DNA samples donated by Jordanian individuals were included in our study which has been used to generate the Chip data (Affymetrix DMETT^{MM} plus platform) in PHBC by professional geneticist's staff, and then it was investigated in this study. Fifty samples were collected before in PHBC since 2010 and they have their chip data and 51 samples collected for this works (Figure 3.2). All volunteers completed a consent form in Arabic (Appendix E).

Primer design

Initially, the No-call SNPs genotypes defined; and there location on





Figure 3.2: Summary of study design. (*) represents the steps done by PHBC and afterwards the completed product can be obtained to use include in our study.

the correspondent genesis fixed. Both database websites http://www. ensembl.org/index.html, and http://www.ncbi.nlm.nih.gov/ utilized in this processes. The Primer3 software online: http://primer3.ut.ee/ used to design PCR primers flanking these No-call SNPs obtained from Affymetrix^{*} DMETT^{MM} plus platform. Primer designed executed carefully to include more than one No-call SNPs in the same amplified region. All primers were synthesized in PHBC.

Methods

The workflow includes sample collection, DNA extraction, polymerase reaction (PCR), gel electrophoresis, purification, DNA sequencing, sequencing analysis and statistical analysis.

Sample collection: After a signed informed consent, 3ml venous blood samples were collected by a vein puncture procedure into 3 ml EDTA tubes from each volunteer and the tubes were mixed properly to ensure the appropriate mixing of the blood with EDTA to prevent clotting. This was done according to the University Review Committee for Research on Humans at Jordan University of Science and Technology. The blood samples were stored at 4°C until the genomic DNA was extracted from each.

DNA Extraction: Genomic DNA from a total of 51 blood samples was isolated from peripheral blood by using a commercially available kit (Qiagen Germany).

DNA extraction procedure

- 1. For each 300 μl sample volume: 900 μl of RBC Lysis Solution was added to a sterile 1.5 ml micro centrifuge tube.
- The tube of blood was gently rocked until thoroughly mixed; then transferred to the tube containing the RBC Lysis Solution. The tube was inverted 5–6 times to mix.
- 3. The mixture was incubated for 10 minutes at room temperature (inverted 2–3 times during the incubation) to lyse the red blood cells, and then was centrifuged at 13,000–16,000 \times g for 20 seconds at room temperature.
- 4. The Supernatant was removed and as much as possible was discarded without disturbing the visible white pellet. Approximately $10-20 \ \mu$ of residual liquid remained in the 1.5 ml tube.
- 5. The tube was vortexed vigorously until the white blood cells were resuspended (10–15 seconds) to obtain efficient cell Lysis.
- 6. About 300µl Cell Lysis Solution was added to the tube containing the resuspended cells. The solution was pipetted 5–6 times to lyse the white blood cells until the solution became very viscous. If clumps of cells were visible after mixing, the solution was incubated at 37°C until the clumps were disrupted.
- About 100 μl of Protein Precipitation Solution was added to the nuclear lysate and vortexed vigorously for 10–20 seconds.
- 8. The mixture was centrifuged at $13,000-16,000 \times g$ for 3 minutes at room temperature. A dark brown protein pellet was visible.
- 9. The supernatant was transferred to a clean 1.5 ml micro centrifuge tube containing 300 μ l of isopropanol incubated at room temperature.

Note: Some supernatant may remain in the original tube containing the protein pellet. This residual liquid was left in the tube to avoid cross-contamination of the DNA solution with the precipitated protein.

- 10. The solution was gently mixed by inversion until the white thread-like strands of DNA formed a visible mass.
- 11. The mixture was centrifuged at $13,000-16,000 \times g$ for 1 minute at room temperature. The DNA was visible as a small white pellet.
- 12. The supernatant was decanted and one sample volume of room temperature 70% ethanol was added to the DNA. The tube was gently inverted several times to wash the DNA pellet and then step 11 was repeated.
- 13. Ethanol was carefully aspirated using either a drawn Pasteur pipette or a sequencing pipette tip. The DNA pellet was very loose at this point and care was taken to avoid aspirating the pellet into the pipette. The tube was inverted onto a clean absorbent paper and the pellet was air-dried for 10–15 minutes.
- 14. About 100 μ l of DNA Rehydration Solution was added to the tube and the DNA was rehydrated by incubation at 65°C for 1 hour. The solution was periodically mixed by gently tapping the tube. Alternatively, the DNA was rehydrated by incubating the solution overnight at room temperature or at 4°C.
- 15. DNA was stored at 2-8°C.

Quantification of isolated genomic DNA: The DNA yield was measured by using a NanoDrop 1000 spectrophotometer (Thermo Scientific). After mixing, 1 μ l from each sample was transferred onto the lower measurement pedestal of the instrument, and using operating software on the computer, the concentration of each sample was measured in ng/ μ l. The TE buffer which was used for rehydration after DNA extraction, was used here as a blank.

Primers optimization: We performed gradient and touchdown PCR reaction using the DNA samples to define the best annealing temperature for the PCR conditions (Table 3.1).

Polymerase chain reaction (PCR): Twenty pairs of primers were tested and have been used to amplify genomic DNA for over 101 DNA samples. Primer sequences and there expected product sizes are listed in [Table 3.1] and [Figure 3.3].

For each pair primers amplification was carried out using GoTaq^{*} Green Master Mix (Promega; USA). Each PCR reaction was carried out in a total volume of 26 μ l containing 12.5 μ l of PCR master mix (Promega, USA), 8.5 μ l of nuclease free water, 1.3 μ l (5-10 pmol/ μ l) of each primer (Table 3.1) and 2-3 μ l of DNA. The amplification protocol was: initial denaturation for 10 min at 95°C, 35 cycles were performed

Gene	Name	Forward primers	Reverse primers	Tm	Product Stp
CD.4	Primer1	TGAAGCCTGAGTGTGTCCAG	OTCTOOCACCTTACTOAGOCT	60 °C	4770p
FM02	Primer1	TGAGGGCGTATTCTCAATCTG	octottottottotoocctttttt	58 °C	519bp
SLC2246	Primer1	TOOCCTITAATGACCTCCTO	OGAGGAAGAGGGTGACATOG	60 °C	ssabp
	Internel		OOTAGATGOGAAOOTOCTOT		
UGTIAI	Primer1	GAAAGCCCACAAATCCAAGA	TOCTITOGOGAAAAAGAATG	58 °C	5996p
CGT28*	Primer1	TOCAGAAACTAOOGOTOGAG	CAGCAACTOGAAAAACAAGCA	60 °C	562bp
8412	Printer1	CTIGCTEAGGGGATCATOGA	AAACCTOGTGATOGATCCCT	60 °C	878bp
	Internel		AAACCOATTOTOOTCAOAOC		
ABCBI	Primer1	GTCCAAGAACTGOCTTTGCT	TCCATGACTTCAGAATGCTGA	58 °C	s7stp
	Primer2	AGGATTTCCCTTCTTCCGATT	CAMOOCAATTCACAGACACAO	62 °C	600bp
	Primer3	AAAAACGCTTCCATGTACCC	ATCTCCACGAAGGCAGAGT	62 °C	5.88bp
	Primer-4	OTCCAOTOCCACTACOOTTT	OCTOAATCATTOOOOACATT	60 °C	550bp

with denaturation at 95°C for 30 seconds, annealing at each primer temperature °C (Table 3.1) which was determined in the optimization step for 30s, and then extension at 72°C for 45 seconds. The final extension step takes 10 min at 72°C (S1000[™] Thermal Cycler, BioRad) (Figure 3.1).

Page 7 of 49

Gel electrophoresis: 5 μ l of each PCR product were loaded into wells of 2% agarose gel, also 5 μ l of 100 bp ladders (Fermentas) was loaded into the gel. 1X Tris borate EDTA (TBE) running buffer was used to run the gel electrophoresis under a constant voltage (140 V) for 40 minutes. 5 μ l from 10 mg/ml Ethidium Bromide (Biorad, USA), an intercalating agent, was added to the buffer to stain the DNA for visualization under UV transluminator provided with a gel documentation system (Biorad, USA). The approximate PCR product size was determined by matching up the bands to known bands of the ladder.

PCR product purification: Before sequencing the PCR products, the products would be purified to remove impurities like nucleotides, unlinked primers, enzyme and Mg^{+2} . For this purpose the EZ- 10 spin^{*} column PCR purification kit (Bio Basic, Canada) was used according to the manufacturer instructions. The PCR product concentration would be measured by means of NanoDrop 1000 (Thermo Scientific, Massachusetts. USA).

Procedure for purification of PCR product

- 1. The PCR reaction mixture was transferred to a 1.5 ml microcentrifuge tube and 5 volumes of Binding Buffer III was added to it.
- 2. The column was placed into a 2.0 ml collection tube and the resulting mixture was transferred to the column. The column stand was left at room temperature for 2 minutes then it was spun at 8000 rpm for 1 minute.



Page 8 of 49



- The flow-through was discarded. About 500 µl Wash Solution was added to the column and spun at 10,000 x g for 1 minute. The flow-through was discarded and column was placed back to the same collection tube.
- 4. 500μ l Wash Solution was then added to the column and then it was spun at 10,000 x g for 1 minute. The flow-through was discarded and spun once more to remove the residue from the Wash Solution.
- 5. The column was then transferred to a clean 1.5 ml micro tube. About 30-50 μl Elution Buffer was added onto the membrane center part of the column. The mixture was incubated at 50°C for 2 minutes then it was spun down at 10,000 x g for 1 minute.

Then the purified PCR product was ready to be used in the Sequencing PCR, or kept at -20° C until needed for use.

PCR sequencing: According to the visualized purified PCR product, 1-3 μ l of the purified product was incorporated with 1 μ l of R-R mix (as provided in Bigdye' terminator V 3.1 Cycle Sequencing kit (Applied bio systems, Foster city, CA, USA), and 5 pmol forward or reverse primer, 4 μ l Big Dye Terminator v1.1/3.1 sequencing buffer (5X), and 4 μ l nuclease free water .The reaction tubes were then loaded into a thermal cycler through the PCR- sequencing amplification protocol.

Cleaning of the PCR sequencing: The PCR-sequencing product was then cleaned using a specially designed kit (NucleoSEQ' supplied from Applied Bio, USA), according to the manufacturer's instructions:

After the addition of nuclease-free water to the agarose columns,

these columns underwent centrifugation for 3 minutes at 1500 rpm. Then the PCR product was loaded at the center of the column and underwent the same previous step of centrifugation, in order to elute the product into a sterile microfuge tube. Finally, 50 μ l of nuclease-free water was added to each tube and mixed. Subsequently, 20 μ l of cleaned sequence PCR product was loaded onto the 3130xl genetic analyzer; capillary electrophoresis system (Applied bio systems, Foster city, California, USA).

Analysis of the samples' electropherograms: The resultant sequences were analyzed by alignment of the standard reference sequence of each primer product according to sample inclusion criteria and the noted differences for each sample were reported by the SNPs position on the Affymetrix^{*} DMET[™] microarray screen and trying to find any kinds of suspects in the sequence that may cause the No-call SNP genotypes, screening other variations and validation of DMET GeneChip by Call genotypes.

Statistical analysis: Measuring call rates of each marker included in this study, before re-genotyping and after. Also, determining error rate through validation process [Figure 3.2,3.3].

Results

Study's participants

A total of 101 healthy Jordanian DNA samples were collected; and hybridized on Affymetrix^{*} DMET^{TMM} plus platform (Affymetrix, Santa Clara, CA, USA). Sixty six variants were detected (39 SNPs of them are No-call target). These SNPs and others (like SNPs that appear in

Page 9 of 49



Mol Biol ISSN: 2168-9547 MBL, an open access journal

Polymerase chain reaction (PCR) results

More than 500 PCR reactions were carried out using pairs of primers to amplify 20 different regions in DMET Genes. PCR was followed by 2% agarose gel electrophoresis to visualize the PCR product, check the size of the amplified target regions, and to ensure that there was no contamination. The amplified product size for SLC22A1-region1 was 528 bp (Figure 4.1), for SLC22A1-region2 was



Figure 4.1: The separation of PCR product of SLC22A1-region1 by 2% agarose gel. Lane M: 100bp ladder. Lane -ve: the negative control. Lanes 1-12: the 528bp amplified products for samples that have No-call data in AM_14350 and AM_14348 probe ID codes.



Figure 4.2: The separation of the PCR product of SLC22A1-region2 using 2% agarose gel. Lane M: 100bp ladder. Lane -ve: the negative control. Lanes 1-12: the 556bp amplified products for samples that have No-call data in AM 14361 and AM 14363 probe ID codes



Figure 4.3: The separation of the PCR product of ABCG2-region1 using 2% agarose gel. Lane M: 100bp ladder. Lane -ve: the negative control. Lanes 1-13: the 660bp amplified products for samples that have No-call data in AM 13689 probe ID code.



Figure 4.4: The separation of the PCR product of CYP1A2-region1 using 2% agarose gel. Lane M: 100bp ladder. Lane -ve: the negative control. Lanes 1-5: the 785bp amplified products for samples that have No-call data in AM_10783, AM_10785 and AM_10784 probe ID codes.



Page 10 of 49

Figure 4.5: The separation of the PCR product of TPMT-region1 using 2% agarose gel. Lane M: 100bp ladder. Lane -ve: the negative control. Lanes 1-13: the 500 bp amplified products for samples that have No-call data in AM_13973 probe ID code.



Figure 4.6: The separation of the PCR product of TPMT-region2 using 2% agarose gel. Lane M: 100bp ladder. Lane -ve: the negative control. Lanes 1-6: the 510bp amplified products for samples that have No-call data in AM_13980 probe ID code.



Figure 4.7: The separation of the PCR product of VKORC1-region2 using 2% agarose gel. Lane M: 100bp ladder. Lane -ve: the negative control. Lanes 1-2: the 567bp amplified products for samples that have No-call data in AM_11045 probe ID code.



556 bp (Figure 4.2), for ABCG2-region1 was 660 bp (Figure 4.3), for CYP1A2-region1 was 785 bp (Figure 4.4), for TPMT-region1 was 500 bp (Figure 4.5), for TPMT-region2 was 510 bp (Figure 4.6), for VKORC1-region2 was 567 bp (Figure 4.7), for SLC15A1-region1 was 495 bp (Figure 4.8), for CYP1A1-region1 was 433 bp (Figure 4.9), for FMO2-region1 was 519 bp (Figure 4.10), for NAT2-region1 was 878 bp (Figure 4.11), for SLC15A2-region1 was 429 bp (Figure 4.12), for CDA-region1 was 477 bp (Figure 4.13), for UGT1A1-region1 was 599 bp (Figure 4.14), for UGT2B7- region1 was 562 bp (Figure 4.15),

Mol Biol

Page 11 of 49



for SLC22A6-region1 was 558 bp (Figure 4.16), for ABCB1-region1 was 578 bp (Figure 4.17), for ABCB1-region2 was 620 bp (Figure 4.18), for ABCB1-region3 was 549 bp (Figure 4.19) and for ABCB1-region4 was 580 bp (Figure 4.20).

DNA sequencing results

After the PCR products were visualized on gels the remaining amplified products were purified and amplified again using the same forward or reverse primers that used in the PCR. DNA was sequenced on a 3130/3130xl Genetic Analyzer (Applied Biosystem).

A total of 64 variations were identified in 20 different amplified regions for 15 ADMEs genes among 101 samples, 39 of them are the target of this study which appear as No-call genotypes in Affymetrix DMETT^{MM} plus platform (Affymetrix, Santa Clara, CA, USA) in some samples and Call in other samples (Table 4.1). Shows these SNPs with their code number in ensembles Database and NCBI, their common name that descripted their location and type, number of reactions that our study covers with No-call samples. While, from (Figure 4.21- 4.53)



Figure 4.12: The separation of the PCR products of *SLC15A2*-region1 using 2% agarose gel. Lane M: 100bp ladder. Lane –ve: the negative control. Lanes 1-6: the 429 bp amplified products for samples that have No-call data in AM_13301 probe ID code.



Figure 4.13: The separation of the PCR products of *CDA*-region1 using 2% agarose gel. Lane M: 100bp ladder. Lane –ve: the negative control. Lanes 1-18: the 477bp amplified products for samples that have No-call data in AM_11499 probe ID code.



Figure 4.14: The separation of the PCR products of *UGT1A1*-region1 using 2% agarose gel. Lane M: 100bp ladder. Lane –ve: the negative control. Lanes 1-14: the 599bp amplified products for samples that have No-call data in AM_13067, AM_13070 and AM_13068 probe ID codes.



Figure 4.15: The separation of the PCR products of *UGT2B7*-region1 using 2% agarose gel. Lane M: 100bp ladder. Lane –ve: the negative control. Lanes 1-15: the 562bp amplified products for samples that have No-call data in AM_13458 and AM_13459 probe ID codes.



Figure 4.16: The separation of the PCR products of *SLC22A6*-region1 using 2% agarose gel. Lane M: 100bp ladder. Lane –ve: the negative control. Lanes 1-6: the 558bp amplified products for samples that have No-call data in AM_10347 probe ID codes.



Figure 4.17: The separation of the PCR products of *ABCB1*-region1 using 2% agarose gel. Lane M: 100bp ladder. Lane –ve: the negative control. Lanes 1-9: the 578bp amplified products for samples that have No-call data in AM_14592probe ID code.



respectively, represents their genotypes differences (partial sequencing Electropherogram) according the sequence references database (we were used ensembles: http://asia.ensembl.org/index.html). Accordingly, Call rates were improve from 89.08% to 95.56% (Table 4.7) in these 39 No-call genotypes probes, details of each marker is present in the previously mentioned table.

Furthermore, 273 No-call genotypes in the Affymetrix[®] DMETT^{MM} plus GeneChips were identified (Appendix E), where approximately



Page 12 of 49

Figure 4.19: The separation of the PCR products of *ABCB1*-region3 using 2% agarose gel. Lane M: 100bp ladder. Lane –ve: the negative control. Lanes 1-8: the 549bp amplified products for samples that have No-call data in AM_14628 probe ID code.



Figure 4.20: The separation of the PCR products of *ABCB1*-region4 using 2% agarose gel. Lane M: 100bp ladder. Lane –ve: the negative control. Lanes 1-6: the 580bp amplified products for samples that have No-call data in AM_14633 probe ID code.

	SNPs is number	Common Name	Probe Set ID	N
1	rs2072671	CD4+2_c.79A>C(K27Q)	AM_11499	18
2	rs7662029	UGT2B7+2_c.327G>A(Promoter)	AM_13458	15
3	rs7668258	UGT2B7#2_c.161C>T(Promoter)	AM_13459	0
4	rs10929303	UGT1A1*76_c.*211C>T(3'UTR)	AM_13067	1
5	rs8330	UGT1A1*79_c.*440C>G(3'UTR)	AM_13070	7
6	rs1042640	UGT1A1*78_c.*339C>G(3'UTR)	AM_13068	5
7	rs1143670	SLC15A2_c.1161A>G(A387A)	AM_13301	16
8	rs6661174	FMO2_23238T>C(X472Q)	AM_11958	-4
9	rs45564337	SLC2246_c.149G>A(R50H)	AM_10347	10
10	rs1142345	TPMT+3C_c.719A>G(Y240C)	AM_13973	4
11	rs1800460	TPMT*3B_c.460G>A(A154T)	AM_13980	4
12	rs12208357	SLC22A1_c.181C>T(R61C)	AM_14348	3
13	rs55918055	SLC22A1_c.262T>C(C88R)	AM_14350	4
14	rs45512393	SLC22A1_c.1201G>A(G401S)	AM_14361	7
15	rs628031	SLC22A1_c.1222G>A(V408M)	AM_14363	17
16	rs1048943	CYP1A1*2C_2454A>G(I462V)	AM_10768	4
17	rs1799814	CYP1A1*4_2452C>A(T461N)	AM_10769	6
18	rs41279188	CYP1A1+5or+9_2460C>A>T(R464SorC)	AM_10766	2
19	rs72547510	CYPIAI*7_2345insT	AM_10771	5
20	rs2069526	CYP1A2*1K739T>G(Promoter)	AM_10783	12
21	rs762551	CYP1A2+1F_+163C>A(Promoter)	AM_10785	5
22	rs12720461	CYP1A2*1K729C>T(Promoter)	AM_10784	1
23	rs1805158	N4T2+19_c.190C>T(R64W)	AM_14998	2
24	rs1041983	NAT2_c.282C>T(Y94Y)	AM_15000	4
25	rs1801280	NAT2*5_c341T>C(I114T)	AM_15001	19
26	rs72554616	NAT2+17_c.434A>C(Q145P)	AM_15005	3
27	rs1799929	NAT2 c.481C>T(L161L)	AM 15006	4

Table 4.1: Summary of the 39 variations that have been identified in our study that are as part of No-call data in some of samples in Affymetrix® DMETT^{MM} plus platform (Affymetrix®, Santa Clara, CA, USA), which represent RS number according ensambles database, AM numbers according Affymetrix® DMETT^{MM} plus platform (Affymetrix®, Santa Clara, CA, USA) product, commen name of this SNPs and numbers of sequencing reaction of each SNPs.

	SNPs is number	Common Name	Probe Set ID	N
28	rs72554617	N4T2+10_c.499G>A(E167K)	AM_15007	2
29	rs1799930	NAT2*6_c.590G>A(R197Q)	AM_15008	4
30	rs1208	NAT2_c.803A>G(K268R)	AM_15010	2
31	rs1339067	SLC15A1_c.1347T>C(A449A)	AM_10650	20
32	rs2274827	SLC15A1_c.1375C>T(R459C)	AM_10647	28
33	rs2235033	ABCB1_c.1554+24C>T	AM_14609	5
34	rs1128503	ABCB1_c.1236C>T(G412G)	AM_14612	17
35	rs3213619	ABCB1_c129T>C	AM_14633	1
36	rs9282564	ABCB1_c.61A>G(N21D)	AM_14628	1
37	rs2032582	ABCB1_c.2677G>T>A(A893SorT)	AM_14592	1
38	rs72552713	ABCG2_c.376C>T(Q126X)	AM_13689	1
39	rs72547529	VKORC1_c.196G>A(V66M)	AM 11044	1

Table 4.1: Summary of the 39 variations that have been identified in our study that are as part of No-call data in some of samples in Affymetrix® DMETT^{MM} plus platform (Affymetrix®, Santa Clara, CA, USA), which represent RS number according ensambles database, AM numbers according Affymetrix® DMETT^{MM} plus platform (Affymetrix®, Santa Clara, CA, USA) product, commen name of this SNPs and numbers of sequencing reaction of each SNPs.

47.25% of No-call are wild-type genotypes, 37% heterozygous and 15.75% were homozygous for variants.

Additionally, 19 of variations identified were reported in the Database; 10 of them are already markers in the Affymetrix^{*} DMETT^{MM} plus platform (Affymetrix^{*}, Santa Clara, CA, USA) (Table 4.2), that are also used to validate the GeneChip by comparing with their Call genotypes in the chips. These SNPs are represented in (Figure 4.54) to (Figure 4.63) respectively to show their genotype variance as Electropherogram results gotten after the direct sequencing method. While the other SNPs reported are listed in (Table 4.3) and represent in (Figure 4.65) to (Figure 4.72) respectively to show their alternative genotypes that we have gotten in our samples that are included in our study.

Our sequencing screening study also found new variations in 8 different positions which are not reported in Ensemble and NCBI databases (Table 4.4). Two of them have an influence on protein level due to amino acid changing: C.-209T > G is a UGT2B7 variant located in the promoter region that changes G to T at position C.-209 (Figure 4.72). Sample code number DM-12-035 was heterozygous (GT), while other samples that were screened have wild type genotypes (GG). We still do not know if this new variation has influences on protein level.

	SNPs code	Probe Set ID	Common Name
1	r\$2020869	AM_11959	FMO2_c.*60A>G
2	r\$2842934	AM_13979	<i>TPMT_c.</i> 474C>T(I158I)
3	rs1867351	AM_14347	SLC22A1_c.156T>C(S52S)
4	rs34447885	AM_14342	<i>SLC22A1_</i> c.41C>T(S14F)
5	rs72552763	AM_14368	SLC22A1_c.1260_1262delGAT(M420del)
6	rs4986996	AM_15002	NAT2*12D_c.364G>A(D122N)
7	rs8187838	AM_10648	SLC15A1_c.1352C>A(T451N)
8	r\$2032588	AM_14610	ABCB1_c.1350+44C>T
9	r\$2214102	AM_14631	ABCB1_c1G>A
10	rs9934438	AM_11045	VKORC1_c.174-136C>T

Table 4.2: Summary of 10 reported SNPs that identified and were used to validate Affymetrix® DMETT^{MM} plus platform (Affymetrix®, Santa Clara, CA, USA) product that used in 101 Jordanian samples in PHBC, Table shows SNPs codes and their commen name with probe ID in this GeneChip.

Page 14 of 49

	SNPs code	Gene	region	Position	variants
1	rs12059454	CDA	Intron variant	Chromosome 1:20915813	G/A
2	rs12731069	CDA	Intron variant	Chromosome 1:20916012	G/A
3	rs7668282	UGT2B7	Promoter	Chromosome 4:69962114	T/C
4	1873823859	UGT2B7	Promoter	Chromosome 4:69962101	G/A
5	rs3762819	SLC15A2	Intron variant	Chromosome 3:121646759	G/A
6	rs1882002	SLC15A2	Intron variant	Chromosome 3:121646495	C/T
7	rs28369911	FMO2	Intron variant	Chromosome 1:171177858	G/T
8	rs113569197	SLC22A1	Splice donar variant delation	Chromosome 6:160560898- 160560905	TGGTAAGT/
9	rs9457843	SLC22A1	Intron variant	Chromosome 6:160560933	C/T

Table 4.3: Summary of 9 reported SNPs that identified. Table shows SNPs codes and their gene, region, position and their variants.

Gene	type	Position	Nomeclature	Sample
UGT2B7	Promoter	4:69962029	c209T>G	DM-12-035
SLC22.46	silent mutation	11:62751911	c.252G>T	DM-10-053
SLC22A1	Intron variant	6:160560962	c.1277+69C>T	DM-10-027
				DM-10-021
				DM-10-028
				DM-10-043
				DM-10-060
SLC22.41	Intron variant	6:160560982	c.1277+82C>T	DM-10-027
				DM-10-021
				DM-10-028
				DM-10-043
				DM-10-060
				DM-10-025
CYP1A2	Promoter	15:75041652	-698C>A	DM-10-026
				DM-10-015
SLC15.41	silent mutation	13:99356603	c.1356T>C	DM-10-040
NAT2	Missence	8:18257572	c.58G>A(D20N)	DM-10-005
NAT2	Missence	8:18257544	c.31G>A(G11S)	DM-10-002
				DM-10-005
				DM-10-014
				DM-10-057

Table 4.4: Summary of about 8 non- reported variants that identified. Table shows these variants and their genes located in, types, position, which samples that identified in and their nomenclature.

Page 15 of 49



TCTTTAGTTTTGTGTCAAATGGACTGCAGAAACAAGATCTGTC ACTGCTACTGTTCTGGACACTCTTCTAAAATATATTGCAT

Figure 4.22: Partial sequencing Electropherogram of UGT2B7 gene using primers that cover one of the No-call SNPs; showing c.327G>A(Promoter) that has probe ID code AM_13458 in the Affymetrix® DMETT^{IMM} plus platform (Affymetrix®, Santa Clara, CA, USA) system. **A.** The wild type allele shows a conserved A at position c.327Gin DM-10-035 sample. Vertical arrow points to the conserved nucleotide A (green peak). **B.** UGT2B7 allele showing a heterozygous A \rightarrow G in DM-10-001 sample. Vertical arrow points to the changed nucleotide A/G (green/ black peak). **C.** UGT2B7 allele showing homozygous A \rightarrow G in DM-12-052 sample at position c.327G. Vertical arrow points to the changed nucleotide G (black peak). **D.** Part of reference sequence of UGT2B7 gene exon

 $A \rightarrow G$ in DM-12-052 sample at position c.327G. Vertical arrow points to the changed nucleotide G (black peak). **D.** Part of reference sequence of *UGT2B7* gene exon red front base pair represents the location of the nucleotide change c.327G. The above reference

UGT2B7 gene sequence is according to the ensemble database which this SNP under rs7662029: http://asia.ensembl.org/Homo_sapiens/Variation/Sequence?db=core;r=4:69961412- 69962412;v=rs7662029;vdb=variation;vf=5391360

Page 16 of 49



c.161C in DM-10- 026 sample. Vertical arrow points to the conserved nucleotide T (red peak). **B**. *UGT2B7* allele showing a heterozygous T \rightarrow C in DM-12-024 sample. Vertical arrow points to the changed nucleotide T \rightarrow C (red/ blue peak). **C**. *UGT2B7* allele showing homozygous T \rightarrow C in DM-12-012 sample at position c.161C. Vertical arrow points to the changed nucleotide C (blue peak). **D**. Part of reference sequence of *UGT2B7* gene exon red front base pair represents the location of the nucleotide change at position c.161C. The above reference *UGT2B7* gene sequence is according to the ensemble database which this SNP under rs7668258: http://asia.ensembl.org/Homo sapiens/Variation/Sequence?db=core;r=4:69961412- 69962412;v=rs7662029;vdb=variation;vf=5391360



TGAAAAATGGCACATGTCATCCTGATCAAAGACACCAGAGGGGGGCAC GATACATATTCAAATATCTTTACTGACTAGCGAGTCTATTATTTTTATTT

Figure 4.24: Partial sequencing Electropherogram of *UGT1A1* gene using primers that cover one of the No-call SNPs; showing c.*211C>T(3'UTR) that has a probe ID code AM_13067 in the Affymetrix® DMETT^{IMM} plus platform (Affymetrix®, Santa Clara, CA, USA) system. NOTE: reverse primer use here in PCR sequencing step. **A.** The wild type allele shows a conserved G at position c.*211C in DM-10-046 sample. Vertical arrow points to the conserved nucleotide G (black peak). **B.** *UGT1A1* allele showing a heterozygous G→A in DM-12-027 sample. Vertical arrow points to the changed nucleotide G→A (black/ green peak). **C.** *UGT1A1* allele showing homozygous G→A in DM-12-027 sample. Vertical arrow points to the changed nucleotide A (green peak). **C.** *UGT1A1* allele showing homozygous G→A in DM-12-024 sample at position c.*211C. Vertical arrow points to the changed nucleotide A (green peak). **D.** Part of reference sequence is according to the ensemble database which this SNP under rs10929303: http://asia.ensembl.org/Homo_sapiens/Transcript/Exons?db=core;g=ENSG0000241635; r=2:234526291-234681956;t=ENST00000373450

Figure 4.25: Partial sequencing Electropherogram of *UGT1A1* gene using primers that cover one of No-call SNPs; showing c.*440C>G(3'UTR) that have probe ID code AM_13070 in the Affymetrix® DMETT^{MM} plus platform (Affymetrix®, Santa Clara, CA, USA) system. **A**. The wild type allele shows a conserved C at position c.*440C in DM-12- 024 sample. Vertical arrow points to the conserved nucleotide C (blue peak). **B**. *UGT1A1* allele showing a heterozygous C \rightarrow G in DM-10-007 sample. Vertical arrow points to the changed nucleotide C \rightarrow G (blue/black peak). **C**. *UGT1A1* allele showing homozygous C \rightarrow G in the DM-10-019 sample at position c.*440C. Vertical arrow points to the changed nucleotide G (black peak). **D**. Part of reference sequence of *UGT1A1* gene 3'UTR red font base pair represents the location of the nucleotide change at position c.*440C. The above reference *UGT1A1* gene sequences is according to the ensemble database which this SNP under rs8330: http://asia.ensembl.org/Homo_sapiens/Transcript/Exons?db=core;g=ENSG0000241635;r=2:234526291_234681456.**ENST0000373450

=2:234526291-234681956:t=ENST00000373450

Figure 4.26: Partial sequencing Electropherogram of $\frac{1}{4}$ GT1A1 gene using primers that cover one of the No-call SNPs; showing c.*339C>G(3'UTR) that has a probe ID code AM_13068 in the Affymetrix DMET plus platform (Affymetrix, Santa Clara, CA, USA) system. NOTE: reverse primer use here in PCR sequencing step. **A.** UGT1A1 allele showing a heterozygous G \rightarrow C in DM-10-015 sample. Vertical arrow points to the changed nucleotide G \rightarrow C (black/blue peak). B. UGT1A1 allele showing homozygous G-C in DM-10-035 sample at position c.*339C. Vertical arrow points to the changed nucleotide C (blue peak). C. Part of reference sequence of UGT1A1 gene 3'UTR red front base pair represents the location of the nucleotide change at position c.*339C. The above reference UGT1A1 gene sequence is according to the ensemble database which this SNP under rs1042640: http://asia.ensembl.org/Homo_sapiens/Transcript/Exons?db=co re;g=ENSG00000241635;r

=2:234526291-234681956;t=ENST00000373450

Figure 4.27: Partial sequencing Electropherogram of *SLC15A2* gene using primers that cover one of the No-call SNPs; showing c.1161A>G(A387A) that has a probe ID code AM_13301 in the Affymetrix® DMETT^{MM} plus platform (Affymetrix®, Santa Clara, CA, USA) system. **A.** The wild type allele shows a conserved A at position c.1161 in DM-12- 047 sample. Vertical arrow points to the conserved nucleotide A (green peak). **B.** *SLC15A2* allele showing a heterozygous A→G in DM-10-015 sample. Vertical arrow points to the changed nucleotide A→G (green/black peak). **C.** *SLC15A2* allele showing homozygous A→G in DM-10-060 sample at position c.1161A. Vertical arrow points to the changed nucleotide G (black peak). **D.** Part of reference sequence of *SLC15A2* gene exon; red front base pair represents the location of the nucleotide change at position c.1161A. The above reference *SLC15A2* gene sequence is according to the ensemble database which this SNP under rs1143670: http://asia.ensembl.org/Homo_sapiens/Transcript/Exons?db=core;g=ENSG0000163406;r = 3:121612936-121662949;t=ENST0000489711

rode AM_11958 in the Affymetrix® DMETT^{MM} plus platform (Affymetrix®, Santa Clara, CA, USA) system. **A.** The wild type allele shows a conserved T at position 23238 in DM-12-052 sample. Vertical arrow points to the conserved nucleotide T (red peak). **B.** *FMO2* allele showing a heterozygous T \rightarrow C in DM-10-054 sample. Vertical arrow points to the changed nucleotide T \rightarrow C (red/blue peak). **C.** Part of reference sequence of *FMO2* gene exon; red front base pair represents the location of the nucleotide change at position 232381. The above reference *FMO2* gene sequence is according to the ensemble database which this SNP under rs6661174: http://asia.ensembl.org/Hom_sapiens/Transcript/Exons?db=core;g=ENSG0000094963;r =1:171154347-171181822;t=ENST00000441535

Page 19 of 49

pair represents the location of the nucleotide change at position c.149G. The above reference SLC22A6 gene sequence is according to the ensemble database which this SNP under rs45564337: http://asia.ensembl.org/Homo_sapiens/Transcript/Exons?db=core;g=ENSG00000197901;r =11:62703857-62752455;t=ENST00000377871

Figure 4.30: Partial sequencing Electropherogram of *TPMT* gene using primers that covers one of the No-call SNPs; showing c.719A>G(Y240C) that has a probe ID code AM_13973 in the Affymetrix® DMETT^{MM} plus platform (Affymetrix®, Santa Clara, CA, USA) system. **A.** The wild type allele shows a conserved A at position c.719A in DM-10- 027 sample. Vertical arrow points to the conserved nucleotide A (green peak). **B.** *TPMT* allele showing a heterozygous A→G in DM-10-017 sample. Vertical arrow points to the changed nucleotide A→G (green/black peak). C. Part of reference sequence of TPMT gene exon; red front base pair represents the location of the nucleotide change at position c.719A. The above reference TPMT gene sequence is according to the ensemble database which this SNP under rs1142345: http://asia. ensembl.org/Homo_sapiens/Transcript/Exons?db=core;g=ENSG00000197901;r

Figure 4.31: Partial sequencing Electropherogram of *TPMT* gene using primers that covers one of the No-call SNPs; showing c.460G>A(A154T)that has a probe ID code AM 13980 in the Affymetrix⁶, Santa Clara, CA, USA) system. A. The wild type allele shows a conserved G at position c.460G

Figure 4.31: Partial sequencing Electropherogram of *TPMT* gene using primers that covers one of the No-call SNPs; showing c.460G>A(A154T) that has a probe ID code AM_13980 in the Affymetrix® DMETT^{MM} plus platform (Affymetrix®, Santa Clara, CA, USA) system. **A**. The wild type allele shows a conserved G at position c.460G in DM-10- 017 sample. Vertical arrow points to the conserved nucleotide G (black peak). **B**. *TPMT* allele showing a heterozygous G \rightarrow A in DM-10-037 sample. Vertical arrow points to the conserved nucleotide G (black peak). **B**. *TPMT* allele showing a heterozygous G \rightarrow A in DM-10-037 sample. Vertical arrow points to the conserved nucleotide G (black peak). **B**. *TPMT* allele showing a heterozygous G \rightarrow A in DM-10-037 sample. Vertical arrow points to the changed nucleotide G \rightarrow A (black/green peak). **C**. Part of reference sequence of *TPMT* gene exon; red front base pair represents the location of the nucleotide change at position c.460G. The above reference *TPMT* gene sequence is according to the ensemble database which this SNP under rs1800460: http://asia. ensembl.org/Homo_sapiens/Transcript/Exons?db=core;g=ENSG00000197901;r = 11:62703857-62752455;t=ENST00000377871

Figure 4.32: Partial sequencing Electropherogram of *SLC22A1* gene using primers that covers one of the No-call SNPs; showing c.181C>T(R61C)that has a probe ID code AM_14348 in the Affymetrix® DMETT^{MM} plus platform (Affymetrix®, Santa Clara, CA, USA) system. **A**. The wild type allele shows a conserved C at position c.181C in DM-10- 060 sample. Vertical arrow points to the conserved nucleotide C (blue peak). **B**. *SLC22A1* allele showing a heterozygous C \rightarrow T in DM-12-035 sample. Vertical arrow points to the changed nucleotide C \rightarrow T (blue/red peak). **C**. *SLC22A1* allele showing homozygous C \rightarrow T in DM-10-015 sample at position c.181C. Vertical arrow points to the changed nucleotide T (red peak). **D**. Part of reference sequence of *SLC22A1* gene exon red front base pair represents the location of the nucleotide change at position c.181C. The above reference *SLC22A1* gene sequence is according to the ensemble database which this SNP under rs12208357: http:// asia.ensembl.org/Homo_sapiens/Variation/Sequence?db=core;r=6:160542648- 160543648;v=rs12208357;vdb=variation;vf=8477839

Page 20 of 49

Figure 4.33: Partial sequencing Electropherogram of SLO22A7 gene using primers that covers one of the No-call SNPs, showing c.20212 (Coort) that has a probe ID code AM_14350 in the Affymetrix® DMETT^{MM} plus platform (Affymetrix®, Santa Clara, CA, USA) system. **A.** The wild type allele shows a conserved T at position c.262T in DM-12- 035 sample. Vertical arrow points to the conserved nucleotide T (red peak). **B.** Part of reference sequence of SLC22A1 gene exert represents the location of the nucleotide change at position c.262T. The above reference SLC22A1 gene sequence is according to the ensemble database which this SNP under rs55918055: http://asia.ensembl.org/Homo_sapiens/Variation/Sequence?db=core;r=6:160542648- 160543648;v=rs12208357;vdb=variation;vf=8477839

Figure 4.34: Partial sequencing Electropherogram of *SLC22A1* gene using primers that covers two of the No-call SNPs; showing c.1201G>A(G401S) (blue vertical arrow) and c.1222G>A (red vertical arrow) that has a probe ID code AM_14361 and AM_14363 (respectively) in the Affymetrix® DMETT[™] plus platform (Affymetrix®, Santa Clara, CA, USA) system. **A.** The wild type allele shows a conserved G at position c.1201G and c.1222G in DM-11-015 sample. Vertical arrows points to the conserved nucleotide G (black peak). **B.** Part of reference sequence of *SLC22A1* gene exon; blue and red front bases represents the location of the nucleotide change at position c.1201G and c.1222G. The above reference *SLC22A1* gene sequence is according to the ensemble database which these SNPs under rs45512393 and rs628031: http://asia.ensembl.org/Homo_sapiens/Transcript/Exons?db=core;g=ENSG00000175003;r=6:160542821-160579750;t=ENST00000366963

Figure 4.36: Partial sequencing Electropherogram of *CYP1A2* gene using primers that covers one of the No-call SNPs; showing -163C>A(Promoter) that has a probe ID code AM_10785 in the Affymetrix® DMETT^{MM} plus platform (Affymetrix®, Santa Clara, CA, USA) system. **A.** The wild type allele shows a conserved C at position -163C in DM-10-030samples. Vertical arrow points to the conserved nucleotide C (blue peak). **B.** *CYP1A2* allele showing homozygous C→A in DM-10-015 sample at position -163C. Vertical arrow points to the changed nucleotide A (green peak). **C.** Part of reference sequence of *CYP1A2* gene Promoter; red front base pair represents the location of the nucleotide change at position -163C. The above reference *CYP1A2* gene sequence is according to ensemble database which this SNP under rs762551: http://asia.ensembl.org/Homo_sapiens/Transcript/Exons?db=core;g=ENSG00000140505;r = 15:75041185-75048543;t=ENST00000343932

00140465:r=15:75011883-75017951:t=ENST00000379727

Figure 4.38: Partial sequencing Electropherogram of *CYP1A1* gene using primers that covers one of the No-call SNPs; showing 2345insT that has a probe ID code AM_10771 in the Affymetrix® DMETT^{MM} plus platform (Affymetrix®, Santa Clara, CA, USA) system. **A.** The wild type allele shows a conserved which T base that should insert in this SNP is absent at position 2345insTin DM-10-015 sample. Vertical arrow points to the conserved nucleotides without T insertion. **B.** Part of reference sequence of *CYP1A1* gene exon red front base pair represents the location of the nucleotide change at position 2345insT. The above reference *CYP1A1* gene sequence is according to ensemble database which this SNP under rs72547510: http://asia.ensembl.org/Homo_sapiens/Variation/Sequence?db=core;r=15:75012594-75013593;v=rs72547510;vdb=variation;vf=16826816

showing a heterozygous C \rightarrow Tin DM-10-025 sample. Vertical arrow points to the changed nucleotide C \rightarrow T (blue/red peak). **C**. *NAT2* allele showing homozygous C \rightarrow T in DM-10-019 sample at position c.282C. Vertical arrow points to the changed nucleotide T (red peak). **D**. Part of reference sequence of *NAT2* gene exon red front base pair represents the location of the nucleotide change at position c.282C. The above reference *NAT2* gene sequence is according to ensemble database which this SNP under rs1041983: http://asia.ensembl.org/Homo_sapiens/Transcript/Exons?db=core;g=ENSG00000156006;r =8:18248755-18258728;t=ENST00000286479

Mol Biol ISSN: 2168-9547 MBL, an open access journal

Page 25 of 49

Figure 4.42: Partial sequencing Electropherogram of *NAT2* gene using primers that covers one of the No-call SNPs; showing c.434A>C(Q145P) that has a probe ID code AM_15005 in the Affymetrix® DMETT[™] plus platform (Affymetrix®, Santa Clara, CA, USA) system. **A.** The wild type allele shows a conserved A at position c.434A in DM-10-002 sample. Vertical arrow points to the conserved nucleotide A (green peak). **B.** Part of reference sequence of *NAT2* gene exon; red front base pair represents the location of the nucleotide change at position c.434A. The above reference *NAT2* gene sequence is according to ensemble database which this SNP under rs72554616: http://asia.ensembl.org/Homo_sapiens/Transcript/Exons?db=core;g=ENSG00000156006;r = 8:18248755-18258728;t=ENST00000286479

Page 26 of 49

The code AM_1500s in the Alignetity² DME 1⁻⁻⁻⁻⁻ plus platform (Alignetity², Santa Clara, CA, OSA) system. A. The wind type aliefe shows a conserved on position (Calignetity², Santa Clara, CA, OSA) system. A. The wind type aliefe shows a conserved not conserved nucleotide G (black peak). **B.** NAT2 allele showing a heterozygous G→A in DM-10-055 sample. Vertical arrow points to the changed nucleotide G→A (black/green peak). **C.** NAT2 allele showing homozygous G→A in DM-10-019 sample at position c.590G. Vertical arrow points to the changed nucleotide G (black/green peak). **C.** NAT2 allele showing homozygous G→A in DM-10-019 sample at position c.590G. Vertical arrow points to the changed nucleotide A (green peak). **D.** Part of reference sequence of NAT2 gene exon red front base pair represents the location of the nucleotide change at position c.590G. The above reference NAT2 gene sequence is according to ensemble database which this SNP under rs1799930: http://asia.ensembl.org/Homo_sapiens/Transcript/Exons?db=core;g=ENSG00000156006;r

=8:18248755-18258728;t=ENST00000286479

Page 27 of 49

Figure 4.45: Partial sequencing Electropherogram of *NAT2* gene using primers that coversone of the No-call SNPs; showing c.803A>G(K268R) that has a probe ID code AM_15010 in the Affymetrix® DMETT^{MM} plus platform (Affymetrix®, Santa Clara, CA, USA) system. **A.** The wild type allele shows a conserved G at position c.803A>G in DM- 10-035 sample. Vertical arrow points to the conserved nucleotide G (black peak). **B.** *NAT2* allele showing a heterozygous G \rightarrow A in DM-10-020 sample. Vertical arrow points to the changed nucleotide G \rightarrow A (black/green peak). **C.** *NAT2* allele showing homozygous G \rightarrow A in DM-10-025 sample at position c.803A>G. Vertical arrow points to the changed nucleotide A (green peak). **D.** Part of reference sequence of *NAT2* gene exon red front base pair represents the location of the nucleotide change at position c.803A>G. The above reference *NAT2* gene sequence is according to ensemble database which this SNP under rs1208: http:// asia.ensembl.org/Homo_sapiens/Transcript/Exons?d=core;g=ENSG0000156006;r = 8:18248755-18258728;t=ENST0000286479

AACGCCACACGCTTCTAGTGTGGGCCCCCAATCACTACCAGGTGgtaa

Figure 4.46:Partial sequencing Electropherogram of *SLC15A1* gene using primers that covers two of the No-call SNPs; showing c.1347T>C(A449A) (blue vertical arrow) and c.1375C>T(R459C) (red vertical arrow) that have probe ID codes AM_10650 and AM_10647 (respectively) in the Affymetrix® DMETT^{MM} plus platform (Affymetrix®, Santa Clara, CA, USA) system. **A.** The wild type allele shows a conserved T at position c.1347T and conserved C at position c.1375Cin DM-10-043 sample. Vertical arrows points to the conserved nucleotides T (red peak) and C (blue peaks). **B.** *SLC15A1* allele showing a heterozygous T→C (red/blue peak) in 1st SNP; whereas 2nd SNP stay in wild type status in DM-10-002 sample. Vertical arrows points to the changed nucleotides in these SNPs. **C.** *SLC15A1* allele showing a heterozygous T→C (red/blue peak) in 1st SNP; whereas 2nd SNP stay in wild type status in DM-10-002 sample. Vertical arrow points to the changed nucleotides T (red peak), SLC15A1 allele showing. **C.** *SLC15A1* allele showing to the changed nucleotide to the SNPs. **C.** *SLC15A1* allele showing to reference sequence of *SLC15A1* gene exon; blue and red front bases represents the location of the nucleotides change at positions c.1347T and c.1347T. T and c.1347T. T and c.1347T and c.1347T. T and c.1347T.

=13:99336055-99404908;t=ENST00000376503

TAAAGTCATAGGCATTGGCTT

Figure 4.48: Partial sequencing Electropherogram of *ABCB1* gene using primers that covers one of the No-call SNPs; showing c.1236C>T(G412G) that has a probe ID code AM_14612 in the Affymetrix® DMETT^{MM} plus platform (Affymetrix®, Santa Clara, CA, USA) system. NOTE that reverse primer used here in PCR sequencing step. **A.** *ABCB1* allele showing a heterozygous A \rightarrow G in DM-10-022 sample. Vertical arrow points to the changed nucleotide A \rightarrow G (green/black peak). **B.** *ABCB1* allele showing homozygous G \rightarrow A in DM-10-017 sample at position c.1236C. Vertical arrow points to the changed nucleotide A (green peak). **C.** Part of reference sequence of *ABCB1* gene exor; red front base pair represents the location of the nucleotide change at position c.1236C. The above reference *ABCB1* gene sequence is according to ensemble database which this SNP under rs1128503:

http://ssia.ensembl.org/Homo_sapiens/Transcript/Exons?db=core;g=ENSG00000085563;r =7:87133175-87342611;t=ENST00000265724

CTCTGCTTCTTTGAGCTTGGAAGAGCCGCTACTCGAATGAGCTCAGG CTTCCTGTGGCAAAGAGAGCGGAAGCGGCTGTGCTCAGCCCACGCCC

Figure 4.49: Partial sequencing Electropherogram of *ABCB1* gene using primers that covers one of the No-call SNPs; showing c.-129T>C that has a probe ID code AM_14633 in the Affymetrix® DMETT^{MM} plus platform (Affymetrix®, Santa Clara, CA, USA) system. NOTE that reverse primer used here in PCR sequencing step. **A**. *ABCB1* allele showing a heterozygous A \rightarrow G in DM-10-001 sample. Vertical arrow points to the changed nucleotide A \rightarrow G (green/black peak) at position c.-129T. **B**. Part of reference sequence of *ABCB1* gene 5'UTR; red front base pair represents the location of the nucleotide change at position c.-129T. The above reference *ABCB1* gene sequence is according to ensemble database which this SNP under rs3213619: http://asia.ensembl.org/Homo_sapiens/Transcript/Exons?db=core;g=E NSG0000085563;r

=7:87133175-87342611;t=ENST00000265724

tgaaaatgaaacaagctagttacCTTTTATTGTTCAGTTTAAAAAAGTTCTTCTTC TTTGCTCCTCCATTGCGGTCCCCTTCAAGATCCATTCCGACetgaagaga **Figure 4.50:** Partial sequencing Electropherogram of *ABCB1* gene using primers that covers one of the No-call SNPs; showing c.61A>G(N21D) that has a probe ID code AM_14628 in the Affymetrix[®] DMETT^{MM} plus platform (Affymetrix[®], Santa Clara, CA, USA) system. NOTE that reverse primer used here in PCR sequencing step. **A.** *ABCB1* allele showing a heterozygous C \rightarrow T in DM-11-013 sample. Vertical arrow points to the changed nucleotide T \rightarrow C (red/ blue peak). **B.** Part of reference sequence of *ABCB1* gene exon red front base pair represents the location of the nucleotide change at position c.61A. The above reference *ABCB1* gene sequence is according to ensemble database which this SNP under rs9282564: http://asia.ensembl.org/Homo_sapiens/Transcript/Exons?db=core;g=ENSG00000085563;r =7:87133175-87342611:t=ENST00000265724 B tttagtttgactcacCTTCCCAGAACCTTCTAGTTCTTTCTTATCTTTCAGTGC TTGTCCAGACAACATTTTCATTTCAACAACTCCTGCTATTGCAATGA Figure 4.51: Partial sequencing Electropherogram of ABCB1 gene using primers that covers one of the No-call SNPs; showing c.2677G>T>A (A893SorT) that has a probe ID code AM_14592 in the Affymetrix® DMETT^{MM} plus platform (Affymetrix®, Santa Clara, CA, USA) system. NOTE that reverse primer used here in PCR sequencing step. A. ABCB1 allele showing a heterozygous C-A in DM-10-014 sample. Vertical arrow points to the changed nucleotide C-A (blue/black peak) at position c.2677G. B. Part of reference sequence of ABCB1gene exon; red front base pair represents the location of the nucleotide change at position c.2677G. The above reference ABCB1 gene sequence is according to ensemble database which this SNP under rs2032582: http://asia.ensembl.org/Homo_sapiens/Transcript/ Exons?db=core;g=ENSG00000085563;r =7:87133175-87342611;t=ENST00000265724 ATTTCAAATGTAATTCAGGTTACGTGGTACAAgtaagtattagtgggtttgcattttctgt tteetetgtttetatatgggtaagtgetttggetgatagtteaatgtgetteeagttgatta Figure 4.52: Partial sequencing Electropherogram of *ABCG2* gene using primers that covers one of the No-call SNPs; showing c.376C>T(Q126X) that has a probe ID code AM_13689 in the Affymetrix® DMETT^{MM} plus platform (Affymetrix®, Santa Clara, CA, USA) system. NOTE that reverse primer used here in PCR sequencing step. **A.** The wild type allele shows a conserved G at position c.376C in DM-10-035 sample. Vertical arrow points to the conserved nucleotide G (black peak). B. Part of reference sequence of ABCG2 gene exon red front base pair represents the location of the nucleotide change at position c.376C. The above reference ABCG2 gene sequence is according to ensemble database which this SNP under rs72552713: http://asia.ensembl.org/Homo sapiens/Transcript/Exons?db=core;g

=ENSG00000118777;r =4:89011416-89152474;t=ENST00000237612 Page 29 of 49

CCTGTCCCAGCACATGCTCCACCAGCCCGAAACCCCCTGCCCCACetg gcagaggggggggggggggggggaaccaggttaggactgtcaacccagtgccttggaccctgcccgagaaagg Figure 4.53: Partial sequencing Electropherogram of VKORC1 gene using primers that covers one of the No-call SNPs; showing c.196G>A(V66M) that has a probe ID code AM_11044 in the Affymetrix® DMETT^{MM} plus platform (Affymetrix®, Santa Clara, CA, USA) system. NOTE that reverse primer used here in PCR sequencing step. A. The wild type allele shows a conserved C at position c.196G in DM-12-047 sample. Vertical arrow points to the conserved nucleotide C (blue peak). B. Part of reference sequence of VKORC1 gene exon red front base pair represents the location of the nucleotide change at position c.196G. The above reference VKORC1 gene sequence is according to ensemble database which this SNP under rs72547529: http://asia.ensembl.org/Homo_sapiens/Transcript/Exons?db=core; g=ENSG00000167397;r

=16:31102163-31107301;t=ENST00000394975

=1:171154347-171181822;t=ENST00000441535

Page 30 of 49

C. Part of reference sequence of *SLC22A1* gene exon red front base pair represents the location of the nucleotide change at position c.156T. The above reference *SLC22A1* gene sequence is according to ensemble database which this SNP under rs1867351: http://asia.ensembl.org/Homo_sapiens/Transcript/Exons?db=core;g =ENSG00000175003;r

=6:160542821-160579750;t=ENST00000366963

CCACAGCCACTGTTTCCAACCAGGGCCACCGTCTGCCCACTCTGC ACCTTCAGGTTCAGACCCTTCAAGATctaccaggacgagtgagaaaaaaaacttcaag

Figure 4.58: Partial sequencing Electropherogram of *SLC22A1* gene using primers that covers one of the No-call SNPs; besides that showing c.1260_1262_ delGAT(M420del) that has a probe ID code AM_14368 in the Affymetrix® DMETT^{MM} plus platform (Affymetrix®, Santa Clara, CA, USA) system, which is a variant that can we validate technique through call data results. **A.** The wild type allele shows sequence at position c.1260_1262delGAT(M420del) without GAT deletion in DM-12-010 sample. Horizontal red box mentions the conserved nucleotides. **B.** *SLC22A1* allele showing a heterozygous GAT \rightarrow - (deletion GAT in one allele) in DM-10-060 sample. Horizontal red box points to thedeletion variant that causes overlapping peaks in electropherogram sequence due to in frame shift effect. **C.** Part of reference sequence of *SLC22A1* gene exon red front base pair represents the location of the nucleotide change at position c.1260_1262delGAT(M420del). The above reference *SLC22A1* gene sequence is

according to ensemble database which this variant under rs72552763:

http://asia.ensembl.org/Homo_sapiens/Transcript/Exons?db=core;g=ENSG00000175003;r

=6:160542821-160579750;t=ENST00000366963

Page 33 of 49

=8:18248755-18258728;t=ENST00000286479

Figure 4.60: Partial sequencing Electropherogram of *SLC15A1* gene using primers that covers No-call SNPs; besides that showing c.1352C>A(T451N) that has a probe ID code AM_10648 in the Affymetrix[®] DMETT^{MM} plus platform (Affymetrix[®], Santa Clara, CA, USA) system, which is a SNPs that can we validate technique through call data results. **A.** The wild type allele shows a conserved C at position c.1352C in DM-10-017 sample. Vertical arrow points to the conserved nucleotide C (blue peak). **B.** *SLC15A1* allele showing a heterozygous C→A in DM-10-037 sample. Vertical arrow points to the changed nucleotide C→A (blue/green peak) at position c.1352C. **C.** Part of reference sequence of *SLC15A1* gene exon; red front base pair represents the location of the nucleotide change at position c.1352C. The above reference *SLC15A1* gene sequence is according to ensemble database which this SNP under rs8187838: http://asia.ensembl.org/ Homo_sapiens/Transcript/Exons?db=core;g=ENSG0000088386;r =13:99336055-99404908;t=ENST00000376503

Page 34 of 49

B TGCTCCTCCATTGCGGTCCCCTTCAAGATCCATTCCGACetgaagagaaac

egcageteattagecaaatgeatgageeteaggegegetggaggtgagaetaacetetagteeceegtegaagee

Figure 4.62:Partial sequencing Electropherogram of *ABCB1* gene using primers that covers the No-call SNPs; besides that showing c.-1G>A that has a probe ID code AM_14631 in the Affymetrix® DMETT^{MM} plus platform (Affymetrix®, Santa Clara, CA, USA) system, which is a SNPs that can we validate technique through call data results. **A.** The wild type allele shows a conserved C at position c.-1G in DM-11-013 sample. Vertical arrow points to the conserved nucleotide C (blue peak). **B.** Part of reference sequence of *ABCB1* gene 5'UTR red front base pair represents the location of the nucleotide change at position c.-1G. The above reference *ABCB1* gene sequence is according to ensemble database which this SNP under rs2214102: http://asia.ensembl.org/Homo_sapiens/Transcript/Exons?db = core;g=ENSG0000085563;r

=7:87133175-87342611;t=ENST00000265724

 $C \rightarrow T$ in DM-10-001 sample. Vertical arrow points to the changed nucleotide $C \rightarrow T$ (blue/red peak). **C.** Part of reference sequence of *CDA* gene intron red front base pair represents the location of the nucleotide change at position. The above reference *CDA* gene sequence is according to ensemble database: http://asia.ensembl.org/Homo_sapiens/Transcript/Exons?db=core;g=ENSG00000158825;r

=1:20915441-20945401;t=ENST00000375071

B C cctctaagcattgaaagaccctggatttaagccacatttgagcgcggtctggggca Figure 4.65: Partial sequencing Electropherogram of the CDA gene using primers that covers one of the No-call SNPs; besides that showing other SNPs that are reported in the data base; here one of these which has a code rs12731069. NOTE that reverse primers are used here in the PCR sequencing step. A. The wild type allele shows a conserved C in DM-10-047 sample. Vertical arrow points to the conserved nucleotide C (blue peak). B. CDA allele showing a heterozygous C-+T in DM-10-046 sample. Vertical arrow points to the changed nucleotide C -> T (blue/red peak). C. Part of reference sequence of CDA gene intron red front base pair represents the location of the nucleotide change at position. The above reference CDA gene sequence is according to ensemble database: http://asia.ensembl.org/ Homo_sapiens/Transcript/Exons?db=core;g=ENSG00000158825;r =1:20915441-20945401;t=ENST00000375071 TTTGACTTATAAGGGTTACATTTTAACTTCTTGGCTAAT Figure 4.66: Partial sequencing Electropherogram of the *UGT2B7* gene using primers that covers one of the No-call SNPs; besides that showing other SNPs that are reported in the data base; here are two of these SNPs which have codes rs73823859 and rs7668282 respectively. A. The wild type allele shows a conserved T in DM-10-040 sample in 2nd SNP. Vertical red arrow points to the conserved nucleotide T (red peak), whereas 1st SNP showing a heterozygous $G \rightarrow A$. Vertical blue

arrow points to the changed nucleotide G \rightarrow A (black/green peak) in same sample. **B**. *UGT2B7* allele showing a heterozygous T \rightarrow C in DM-10-016 sample in 1st SNPs. Vertical blue arrow points to the changed nucleotide T \rightarrow C (red/blue peak), whereas 2nd SNP shows a conserved G in DM-10-047 sample in. Vertical red arrow points to the conserved nucleotide G (black peak) in same sample. C. Part of reference sequence of UGT2B7 gene Promoter blue and red front base pair represents the location of the nucleotides changes inthese SNPs position. The above reference UGT2B7 gene sequence is according to ensemble database: http://asia. ensembl.org/Homo_sapiens/Transcript/Exons?db=core;g=ENSG00000171234;r =4:69917081-69978705;t=ENST00000305231

Page 36 of 49

Page 37 of 49

Figure 4.0: Partial sequencing Electropherogram or the *SLC15A2* gene using primers that covers one of the No-call SNPs; besides that showing other SNPs that are reported in the data base; here isone of these which has a code rs3762819. **A**. The wild type allele shows a conserved G in DM-12-025 sample. Vertical arrow points to the conserved nucleotide G (black peak). **B**. *SLC15A2* allele showing a heterozygous $G \rightarrow A$ in DM-12-010 sample. Vertical arrow points to the changed nucleotide $G \rightarrow A$ (black/green peak). **C**. *SLC15A2* allele showing homozygous $G \rightarrow A$ in DM-10-060 sample. Vertical arrow points to the changed nucleotide G (black gene sequence is according to ensemble database: http://asia.ensembl.org/Homo_sapiens/Transcript/Exons?db=core;g=ENSG00000163406;r =3:121612936-121662949:t=ENST00000489711

Figure 4.68: Partial sequencing Electropherogram of the SLC15A2 gene using primers that covers one of the No-call SNPs; besides that showing other SNPs that are reported in the data base; here is one of these which has a code rs1882002. **A.** The wild type allele shows a conserved C in DM-10-047 sample. Vertical arrow points to the conserved nucleotide C (blue peak). **B.** *SLC15A2* allele showing a heterozygous C \rightarrow T in DM-10-029 sample. Vertical arrow points to the changed nucleotide C \rightarrow T (blue/red peak). **C.** *SLC15A2* allele showing homozygous C \rightarrow T in DM-10-060 sample. Vertical arrow points to the changed nucleotide T (red peak). **D.** Part of reference sequence of *SLC15A2* gene intron; red front base pair represents the location of the nucleotide change. The above reference *SLC15A2* gene sequence is according to ensemble database: http://asia.ensembl.org/Homo_sapiens/Transcript/Exons?db=core;g=ENSG00000163406;r =3:121612936-121662949;t=ENST00000489711

Page 38 of 49

above reference SLC22A1 gene sequence is according to ensemble database which this variant under rs113569197: http://asia.ensembl.org/Homo_sapiens/ Transcript/Exons?db=core;g=ENSG00000175003;r

=6:160542821-160579750;t=ENST00000366963

Page 39 of 49

Figure 4.72: Partial sequencing Electropherogram of the *UGT2B7* gene using primers that covers one of the No-call SNPs; besides that showing other variations that are new and not reported in data base; C.-209T>G. **A.** The wild type allele shows a conserved G in DM-12-035 sample. Vertical arrow points to the conserved nucleotide G (black peak) in position C.-209. **B.** *UGT2B7* allele showing a heterozygous $G \rightarrow T$ in DM-12-035 sample. The vertical arrow points to the changed nucleotide $G \rightarrow T$ (black/red peak) in same position. **C.** Part of reference sequence of *UGT2B7* gene promoter; red front base pair represents the location of the nucleotide change C.-209T position. The above reference *UGT2B7* gene sequence is according to ensemble database: http://asia.ensembl.org/Homo_sapiens/Transcript/Exons?db=core;g=ENSG0000171234;r =4:69917081-69978705;t=ENST00000305231

c.252G > T is a SLC22A6 Variant located in exon1 region that changes G to T at position c.252 (Figure 4.73). Sample code number DM-12-053 was heterozygous (GT). While other samples that were screened had wild type genotypes (GG). This variation doesn't change the amino acid CCG (Ser) \rightarrow CCT (Ser) (synonymous mutation) at codon 84. c.1277 + 69°C > T and c.1277 + 82°C > T are SLC22A1 variants located close to each other in the intron7 region that changes C to T at both position c.1277 + 69 and c.1277 +82 (Figure 4.74). Samples codes DM-12-027, DM-12-021, DM-12-028, DM-12-043 and DM-12-060 were heterozygous (CT) in both variants that looked associated with each other and DM- 12-025 in the 2ND variant only, while other samples that were screened had wild types genotypes (CC) in both variants again. $-698^{\circ}C > A$ is a CYP1A2 variant located in promoter region that changes C to A at position C.-209 (Figure 4.75). Sample code number DM-12-026 and DM-10-015 were heterozygous (CA), while other samples that were screened had wild type genotypes (CC). We still do not know if this new variation has influences on protein level

c.1356T > C is a SLC15A1 Variant located in exon17 region that changes T to C at position c.1356 (Figure 4.76). Sample code number DM-10-040 was heterozygous (TC). While other samples that were screened, had wild type genotypes (CC). That variation doesn't change the amino acid GAC (Asp) \rightarrow GAT (Asp) (synonymous mutation) at codon 453. c.58G > A is a NAT2 Variant located in exon2 region that changes G to A at position c.58 (Figure 4.77). Sample code number

DM-10-005 was heterozygous (GA). While other samples that were screened, had wild type genotypes (GG). That variation changed the amino acid GAC (Asp) \Rightarrow AAC (Asn) (non-synonymous mutation) at codon 20.

Page 40 of 49

c.31G > A is a NAT2 Variant located in the exon2 region that changes G to A at position c.31 (Figure 4.78). Sample code number DM-10-005, DM-10-002, DM-10-014 and DM- 10-057 were heterozygous (GA). While other samples that were screened, had wild type genotypes (GG). That variation changed the amino acid GGC (Gly) \Rightarrow AGC (Ser) (non- synonymous mutation) at codon 11.

Data Analysis for validation

This study has a validation purpose and has been done by three sources: by call genotypes that are close to No-call genotypes and share the same reactions (Appendix B), by SNPs that appear in the reaction and have probes markers in Affymetrix^{*} DMET^{TMM} plus platform (Affymetrix^{*}, Santa Clara, CA, USA) (Table 4.5) and by sequencing call genotypes directly (Table 4.5), to get about 375 call genotypes which score 370 (98.93%) with 4 wrong call genotypes (error rate is 1.07%). [Figure 4.1-4.78] [Table 4.1-4.7]

Discussion

In a high-throughput GeneChip array which based on hybridization with allele- specific probes; genotyping errors is very common which

that are new and not reported in the data base; c.252G>T shows here a silent mutation. **A.** The wild type allele shows a conserved G in DM-10-053 sample. Vertical arrow points to the conserved nucleotide G (black peak) in position c.252G. **B.** *SLC22A6* allele showing a heterozygous $G \rightarrow T$ in DM-10-053 sample. Vertical arrow points to the changed nucleotide $G \rightarrow T$ (black/red peak) in same position. **C.** Part of reference sequence of *SLC22A6* gene promoter; red front base pair represents the location of the nucleotide change c.25G2. The above reference *SLC22A6* gene sequence is according to ensemble database: http://asia.ensembl.org/Homo_sapiens/Transcript/Exons?db=core;g=ENSG0000197901;r = 11:62703857-62752455;t=ENST00000377871

Page 41 of 49

=6:160542821-160579750:t=ENST00000366963

that are new and not reported in the data base; -698C>G(Promoter) showing here. **A.** The wild type allele shows a conserved Cin DM-10-015 sample. Vertical arrow points to the conserved nucleotide C (blue peak) in position -698C. **B.** *CYP1A2* allele showing a heterozygous C \rightarrow Gin DM-10-026 sample. Vertical arrow points to the changed nucleotide C (blue/black peak) in same position. **C.** Part of reference sequence of *CYP1A2* gene promoter; red front base pair represents the location of the nucleotide change-698C position. The above reference CYP1A2 gene sequence is according to ensemble database: http://asia. ensembl.org/Homo_sapiens/Transcript/Exons?db=core;g=ENSG00000140505;r =15:75041185-75048543;t=ENST00000343932

Page 42 of 49

Figure 4.77: Partial sequencing Electropherogram of the *NAT2* gene using primers that covers one of the No-call SNPs; besides that showing another variation that is new and not reported in the data base; c.58G>A(D20N) showing here as silent mutation. **A.** The wild type allele shows a conserved G in DM-10-005 sample. Vertical arrow points to the conserved nucleotide G (black peak) in position c.58G. **B.** *NAT2* allele showing a heterozygous G→Ain DM-10-005 sample. Vertical arrow points to the changed nucleotide G→A (black/green peak) in same position. **C.** Part of reference sequence of *NAT2* gene promoter; red front base pair represents the location of the nucleotide change c.58G position. The above reference *NAT2* gene sequence is according to ensemble database: http://asia.ensembl.org/Homo_sapiens/Transcript/Exons?db=core;g=ENSG0000156006;r = 8:18248755-18258728;t=ENST00000286479

Mol Biol ISSN: 2168-9547 MBL, an open access journal

=13:99336055-99404908;t=ENST00000376503

the changed nucleotide G-A (black/green peak) in same position. C. Part of reference sequence of NAT2 gene promoter; red front base pair represents the location

of the nucleotide change c.31G position. The above reference NAT2 gene sequence is according to ensemble database: http://asia.ensembl.org/Homo_sapiens/ Transcript/Exons?db=core;g=ENSG00000156006;r

=8:18248755-18258728;t=ENST00000286479

	Sample Code	Ref/Var	Call	In Seq	Validation
1	DM10-C1N_(DMET_Plus).dmet.chp	C/A	C/C	CIC	Correct
2	DM-10-020_(DMET_Plus).dmet.chp	C/A	C/C	CIC	Correct
3	DM-10017R_(DMET_Plus).dmet.chp	C/A	C/C	CIC	Correct
4	DM-10-025_(DMET_Plus).dmet.chp	C/A	C/A	C/A	Correct
5	DM-10-026_(DMET_Plus).dmet.chp	C/A	C/C	CIC	Correct
6	DM-10028R_(DMET_Plus).dmet.chp	C/A	C/C	CIC	Correct
7	DM-10030R_(DMET_Plus).dmet.chp	C/A	C/C	CIC	Correct
8	DM-10037_(DMET_Plus) dmet.chp	C/A	C/A	C/A	Correct
9	DM10-040N_(DMET_Plus) dmet.chp	C/A	CC	CIC	Correct
10	DM10-047N_(DMET_Phus).dmet.chp	C/A	C/C	CIC	Correct
11	DM10-054_(DMET_Plus) dmet.chp	C/A	CIC	CC	Correct
12	DM12-012_(DMET_Plus).dmet.chp	C/A	C/A	C/A	Correct
13	DM-10-002_(DMET_Plus).dmet.chp	C/A	C/C	CIC	Correct
14	test dm10-007_(DMET_Plus) dmet.chp	C/A	CC	CIC	Correct
15	DM-10014R_(DMET_Plus) dmet.chp	C/A	C/C	CIC	Correct
16	DM-10-015_(DMET_Plus).dmet.chp	C/A	CC	CIC	Correct
17	DM-10019R_(DMET_Phas).dmet.chp	C/A	C/C	CIC	Correct
18	DM-10027R_(DMET_Plus).dmet.chp	C/A	CIC	CIC	Correct
19	DM-10029R_(DMET_Plus) dmet.ehp	C/A	CC	CC	Correct
20	DM-10035_(DMET_Plus).dmet.chp	C/A	C/C	CIC	Correct
21	DM10-041N_(DMET_Plus).dmet.chp	C/A	CIC	CIC	Correct
22	DM10-043N_(DMET_Plus).dmet.chp	C/A	CIC	CIC	Correct
23	DM10-052_(DMET_Plus).dmet.chp	C/A	CC	CIC	Correct
24	DM10-046N (DMET Plus).dmet.chp	C/A	C/C	CIC	Correct
25	DM10-053_(DMET_Plus) dmet.chp	C/A	CIC	CIC	Correct
26	DM10-060 (DMET Plus) 2.dmet.chp	C/A	CIC	CIC	Correct
27	DM11-015_(DMET_Plus) dmet.chp	C/A	CC	CC	Correct
28	DM11-017 (DMET_Plus).dmet.chp	C/A	C/C	CIC	Correct
29	DM-10-016r_(DMET_Plus) dmet.chp	C/A	C/C	CC	Correct
30	DM-10031R_(DMET_Plus) dmet.chp	C/A	CC	CC	Correct
41	Dm12-47 (DMET Phy) dnet cho	C/T	T/T	CT	Incorrect

Page 44 of 49

Sample Code	Probe Set ID	Ref	Var	Call	Seq	Valid
DM-10-041_(DMET_Plus).dmet.chp	AM_10783	Т	G	T/T	T/T	Correct
DM-10-052_(DMET_Plus).dmet.chp	AM_10783	т	G	T/T	T/T	Correct
DM-10-015 (DMET_Plus) dmet chp	AM_10783	т	G	T/T	T/T	Correct
DM-11-017_(DMET_Plus).dmet.chp	AM_10783	т	G	T/T	T/T	Correct
DM-10-015_(DMET_Plus).dmet.chp	AM_10783	т	0	T/T	T/T	Correct
DM-12-026_(DMET_Plus).dmet.chp	AM_10783	т	G	T/G	T/G	Correct
DM-12-032_(DMET_Plus).dmst.chp	AM_10783	т	G	T/T	T/T	Correct
DM-10-055_(DMET_Plus).dmet.chp	AM_10783	т	G	T/T	TT	Correct
DM-12-047_(DMET_Plus).dmet.chp	AM_10783	т	G	T/T	TT	Correct
DM-10-023_(DMET_Plus) dmet chp	AM_10783	т	G	T/T	T/T	Correct
DM-10-025 (DMET_Plus) dmet chp	AM_10783	т	G	T/T	TT	Correct
DM-10-051_(DMET_Plus).dmet.chp	AM_10783	т	G	T/T	TT	Correct
DM-10-024_(DMET_Plus).dmet.chp	AM_10785	с	A	C/C	CC	Correct
DM-10-046_(DMET_Plus).dmet.chp	AM_10785	с	A	C/C	C/C	Correct
DM-12-001_(DMET_Plus).dmet.chp	AM_10785	c	A	c/c	C/C	Correct
DM-12-023_(DMET_Plus).dmet.chp	AM_10785	с	A	c/c	C/C	Correct
DM-12-033_(DMET_Plus) dmet chp	AM_10785	с	A	C/C	CC	Correct
DM-10-056 (DMET_Plus) dmet chp	AM_10785	с	А	C/C	CC	Correct
DM-10-053_(DMET_Plus).dmet.chp	AM_10785	с	A	C/C	CC	Correct
DM-12-029_(DMET_Plus).duret.chp	AM_10785	с	A	C'A	CA	Correct
DM-10-043 (DMET_Plus).dmet.chp	AM_10785	с	л	C/A	CA	Correct
DM-12-008_(DMET_Plus).dmet.chp	AM_10785	с	л	C/A	CA	Correct
DM-10-034_(DMET_Plus) dmet chp	AM_10785	с	A	C/A	CA	Correct
DM-10-041 (DMET_Plus) dmet chp	AM_10785	с	A	C/A	C/A	Correct
DM-10-052 (DMET_Plus) dmet chp	AM_10785	C	A	C/A	C/A	Correct
DM-10-015 (DMET_Pins) dm# chp	AM_10785	с	A	C'A	C/A	Correct
DM-11-017 (DMET_Plus) dmet chp	AM_10785	с	A	A/A	A/A	Correct
DM-10-016 (DMET Plus) dust chp	AM 10785	с	л	A/A	A/A	Correct
DM-12-026 (DMET_Plus).dmet.chp	AM_10785	с	л	A/A	AA	Correct
DM-12-032 (DMET Plus).dmet.chp	AM 10785	с	Α	Λ/Λ	AA	Correct
DM-10-055 (DMET_Plus) duret chp	AM_10785	С	Λ	A/A	A/A	Correct
DM-12-047 (DMET Plus) dmet chp	AM 10785	C	A	A/A	AA	Correct
DM-10-023 (DMET Plus) dmit chp	AM 10785	C	A	A/A	AA	Correct
DM-10-025 (DMET Pins) dmrt chp	AM 10785	c	A	A/A	AA	Correct
DM-10-051 (DMET Phy) dust cho	AM 10785	C	A	A/A	AlA	Conect

Table 4.5b: Validation results for SNPs that have probe markers in Affymetrix[®] DMETT[™] plus platform.

Probe Set ID	Ref	Var	Call	Seq	Valid
AM_10784	С	т	CC	CC	Correct
AM_10784	с	т	CC	CC	Correct
AM_10784	с	т	CC	CC	Correct
AM_10784	С	т	OC	CC	Correct
AM_10784	с	т	CC	CC	Corpect
AM_10784	с	т	CC	CC	Correct
AM_10784	с	т	CC	CIC	Corpect
AM_10784	с	т	CC	CC	Conect
AM_10784	С	т	OC	CC	Correct
AM_10784	с	т	cc	CC	Correct
AM_10784	с	т	CC	CC	Correct
AM_10784	с	т	CC	CC	Comect
AM_10784	с	т	oc	CC	Corpect
AM_10784	С	т	CC	CC	Correct
AM_10784	с	т	cc	CIC	Correct
AM_10784	с	т	oc	cc	Correct
AM 10784	c	т	oc	CC	Correct
AM_10784	с	т	cc	CC	Correct
AM 10784	с	т	CC	CC	Correct
AM_10784	с	т	cic	CC	Correct
AM 10784	с	т	CC	CC	Correct
AM_10784	c	т	CC	CIC	Correct
AM 10784	с	т	CC	CC	Contect
AM 10783	т	G	T/T	TT	Correct
AM 10783	т	G	T/T	TT	Correct
AM 10783	т	G	TT	TT	Correct
AM 10783	т	G	T/T	TT	Correct
AM 10783	т	G	T/T	TT	Comect
AM 10783	т	G	T/T	TT	Correct
AM 10783	т	G	T/T	IT	Comect
AM 10783	т	0	T/T	TT	Correct
AM 10783	т	G	T/T	TT	Correct
AM 10783	т	G	T/T	TT	Corpect
	ID AM_10734 AM_10734 AM_10734 AM_10734 AM_10734 AM_10734 AM_10734 AM_10734 AM_10734 AM_10734 AM_10734 AM_10734 AM_10734 AM_10734 AM_10734 AM_10734 AM_10733 AM	ID Ref AM_10734 C AM_10735 T AM_10733 T AM_10733 T AM_10733 T AM_10733 T	ID Ref Var AM_10734 C T AM_10733 T G AM_10733 T G AM_10733 T	ID Ref Var Call AM_10734 C T CC AM_10734 C T CC	Ref Var Call seq AM_10734 C T C/C C/C AM_10734 C <td< td=""></td<>

Page 45 of 49

	SNPs code	Probe Set ID	Call rates	No-call rates	Call rates*	No-call rates*	
1	rs2072671	AM_11499	74.26%	25.74%	82.18%	17.82%	
2	rs7662029	AM_13458	69.31%	30,69%	84.16%	15.84%	
3	ns7668258	AM 13459	99.01%	0.99%	99.01%	0.99%	
4	rs10929303	AM 13067	96.04%	3.96%	97.03%	2.97%	
5	rs8330	AM 13070	86,14%	13.86%	93.07%	6.93%	
6	rs1042640	AM_13068	90.10%	9.90%	95.05%	4.93%	
2	rs1143670	AM_13301	78.22%	21.78%	95.05%	4.95%	
8	rs6661174	AM_11958	96.04%	3.96%	100%	0%	
9	rs45564337	AM_10347	85,15%	14.85%	95.05%	4.9.5%	
10	rs1142345	AM_13973	95.05%	4.95%	99.01%	0.99%	
11	rs1800460	AM_13900	92.08%	7.92%	96.04%	3.96%	
12	rs12208357	AM_14348	96.04%	3.96%	99.01%	0.99%	
13	n55918055	AM_14350	92.08%	7.9294	96.04%	3.96%	
14	rs45512393	AM_14361	94.06%	5.94%	100%	0%	
15	rs628031	AM_14363	80,20%	19.00%	97.03%	2.97%	
16	rs1043943	AM_10768	90,10%	9,90%	94.06%	5.94%	
17	rs1799814	AM_10769	92.08%	7.92%	98.02%	1.98%	
18	1941279188	AM_10766	91.09%	8,9196	93.07%	6.93%	
19	rs72547510	AM_10771	91.09%	8.91%	96.04%	3.96%	
20	rs2069526	AM_10783	80.20%	19.80%	92.08%	7.92%	
21	rs762551	AM_10785	87.13%	12.87%	92.08%	7.92%	
22	rs12720461	AM_10784	93.07%	6.93%	94.06%	5.94%	
23	rs1805158	AM_14998	96.04%	3,96%	98.02%	1.98%	
24	rs1041963	AM_15000	95,05%	4.95%	99.01%	0.99%	
25	rs1801280	AM_15001	76.24%	23.76%	95.05%	4.93%	
26	rs72554616	AM_15005	96.04%	3.96%	99.01%	0.99%	
27	rs1799929	AM_15006	90.10%	9.90%	94.06%	5.94%	
28	1572554617	AM_15007	95.05%	4.95%	97.03%	2.97%	
29	rs1799930	AM_15008	96.04%	3,96%	100%	0%	
30	rs1208	AM_15010	96.04%	3,96%	98.02%	1.98%	
31	n1339067	AM_10650	73.27%	26.73%	93.07%	6.93%	
32	ps2274827	AM_10647	65.35%	34.65%	93.07%	6.93%	
33	rs2235033	AM_14609	95.05%	4.95%	100%	0%	
34	rs1128503	AM_14612	68.32%	31.68%	85.15%	14.85%	
35	rs3213619	AM_14633	92.08%	7.92%	93.07%	6.93%	
36	rs9282564	AM_14628	98.02%	1.98%	99.01%	0.99%	
37	ni2032582	AM_14592	98.02%	1.98%	99.01%	0.99%	
38	rs72552713	AM_13689	98.02%	1.93%	99.01%	0.99%	
39	rs72547529	AM_11044	97.03%	2.97%	98.02%	1.98%	
	Avera	P751	89.08%	10.92%	95.56%	4.44%	

Table 4.7: Call rates improvement details in each marker probes.

limited the technology application; in addition No-call genotypes for many SNPs on the chip immerge as bigger and more serious problem in high-throughput genotyping methods which currently using at PHBC with Affymetrix^{*} DMET^{TMM} plus platform (Affymetrix^{*}, Santa Clara, CA, USA) to genotyped 1936 SNPs in Jordanian samples.

Total of 101 samples were genotyping through this GeneChip type that specific for most important genes variations as pharmacogenomics study to determines the haplotypes of the major of reported SNPs clusters like these are functional in metabolizing of the most widely prescribed anticancer drug in the world including aromatase inhibitors, tamoxifen and thiopurines groups.

Accordingly, we were Genotyping these No-call s data by direct sequencing method that gives us an opportunity to complete some of SNPs clusters (haplotypes) which can then determine the exact phenotypes for the study samples weather the phenotype of these samples will be an UM, EM, IM or PM metabolizer. For validation purposes, we carried re-genotyping some of the SNPs which originally reported as a call genotype using the

Affymetrix^{*} DMET^{TMM} plus platform experiments. Twenty different genomic regions in 15 different ADMEs genes (NAT2, ABCB1, ABCG2, SLC15A1, SLC15A2, SLC22A1, SLC22A6, UGT2B7, UGT1A1, CDA, FMO2, CYP1A1, CYP1A2, TPMT and VKORC1) were amplified by PCR reactions and sequenced. We identified 66 different SNPs, 39 of these variation are belong to the No-calls genotypes markers in some samples of Affymetrix^{*} DMET^{TMM} plus results data. In all SNPs; there were no any new variants for these SNPs and all were the reference or the variant SNPs originally identified by Affymetrix^{*} DMET^{TMM} plus chip. In sequencing a total of 273 No-call s; there was no biological reason for the No-call which suggests other problems.

On the other hand, we have found that there are some of No-call genotypes samples have other SNPs close to the original SNP. These SNPs are 25 bp distance upstream or downstream of the No-call SNPs. Fifteen SNPs out of the 39 SNPs tested encounter this problem and have neighboring SNPs (Figure 3.2) probes of the No-call (probes of the No-call SNPs Figure 5.1B). The existence of the target SNPs in proximity to another variants might cause some of the No-call results (Figure 5.1C) [63]. For example, we found in CYP1A1 gene 3 of No-call SNPs located very close to each other (Figure 5.1A). Several samples have No-call s genotypes in all of these 3 SNPs, which resulted in weaken the intensity of fluorescence signals over the background and prevent signing the genotype. SNPs located in proximity to each other's might explain some of the missing data however some other reasons including poor quality DNA samples, problem in Cut-off percentage of the chip data and triallelic variations are also reported [87].

Our validation study shows in the No-call genotypes; 47.25% of the No-call s was for wild-type alleles, 37% were for heterozygous and 15.75% for homozygous variant. Where for the call allele validation the wild-type alleles corrected were 87.35%, heterozygous 9.57% and 3.08% homozygous variant respectively (Figure 5.2).

To overcome this problem; one way is to increase the Cut-off percentage for the chip; although this would decrease the samples numbers. The re-genotyping of No-call will be practical to achieve final conclusion of the core ADMEs genes.

Besides genotyping the No-call SNPs, this study have another purpose that focusing on validation of this new revolutionary techniques that still be as a research tool and not for diagnostic or clinical purpose. From 375 Call genotypes, DMET^{TMM} plus platform (Affymetrix^{*}, Santa Clara, CA, USA) is scores 371 (98.93%) with 4 errors (1.07%), which is

above accepted error rate (less than 1%) [10,88,89]. The bias in No-call genotype/alleles and the rate of error that above 1% are may be resulting from cut-off of QC that used in PHBC (87%). Therefore, we suggested that the ongoing QC cut-off for call-rate should be increased, while the cut-off for genotyping error rate can be reduced properly [10].

Furthermore, study identified 8 novel variant located in this important genes: c- 209T > G in UGT2B7 and -69°8C > A in CYP1A2 are non-reported promoter variants located close to SNPs that Affymetrix[®] DMETT^{MM} gene chip has probe for them. c.252G > T in SLC22A6 and c.1356T > C in SLC15A1 are non-reported exonic variants which are silent mutations (synonymous) that no effect on protein level. c.1277 + 69°C > T and c.1277 + 82°C > T in SLC22A1 are intronic variants, its associate to each other in same 5 samples, which look as markers for Jordanian, and this need more investigation to look for which subpopulation are associate with.

The most important Novel variants that our study discover are c.58G > A and c.31G > A in NAT2 gene as non- synonymous or missense variants (1st one D20N and 2ND one G11S) which NAT2 gene encodes an enzyme that function to activate or deactivate aryl amine and hydrazine drugs and carcinogens [90]. Polymorphisms in this gene are associated with higher incidences of cancer and drug toxicity [91]. These non-reported variants repeat on 4 different samples (four alleles for 1st and one allele for 2nd) and appear associate to each other in one sample which may be recorded as new haplotype.

Therefore, these non-reported variants can be change the whole haplotype and phenotypic conclusion that chip determined, these will be a very clear after flow cloning procedure for variants to determine their location in which allele. For instance, in sample DM10-005, there are two new variants in NAT2 gene in coding region and non-synonymous that change 2 amino acids. According reported SNPs, and specific allele hybridization that DMETT^{MM} plus platform (Affymetrix^{*}, Santa Clara, CA, USA), the haplotype is *5/*5 and phenotype for this sample is SA (Slow acetylator), But when include the new variants which still don't know yet about the association to other SNPs and with each other (Table 5.1).

Conclusion

- This study identified 66 variations in 15 ADMEs genes (CDA, FMO2, SLC22A6, UGT1A1, UGT2B7, CYP1A1, SLC22A1, NAT2, ABCB1, ABCG2, SLC15A2, VKORC1, CYP1A2, SLC15A1 and TPMT).
- 39 of the variations had No-call genotypes in the Affymetrix DMETT^{MM} plus platform (Affymetrix, Santa Clara, CA, USA), which improved call rates in these markers from 89.08% to 95.56% in 101 samples and that will help to get a more accurate conclusion about Jordanian haplotypes in these genes. All of the No-call genotypes SNPs in this study (39 SNPs) are reported, and this means that our study which includes about 273 No-call s genotypes didn't find any relation between the

Page 47 of 49

```
Gene
          Variant
                        Sample
                                   Haplotype
                                                         Phenotyp
UOT287
          C-209T>G
                        DM-12-035
                                   *1a/*2c
                                                         UM or EM
SLC2246 c.252GPT
                       DM-10-053 NoRAP
SLC2241 c.1277+82C>T DM-10-027 NoHAP
          c.1277+82C>T
                        DM-10-021
                                   NoHAP
                        DM-10-028
                                   NoHAP
                        DM-10-043 NoRAP
                        DM-10-060
                                   NoRAP
                        DM-10-025 NoRAP
CTP142
                        DM-10-026 *1E-1E
                                                         EM
          -600C>A
                        DM-10-015 *LA/*1F,*1F/*6,*1F/*7
                                                         EM.EM or IM.IM
SEC1547 c.1356T>C
                        DM-10-040 *L/*4.*L/*4B.*L/*12.
                                   *3/*4.*3/*48.*3/*12.
                                   *3B/*4,*3B/*4B.*3B/*12.
                                   +4/*5,*4/*21,*418*5
                                   *48/*21,*5/*12,*12/*21
NAT2
                        DM-10-005 *5/*5
          c.580>A
          c.31G>A
                        DM-10-002 +4/+5,+5/+17
                                                         LA.SA
                        DM-10-014
                                  +4+4.+4+5,+4+52
                                                         RAIAIASASA
                                   *4/*6,*5/*5,*5/*5E
                                                          SA.SA.SA
                                   +5/*6.*5E/*5E.*5E/*6
                        DM-10-057 +4/+5
                                                         EA.
```

Table 5.1: Novel variants and their sample currently haplotype/phenotype.

missing genotypes in the high-throughput GeneChip and Novel variations in the same positions that the probes binds with. Also, we found that 15 of these SNPs are neighbored by other variations upstream/downstream and this may be one reason for No-call out-puts.

- Among these variations, 8 were non-reported: C.-209T > G in UGT2B7 and 698°C > A in CYP1A2 as promoter variants, c.252G > T in SLC22A6 and c.1356T > C in SLC15A1 as silent mutation or synonymous variants, c.1277 + 69°C > T and c.1277 + 82°C > T in SLC22A1 as intronic variants associated with each other and the most important variations were two missense (non- synonymous) in NAT2 gene: D20N c.58G > A and G11S c.31G > A.
- The other variations were reported so some of them were used to validate the Affymetrix DMET^{TMM} plus platform.
- Error rate was 1.07% which was above the accepted rate (less than 1%), and there was a clear bias between call and No-call genotypes (after re-genotyping). This will be eliminated just when the increase current cuts-off.

Recommendations

After analyzing and reviewing the results of the current study I recommend to:

- Increase the current cut-off percentage of QC of chips to decrease the error rates and bias conclusion.
- Focus on core markers in the Affymetrix^{*} DMETT^{MM} plus platform (Affymetrix^{*}, Santa Clara, CA, USA) as a priority. This will help in managing how to deal and benefit from this high through-out technique.
- Eliminate DNA quality factors that increase No-call rates in same samples, this also will decrease the numbers of missing genotypes, even after increasing the cut- off percentage.
- Flowing Cloning procedure technique is used to determine

Novel variant association alleles, to obtain new haplotypes that change the final conclusions for Jordanian phenotypic groups for ADME genes.

- Continue to screen other DMET Genes (especially gene associations with ADRs in Jordanians) by direct sequencing of the Affymetrix^{*} DMET^{TMM} plus platform data to complete the gap in the No-call genotypes to know the phenotypics.
- This performed study can be extended to include Jordanians from different ethnic origins like (Circassians, Chechens, Bedouins, Gypsy, etc.) as sub-population studies to looks how these groups differ from the main population groups.
- Eight novel mutations that were identified in this study needed to be studied to determine whether they are neutral polymorphisms or clinically relevant mutations.
- Due to that, further analysis is important here (such as, functional analysis and splicing assays) which is required to investigate their effects on both the RNA and protein level.
- Continue screening of DMET genes, even outside the Affymetrix[°] DMETT^{MM} plus platform data project, when these genes look highly polymorphic in Jordanians, and this is very important to discover the differences between us and other populations, which has influences on our drug responses and therapies.
- Testing drugs metabolism, elimination and transporters through the DMETT^{MM} plus platform technique is very helpful to determine suspected individuals with ADRs or non-response people, so that association studies can be included in the same project.

Reference

 Yoshizawa M, Hayashi H, Tashiro Y, Sakawa S, Moriwaki H, et al. (2009) Effect of VKORC1-1639 G>A polymorphism, body weight, age, and serum albumin alterations on warfarin response in Japanese patients. Thromb Res 124: 161-166.

- Kang JS, Lee MH (2009) Overview of therapeutic drug monitoring. Korean J Intern Med 24: 1-10.
- Cooper GM, Johnson JA, Langaee TY, Feng H, Stanaway IB, et al. (2008) A genome-wide scan for common genetic variants with a large influence on warfarin maintenance dose. Blood 112: 1022-1027.
- Hamdy S, Hiratsuka M, Narahara K (2002) Allele and genotype frequencies of polymorphic cytochromes P450 (CYP2C9, CYP2C19, CYP2E1) and dihydropyrimidine dehydrogenase (DPYD) in the Egyptian population. Br J. Clin Pharmacol 53: 596–603.
- Mizutani T (2003) PM frequencies of major CYPs in Asians and Caucasians. Drug Metab Rev 35: 99-106.
- Wilke RA, Lin DW, Roden DM, Watkins PB, Flockhart D, et al. (2007) Identifying genetic risk factors for serious adverse drug reactions: current progress and challenges. Nat Rev Drug Discov 6: 904-916.
- Nakai K, Tsuboi J, Okabayashi H (2007) Ethnic differences in the VKORC1 gene polymorphism and an association with warfarin dosage requirements in cardiovascular surgery patients. Pharmacogenomics 8: 713–719.
- Lo' pez-Correa C, Gelbert L (2013) Clinical Pharmacogenomics. In: Molecular Genetic Pathology. Edits: Cheng, D and Zhang, J. Springer Science. New York 1009-1030.
- Burmester JK, Sedova M, Shapero MH, Mansfield E (2010) DMET microarray technology for pharmacogenomics-based personalized medicine. Methods Mol Biol 632: 99-124.
- Wenqing, Fu W, Wang Y, Wang Y, Li R, Li R, et al. (2009) Missing call bias in high-throughput genotyping. BMC Genomics 10: 106.
- Daly TM, Dumaual CM, Miao X, Farmen MW, Njau RK, et al. (2007) Multiplex assay for comprehensive genotyping of genes involved in drug metabolism, excretion, and transport. Clin Chem 53: 1222-1230.
- Sissung TM, English BC, Venzon D, Figg WD, Deeken JF (2010) Clinical pharmacology and pharmacogenetics in a genomics era: the DMET platform. Pharmacogenomics 11: 89-103.
- Guzzi PH, Agapito G, Di Martino MT, Arbitrio M, Tassone P, et al. (2012) DMETanalyzer: automatic analysis of Affymetrix DMET data. BMC Bioinformatics 13: 258.
- 14. Shastry BS (2006) Pharmacogenetics and the concept of individualized medicine. Pharmacogenomics J 6: 16-21.
- 15. Huang RS, Ratain MJ (2009) Pharmacogenetics and pharmacogenomics of anticancer agents. CA Cancer J Clin 59: 42-55.
- 16. Dervieux T, Meshkin B, Neri B (2005) Pharmacogenetic testing: proofs of principle and pharmacoeconomic implications. Mutat Res 573: 180-194.
- 17. Marsh S, McLeod HL (2004) Pharmacogenetics of irinotecan toxicity. Pharmacogenomics 5: 835-843.
- Wilson JF, Weale ME, Smith AC, Gratrix F, Fletcher B, et al. (2001) Population genetic structure of variable drug response. Nat Genet 29: 265-269.
- Borgiani P, Ciccacci C, Forte V, Sirianni E, Novelli L, et al. (2009) CYP4F2 genetic variant (rs2108622) significantly contributes to warfarin dosing variability in the Italian population. Pharmacogenomics 10: 261-266.
- Roper N, Storer B, Bona R, Fang M (2010) Validation and comparison of pharmacogenetics-based warfarin dosing algorithms for application of pharmacogenetic testing. J Mol Diagn 12: 283-291.
- Thacker SM, Grice GR, Milligan PE, Gage BF (2008) Dosing anticoagulant therapy with coumarin drugs: is genotyping clinically useful? Yes J Thromb Haemost 6: 1445-1449.
- Need AC, Motulsky AG, Goldstein DB (2005) Priorities and standards in pharmacogenetic research. Nat Genet 37: 671-681.
- Sadée W, Dai Z (2005) Pharmacogenetics/genomics and personalized medicine. Hum Mol Genet 14 Spec No.
- Burroughs V, Maxey R, Levy R (2002) Racial and ethnic differences in response to medicines: towards individualized pharmaceutical treatment. J Natl Med 94: 1-26.
- Xie HG, Kim RB, Wood AJ, Stein CM (2001) Molecular basis of ethnic differences in drug disposition and response. Annu Rev Pharmacol Toxicol 41: 815-850.

 Burchard EG, Ziv E, Coyle N, Gomez SL, Tang H, et al. (2003) The importance of race and ethnic background in biomedical research and clinical practice. N Engl J Med 348: 1170-1175.

Page 48 of 49

- Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, et al. (2001) Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. Science 294: 1719-1723.
- Flockhart DA, Tanus-Santos JE (2002) Implications of cytochrome P450 interactions when prescribing medication for hypertension. Arch Intern Med 162: 405-412.
- 29. Craig C and stitzel R (2003) Modern Pharmacology with Clinical Applications. BrownCo. Boston, 832.
- Bosch TM, Meijerman I, Beijnen JH, Schellens JH (2006) Genetic polymorphisms of drug-metabolising enzymes and drug transporters in the chemotherapeutic treatment of cancer. Clin Pharmacokinet 45: 253-285.
- Lacaná E1, Amur S, Mummanneni P, Zhao H, Frueh FW (2007) The emerging role of pharmacogenomics in biologics. Clin Pharmacol Ther 82: 466-471.
- Frueh FW, Gurwitz D (2004) From pharmacogenetics to personalized medicine: a vital need for educating health professionals and the community. Pharmacogenomics 5: 571-579.
- 33. Frueh FW, Amur S, Mummaneni P, Epstein RS, Aubert RE, et al. (2008) Pharmacogenomic biomarker information in drug labels approved by the United States food and drug administration: prevalence of related drug use. Pharmacotherapy 28: 992-998.
- Akey J, Jin L, Xiong M (2001) Haplotypes vs single marker linkage disequilibrium tests: what do we gain? Eur J Hum Genet 9: 291-300.
- Harding R, Fullerton S, Griffiths R, Bond J, Cox M, et al. (1997) Archaic African and Asian lineages in the genetic ancestry of modern humans. Am J. Hum Genet 60: 772–789.
- 36. Kidd JR, Pakstis AJ, Zhao H, Lu RB, Okonofua FE, et al. (2000) Haplotypes and linkage disequilibrium at the phenylalanine hydroxylase locus, PAH, in a global representation of populations. Am J Hum Genet 66: 1882-1899.
- 37. Tishkoff SA, Pakstis AJ, Ruano G, Kidd KK (2000) The accuracy of statistical methods for estimation of haplotype frequencies: an example from the CD4 locus. Am J Hum Genet 67: 518-522.
- Dai JY, Ruczinski I, LeBlanc M, Kooperberg C (2006) Imputation methods to improve inference in SNP association studies. Genet Epidemiol 30: 690-702.
- Botstein D, Risch N (2003) Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat Genet 33 Suppl: 228-237.
- 40. Fallin D, Cohen A, Essioux L, Chumakov I, Blumenfeld M, et al. (2001) Genetic analysis of case/control data using estimated haplotype frequencies: application to APOE locus variation and Alzheimer's disease. Genome Res 11: 143-151.
- 41. Douglas JA, Boehnke M, Gillanders E, Trent JM, Gruber SB (2001) Experimentally-derived haplotypes substantially increase the efficiency of linkage disequilibrium studies. Nat Genet 28: 361-364.
- Michalatos-Beloin S, Tishkoff SA, Bentley KL, Kidd KK, Ruano G (1996) Molecular haplotyping of genetic markers 10 kb apart by allele-specific longrange PCR. Nucleic Acids Res 24: 4841-4843.
- Tregouet DA, Escolano S, Tiret L, Mallet A, Golmard JL (2004) A new algorithm for haplotype-based association analysis: the Stochastic-EM algorithm. Ann Hum Genet 68: 165-177.
- Zhang K, Sun F, Zhao H (2005) HAPLORE: a program for haplotype reconstruction in general pedigrees without recombination. Bioinformatics 21: 90-103.
- 45. Catanzaro D and Labb M (2009) The pure parsimony haplotyping problem: overview and computational advances. Intl. Trans. in Op. Res 16: 561–584.
- 46. Tishkoff S, Pakstis A, Stoneking M, Kidd J, Destro-Bisol G, et al., (2000) Short tandem-repeat polymorphism/alu haplotype variation at the PLAT locus: implications for modern human origins. Am J. Hum Genet 67: 901–925.
- Aspinall MG, Hamermesh RG (2007) Realizing the promise of personalized medicine. Harv Bus Rev 85: 108-117, 165.
- Lyssenko V, Jonsson A, Almgren P, Pulizzi N, Isomaa B, et al. (2008) Clinical risk factors, DNA variants, and the development of type 2 diabetes. N Engl J Med 359: 2220-2232.

Page 49 of 49

- Meigs JB, Shrader P, Sullivan LM, McAteer JB, Fox CS, et al. (2008) Genotype score in addition to common risk factors for prediction of type 2 diabetes. N Engl J Med 359: 2208-2219.
- Wadelius M, Pirmohamed M (2007) Pharmacogenetics of warfarin: current status and future challenges. Pharmacogenomics J 7: 99-111.
- Strachan T, Read A (2011) Human Molecular Genetics. Grand Science. New York 4, 781.
- Mega JL, Close SL, Wiviott SD, Shen L, Hockett RD, et al. (2009) Cytochrome p-450 polymorphisms and response to clopidogrel. N Engl J Med 360: 354-362.
- 53. Weinshilboum R (2003) Inheritance and drug response. N Engl J Med 348: 529-537.
- Zanger UM, Turpeinen M, Klein K, Schwab M (2008) Functional pharmacogenetics/genomics of human cytochromes P450 involved in drug biotransformation. Anal Bioanal Chem 392: 1093-1108.
- 55. Zhao L, Li S, Khalid N (2003) Amethod for the assessment of disease associations with single-nucleotide polymorphism haplotypes and environmental variables in case-control studies. Am J. Hum Genet 72: 1231–1250.
- Evans WE, McLeod HL (2003) Pharmacogenomics--drug disposition, drug targets, and side effects. N Engl J Med 348: 538-549.
- Huentelman MJ, Craig DW, Shieh AD, Corneveaux JJ, Hu-Lince D, et al. (2005) SNiPer: improved SNP genotype calling for Affymetrix 10K GeneChip microarray data. BMC Genomics 6: 149.
- Marchini J, Howie B, Myers S, McVean G, Donnelly P (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet 39: 906-913.
- Wang DG, Fan JB, Siao CJ, Berno A, Young P, et al. (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 280: 1077-1082.
- 60. Kruglyak L, Nickerson DA (2001) Variation is the spice of life. Nat Genet 27: 234-236.
- Bolstad BM, Irizarry RA, Astrand M, Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 19: 185-193.
- 62. Li C, Hung Wong W (2001) Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. Genome Biol 2: RESEARCH0032.
- Dumaual C, Miao X, Daly TM, Bruckner C, Njau R, et al. (2007) Comprehensive assessment of metabolic enzyme and transporter genes using the Affymetrix Targeted Genotyping System. Pharmacogenomics 8: 293-305.
- Caldwell MD, Awad T, Johnson JA, Gage BF, Falkowski M, et al. (2008) CYP4F2 genetic variant alters required warfarin dose. Blood 111: 4106-4112.
- 65. Brandt JT, Close SL, Iturria SJ, Payne CD, Farid NA, et al. (2007) Common polymorphisms of CYP2C19 and CYP2C9 affect the pharmacokinetic and pharmacodynamic response to clopidogrel but not prasugrel. J Thromb Haemost 5: 2429-2436.
- 66. Hardenbol P, Yu F, Belmont J, Mackenzie J, Bruckner C, et al. (2005) Highly multiplexed molecular inversion probe genotyping: over 10,000 targeted SNPs genotyped in a single tube assay. Genome Res 15: 269-275.
- Ji H, Kumm J, Zhang M, Farnam K, Salari K, et al., (2006) Molecular inversion probe analysis of gene copy alterations reveals distinct categories of colorectal carcinoma. Cancer Res 66: 7910–7919.
- Wang Y, Moorhead M, Karlin-Neumann G, Falkowski M, Chen C, et al. (2005) Allele quantification using molecular inversion probes (MIP). Nucleic Acids Res 33: e183.
- Wang Y, Moorhead M, Karlin-Neumann G, Wang NJ, Ireland J, et al. (2007) Analysis of molecular inversion probe performance for allele copy number determination. Genome Biol 8: R246.
- Ji H, Welch K (2009) Molecular inversion probe assay for allelic quantitation. Methods Mol Biol 556: 67-87.
- Robarge JD, Li L, Desta Z, Nguyen A, Flockhart DA (2007) The star-allele nomenclature: retooling for translational genomics. Clin Pharmacol Ther 82: 244-248.
- 72. Altshuler D, Daly MJ, Lander ES (2008) Genetic mapping in human disease. Science 322: 881-888.

- O'Connell JR, Weeks DE (1998) PedCheck: a program for identification of genotype incompatibilities in linkage analysis. Am J Hum Genet 63: 259-266.
- Douglas JA, Boehnke M, Lange K (2000) A multipoint method for detecting genotyping errors and mutations in sibling-pair linkage data. Am J Hum Genet 66: 1287-1297.
- Hosking L, Lumsden S, Lewis K, Yeo A, McCarthy L, et al., (2004) Detection of genotyping errors by Hardy-Weinberg equilibrium testing. Eur J. Hum Genet 12: 395-399.
- Leal SM (2005) Detection of genotyping errors and pseudo-SNPs via deviations from Hardy-Weinberg equilibrium. Genet Epidemiol 29: 204-214.
- 77. Matsuzaki H, Dong S, Loi H, Di X, Liu G, et al. (2004) Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. Nat Methods 1: 109-111.
- Yu Z, Garner C, Ziogas A, Anton-Culver H, Schaid DJ (2009) Genotype determination for polymorphisms in linkage disequilibrium. BMC Bioinformatics 10: 63.
- Holt B, Benfer RA (2000) Estimating missing data: an iterative regression approach. J Hum Evol 39: 289-296.
- Su SC, Kuo CC, Chen T (2005) Inference of missing SNPs and information quantity measurements for haplotype blocks. Bioinformatics 21: 2001-2007.
- Li Y, Ding J, Abecasis G (2006) rapid haplotype reconstruction and missing genotype inference. Am J, Human Genet 79: 2290-2299.
- Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68: 978-989.
- Yu Z, Schaid DJ (2007) Methods to impute missing genotypes for population data. Hum Genet 122: 495-504.
- Fallin D and Schork N (2000) Accuracy of haplotype frequency estimation for biallelic loci, via the expectation–maximization algorithm for unphased diploid genotype data. Am J. Hum Genet, 67, 947–959.
- Kelly ED, Sievers F, McManus R (2004) Haplotype frequency estimation error analysis in the presence of missing genotype data. BMC Bioinformatics 5: 188.
- 86. Efros A, Halperin E (2012) Haplotype reconstruction using perfect phylogeny and sequence data. BMC Bioinformatics 13 Suppl 6: S3.
- Huebner C, Petermann I, Browning B, Shelling A, Ferguson L (2007) Triallelic Single Nucleotide Polymorphisms and Genotyping Error in Genetic Epidemiology Studies: MDR1 (ABCB1) G2677/T/A as an Example. Cancer Epidemiol Biomarkers Prev 16: 1185-1192.
- Pompanon F, Bonin A, Bellemain E, Taberlet P (2005) Genotyping errors: causes, consequences and solutions. Nat Rev Genet 6: 847-859.
- Easton D, Pooley K, Dunning A, Pharoah P, Thompson D, et al. (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 447: 1087-1093.
- Sabbagh A, Langaney A, Darlu P, Gérard N, Krishnamoorthy R, et al. (2008) Worldwide distribution of NAT2 diversity: implications for NAT2 evolutionary history. BMC Genet 9: 21.
- Patin E, Harmant C, Kidd KK, Kidd J, Froment A, et al. (2006) Sub-Saharan African coding sequence variation and haplotype diversity at the NAT2 gene. Hum Mutat 27: 720.