3-2017

Genetic Polymorphism of Cytochrome P450-1A2 (CYP1A2) and N-Acetyltransferase-2 (NAT2) Among Emiratis

Mohammed Majed Al-Ahmad

Follow this and additional works at: https://scholarworks.uaeu.ac.ae/all_theses

Part of the Medicine and Health Sciences Commons

Recommended Citation

Al-Ahmad, Mohammed Majed, "Genetic Polymorphism of Cytochrome P450-1A2 (CYP1A2) and N-Acetyltransferase-2 (NAT2) Among Emiratis" (2017). Theses. 642.
https://scholarworks.uaeu.ac.ae/all_theses/642

This Dissertation is brought to you for free and open access by the Electronic Theses and Dissertations at Scholarworks@UAEU. It has been accepted for inclusion in Theses by an authorized administrator of Scholarworks@UAEU. For more information, please contact fadl.musa@uaeu.ac.ae.
GENETIC POLYMORPHISM OF CYTOCHROME P450-1A2 (CYP1A2) AND N-ACETYLCYSTEINE-2 (NAT2) AMONG EMIRATIS

Mohammad Majed Al-Ahmad

This dissertation is submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Under the Supervision of Professor Salim Bastaki

March 2017
Declaration of Original Work

I, Mohammad Majed Al-Ahmad, the undersigned, a graduate student at the United Arab Emirates University (UAEU), and the author of this dissertation entitled "Genetic Polymorphism of Cytochrome P450-1A2 (CYP1A2) and N-Acetyltransferase-2 (NAT2) Polymorphism among Emiratis" hereby, solemnly declare that this dissertation is an original research work that has been done and prepared by me under the supervision of Professor Salim Bastaki, in the College of Medicine and Health Sciences at UAEU. This work is entirely original and has not previously been presented or published, or formed the basis for the award of any academic degree, diploma or a similar title at this or any other university. Any materials borrowed from other sources (whether published or unpublished) and relied upon or included in my dissertation have been properly cited and acknowledged in accordance with appropriate academic conventions. I further declare that there is no potential conflict of interest with respect to the research, data collection, authorship, presentation and/or publication of this dissertation.

Student’s Signature: [signature] Date: 14/5/2017
Advisory Committee

1) Advisor (Committee Chair): Professor Salim Bastaki
   Title: Professor of Pharmacology and Therapeutics
   Department of Pharmacology
   College of Medicine and Health Sciences

2) Co-advisor: Professor Bassam Ali
   Title: Professor of Molecular and Genetic Medicine
   Department of Pathology
   College of Medicine and Health Sciences

3) Member: Professor Yousef Abdulrazzaq
   Title: Professor of Paediatrics
   Department of Paediatrics
   College of Medicine and Health Sciences
Approval of the Doctorate Dissertation

This Doctorate Dissertation is approved by the following Examining Committee Members:

1) Advisor (Committee Chair): Professor Salim Bastaki

Title: Professor of Pharmacology

Department: Pharmacology

College: Medicine and Health Sciences

Signature ____________________________ Date 27/3/2017

2) Member: Professor Bassam Ali

Title: Professor of Molecular and Genetic Medicine

Department: Pathology

College: Medicine and Health Sciences

Signature ____________________________ Date 27/3/2017

3) Member: Assistant Professor Suraiya Anjum Ansari

Title: Assistant Professor of Biochemistry

Department: Biochemistry

College: Medicine and Health Sciences

Signature ____________________________ Date 27/3/2017
4) Member (External Examiner): Professor Ann Katherine Daly

Title: Professor of Pharmacogenetics

Department: Institute of Cellular Medicine

Institution: Newcastle University, UK

Signature [Signature] Date 27/3/17

This Doctorate Dissertation is accepted by:

Acting-Dean of the College of Medicine and Health Sciences: Professor Ruth Langer

Signature [Signature] Date 10/3/17

Dean of the College of the Graduate Studies: Professor Nagi T. Wakim

Signature [Signature] Date 14/5/2017

Copy 7 of 15
Abstract

Brief introduction: There are limited studies on CYP1A2 and NAT2 polymorphisms among Emiratis. Aims: This study aims to determine CYP1A2 and NAT2 alleles and genotypes and correlate these genotypes with caffeine metabolism phenotypes among Emiratis. Methods: After obtaining informed consent, five hundred and eighty one non-smoker subjects were given 300ml of caffeinated soft drink and were asked to provide a buccal swab and a urine sample two hours later. TaqMan Real Time PCR, PCR-RFLP, DNA sequencing were performed to determine CYP1A2 and NAT2 alleles and genotypes. Phenotyping was carried out by analysing the caffeine metabolites using HPLC analysis. Results: We found that 1.4%, 16.3% and 82.3% of the recruited subjects were slow, intermediate and rapid CYP1A2 metabolisers, respectively. Only 1.4% of the subjects were homozygotes for CYP1A2 mutant alleles while 16.1% were heterozygotes and 82.5% were homozygotes for the CYP1A2 wild type genotype. Therefore, the frequency of the wild type genotype CYP1A2*1A/*1A was 0.825 followed by CYP1A2*1A/*1C and CYP1A2*1A/*1K, with frequencies of 0.102 and 0.058, respectively. The degree of phenotype/genotype concordance was 81.6%. The CYP1A2*1C/*1C and CYP1A2*3/*3 genotypes showed the lowest phenotype status. With regards to NAT2, we found that 78.5%, 19.1% and 2.4% of the subjects were slow, intermediate and rapid NAT2 acetylators, respectively, with 77.4% being homozygote or heterozygote for NAT2 mutant alleles and 18.4% and 4.2% were heterozygote and homozygote for the NAT2 wild type genotypes, respectively. The most common genotypes found were NAT2*5B/*7B, NAT2*5B/*6A, NAT2*7B/*14B and NAT2*4/*5B with frequencies of 0.255, 0.135, 0.105 and 0.09, respectively. The degree of phenotype/genotype concordance was equal to 96.2%. The NAT2*6A/*6A, NAT2*6A/*7B, NAT2*7B/*7B, NAT2*5A/*5B and NAT2*5A/*5A...
genotypes were found to be associated with the lowest 5-Acetylamino-6-Formylamino-3-Methyluracil/1-Methylxanthine (AFMU/1X) ratios. Significant contributions: The majority of the studied Emiratis are slow NAT2 acetylators with implications for the prescription of medications that are metabolised by this enzyme. In addition, a small percentage of Emiratis have slow CYP1A2 enzyme activity which again should be taken into consideration when prescribing medications that are partially metabolised by this enzyme. In Emirati population, the frequency of the CYP1A2*1A and NAT2*5 alleles were the highest relative to other alleles frequencies. Moreover, Individuals who carried NAT2*6A/*6A, NAT2*6A/*7B, NAT2*7B/*7B, NAT2*5A/*5B, NAT2*5A/*5A, CYP1A2*1C/*1C or CYP1A2*3/*3 genotypes might be at high risk of toxicity with some drugs and some diseases compared to others as these genotypes are associated with the slowest phenotype status. Consequently, genetic testing is recommended prior to prescribing medications that are largely metabolized by CYP1A2 or NAT2. Gap filled: This is the first detailed study of CYP1A2 and NAT2 alleles and genotypes among Emiratis.

Keywords: Cytochrome P450-1A2 (CYP1A2) polymorphisms, N-acetyltransferase-2 (NAT2) polymorphisms, Emiratis genetics, drug metabolism.
دراسة التنوع في أشكال التنميط الظاهري والجيني السايتوكروم و والاستابيلي بين مواطني دولة الإمارات العربية المتحدة

الملخص

المقدمة: هناك القليل من الدراسات حول تنوع أشكال التنميط الظاهري والجيني السايتوكروم ل- NAT2 والاستابيلي ل- CYP1A2 بين مواطني دولة الإمارات العربية المتحدة. أهداف البحث: تمت هذه الدراسة لمعرفة أشكال التنميط الظاهري والجيني السايتوكروم ل- CYP1A2 والاستابيلي ل- NAT2 في مواطني الدولة. طريقة البحث: لقد تم اختيار 581 شخص غير مدخن لتناول 300 مل من المشروبات التي تحتوي على الكافيين ومن ثم تم أخذ مسحة من الشدق وعينة من البول. بعد ذلك تم عملية HPLC وتحليل ال- DNA وPCR-RFLP و Real time PCR لمعرفة أشكال التنميط CYP1A2 والسايتوكروم و نتائج البحث. لقد وجدنا أن 1.4% من العينات كانت بطيئة التنميط الظاهري السايتوكروم ل- CYP1A2 و 16.3% منها كانت متوسطة CYP1A2. بينما كانت 82.3% منها كانت سريعة التنميط الظاهري السايتوكروم ل- CYP1A2. المتماثلة للائيات الفاقدة ل- 16.1% متباناة لاليال الغير Methylated CYP1A2 و 82.5% متماثلة لاليال الغير متحول البايتوكروم ل- CYP1A2. ونجد أيضاً أن الالي الغير متحول هو الالي الأكثر انتشارا بنسبة 0.825 ، يتبعه الالي متحول، حيث يوجد ارتباط ذو أهمية بين التنميط الظاهري السايتوكروم ل- CYP1A2 والتمييز الجيني له بقيمة 81.6% ، إلى جانب ذلك، نجد أيضاً أن الأشخاص الذين لديهم الائيات المتماثلة و 3.3% من الاليات البارزة لديهم استابيلات NAT2 و 8.5% من الأشخاص لديهم أسفلات NAT2. بطيئة و 19.1% منهم متوسطة و 2.4% منهم سريعة. إلى جانب ذلك، 77.5% من العينات متماثلة أو متباناة لاليات البارزة، أما لاليات الاستابيلات و 18.4% منها متباناة لاليات الاستابيلات الغير لاليات البارزة. متقابلة للائيات الاستابيلات لاليات البارزة و 4.2% منها متباناة لاليات الاستابيلات الغير لاليات البارزة. ل- NAT2.
ووجد أيضًا أن الأليلات المتحولة مثل NAT2 الأكثر انتشارًا بالنسب الأدنى 0.225 للأليل 5B/6A و 0.135 للاليل 7B/1B بالكامل و 0.105 للاليل 5B/6A. وعلاوة على ذلك، كان هناك ارتباط ذو أهمية بين التنميط الظاهري والتنميط الجيني له قيمة 96.2%. ولقد وجدنا أن الأنسام الجينية الأدنى NAT2*6A/6A, NAT2*6A/7B, NAT2*7B/7B, NAT2*5A/5B and (M) مرتبطة مع أقل نتائج أو نسبة التنميط الظاهري عن طريق حساب نسبة CYP1A2 والنية CYP1A2 (AFMU/1X). مساهمات ذات أهمية: هناك نسبة قليلة من نشاطية إنزيم الساينتاكرومي A2 بين مواطني الدولة والتي ارتبطت بشكل مباشر بوجود نسبة قليلة من الأليلات المتحولة الساينتاكرومي L-2 ونسب عالية من الأليلات المتحولة الاستايلية L-3. كانت نسبة التنميط الجيني الساينتاكرومي A2 في الأعلى في الدولة. علاوة على ذلك، يعتبر الأشخاص الذين لديهم الأنسام الجينية الأدنى (NAT2*6A/6A, NAT2*6A/7B, NAT2*7B/7B, NAT2*5A/5B, NAT2*5A/5A, CYP1A2*1C/1C) هم إلى حد كبير أكثر المجموعات عرضة لخطر السمية الناتجة عن الأدوية وخطر الإصابة بالأمراض. وعليه فإنه ينصح بالإجراء الاختبارات الجينية اللازمة قبل وصف الأدوية التي تخضع للعملية الاستقلالية بواسطة NAT2 أو CYP1A2. وهنا يكون هذا البحث الأول من نوعه لدراسة التنوع في أنماط التنميط الظاهري والجيني الساينتاكرومي والاستايلسي بين مواطني دولة الإمارات العربية المتحدة.

مफاهيم البحث الرئيسية: التنميط الجيني الساينتاكرومي (CYP1A2), التنميط الجيني الاستايلسي (NAT2).
Acknowledgments

I express my deep sense of gratitude to Professor Salim Bastaki and Professor Bassam Ali. Their dedication and keen interest above all their overwhelming attitude to help their students had been solely and mainly responsible for completing my work.

I thank profusely Professor Salim Bastaki, Professor Bassam Ali and Professor Yousef Abdulrazzaq for their kind help, cooperation, invaluable assistance, support and guidance throughout my study period.

I owe a deep sense of thanks to the chairman, all members of the Department of Pharmacology, Pharmacology and Pathology lab staff at the United Arab Emirates University for assisting me in my studies and research. My special thanks are extended to Ms Naheed Amir who helped in ordering and maintaining the required materials for my studies. Nevertheless, my thanks extended to others who provided help and supported me during my study particularly Mr Dhanasekhar, Mr. Azim Ullah Shamsulislam and Ms Anne John.

I would like to convey thanks to the College of Medicine and Health Sciences for giving me this opportunity to get my PhD degree.

Finally, I acknowledge with thanks to my family and friends for their constant encouragement during my research period.
Dedication

To my family who helped and supported me from day one, to my colleagues, friends and supervisors; Professor Salim Bastaki, Professor Bassam Ali and Professor Yousef Abdulrazzaq, who guided and helped me undertake this project.

I thank them all for the support and help they offered me throughout my journey in accomplishing this thesis/dissertation and getting my PhD degree.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>i</td>
</tr>
<tr>
<td>Declaration of Original Work</td>
<td>ii</td>
</tr>
<tr>
<td>Copyright</td>
<td>iii</td>
</tr>
<tr>
<td>Advisory Committee</td>
<td>iv</td>
</tr>
<tr>
<td>Approval of the Doctorate Dissertation</td>
<td>v</td>
</tr>
<tr>
<td>Abstract</td>
<td>vii</td>
</tr>
<tr>
<td>Title and Abstract (in Arabic)</td>
<td>ix</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>xi</td>
</tr>
<tr>
<td>Dedication</td>
<td>xii</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>xiii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xvii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xix</td>
</tr>
<tr>
<td>List of Symbols, Acronyms, Abbreviations</td>
<td>xxii</td>
</tr>
<tr>
<td>List of Glossary</td>
<td>xxv</td>
</tr>
<tr>
<td>Chapter 1: Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Overview</td>
<td>1</td>
</tr>
<tr>
<td>1.1.1 Background</td>
<td>1</td>
</tr>
<tr>
<td>1.1.2 Factors Influence Drug Metabolism</td>
<td>2</td>
</tr>
<tr>
<td>1.1.2.1 Genetic Polymorphism</td>
<td>2</td>
</tr>
<tr>
<td>1.1.2.2 Diseases</td>
<td>4</td>
</tr>
<tr>
<td>1.1.2.3 Age</td>
<td>4</td>
</tr>
<tr>
<td>1.1.2.4 Concomitant Drugs</td>
<td>4</td>
</tr>
<tr>
<td>1.1.3 Cytochrome P450 (CYP450)</td>
<td>6</td>
</tr>
<tr>
<td>1.1.3.1 Cytochrome P450 Family</td>
<td>7</td>
</tr>
</tbody>
</table>
1.1.4 CYP1A2 Isoform ................................................. 9
1.1.5 Acetylation ..................................................... 10
1.2 Statement of the Problem ....................................... 12
1.3 Literature Review ................................................ 15
1.3.1 CYP1A2 .......................................................... 15
  1.3.1.1 Gene Localization, Protein Structure and Expression ..... 15
    1.3.1.1.1 The CYP1A2 Gene Family .......................... 15
    1.3.1.1.2 Reaction Mechanism, Structural Features and Functional
               Relevance of CYP1A2 .................................. 16
    1.3.1.1.3 CYP1A2 Regulation .................................. 18
    1.3.1.1.4 CYP1A2 Variants and its Molecular Epidemiology ...... 18
    1.3.1.1.5 CYP1A2 Important Haplotypes ....................... 24
    1.3.1.1.6 CYP1A2 Alleles Nomenclature/Variant Annotations ...... 27
    1.3.1.1.7 CYP1A2 Genotype/Phenotype Relationship ............. 28
    1.3.1.1.8 Clinical Impact, Drug Metabolism and Substrate
               Specificity of CYP1A2 Polymorphism .................... 30
1.3.2 N-Acetyltransferase-2 (NAT2) ................................ 36
  1.3.2.1 Gene Localization, Structure and Expression .............. 36
    1.3.2.1.1 N-acetyltransferase Gene Family ................... 36
    1.3.2.1.2 Reaction Mechanism, Structural Features, Variations and
               Functional Relevance of NAT2 ........................ 37
    1.3.2.1.3 NAT2 Regulation .................................... 39
    1.3.2.1.4 NAT2 Variants and its Molecular Epidemiology .......... 40
    1.3.2.1.5 NAT2 Important Haplotypes ........................ 45
    1.3.2.1.6 NAT2 Alleles Nomenclature/Variant Annotations .......... 45
    1.3.2.1.7 NAT2 Genotype/Phenotype Relationship ............... 46
3.2.4.3 Polymerase Chain Reaction – Restriction Fragment Length Polymorphism (PCR – RFLP) ............................................. 78

3.2.4.3.1 Polymerase Chain Reaction (PCR) ............................................. 78

3.2.4.3.2 Restriction Fragment Length Polymorphism (RFLP) .......... 81

3.3 Data Analysis ................................................................................. 87

Chapter 4: Results ........................................................................... 88

4.1 Population Data ........................................................................... 88

4.2 Descriptive Analysis of the Recruited Sample .............................. 89

4.3 HPLC Analysis ............................................................................ 90

4.4 Statistical Analysis ....................................................................... 92

4.4.1 CYP1A2 .................................................................................. 92

4.4.1.1 CYP1A2 Genotype ................................................................. 92

4.4.1.2 Caffeine Metabolites Ratio and CYP1A2 Phenotype Status .... 97

4.4.2 NAT2 .................................................................................... 103

4.4.2.1 NAT2 Genotype ................................................................. 103

4.4.2.2 Caffeine Metabolites Ratio and NAT2 Phenotype Status ....... 113

Chapter 5: Discussions ..................................................................... 121

5.1 CYP1A2 .................................................................................... 122

5.2 NAT2 ....................................................................................... 124

Chapter 6: Conclusions ................................................................... 131

6.1 CYP1A2 .................................................................................... 131

6.2 NAT2 ....................................................................................... 131

References ....................................................................................... 132

List of Publications ........................................................................... 156

Appendices ....................................................................................... 157
List of Tables

Table 1: The summary of the cytochrome P450 protein families ............................................................... 8

Table 2: Frequencies of important SNPs in Human CYP1A2 Gene among various populations ................................................................. 21

Table 3: CYP1A2 alleles, their references SNPs and catalytic activity based on the Human Cytochrome P450 Allele Nomenclature Committee .......................................................................................... 27

Table 4: Frequency of important SNPs in human NAT2 gene ................................................................. 42

Table 5: Frequency of other SNPs in human NAT2 gene ........................................................................... 43

Table 6: NAT2 alleles and its references SNPs and catalytic activity ................................................................. 45

Table 7: Effect of acetylator status on drug response and toxicity ......................................................... 57

Table 8: The expected distribution of our samples based on the distribution of UAE (Emiratis) between the Emirates ................................................................. 62

Table 9: Classification of CYP1A2 enzyme activity based on the ratio of (17MX+17MU)/137MX ................................................................. 66

Table 10: The acetylator status of NAT2 activity was classified based on the ratio of AFMU/1MX ................................................................. 66

Table 11: Fluorescence signal correlations ........................................................................... 69

Table 12: Assay ID or design that has been used for each SNP ................................................................. 72

Table 13: Preparation of TBE solution ........................................................................... 81

Table 14: The restriction cleavage for the amplificates of the PCR product ................................................................. 82

Table 15: The restriction cleavage site and its DNA fragment lengths from the restriction cleavage ................................................................. 84

Table 16: Evaluation sheet (✓ or • = fragment length combination found) of the restriction enzymes with mutation sites and fragment length for NAT2 ................................................................. 86
Table 17: The CYP1A2 genotype frequencies and distribution

Table 18: The characteristic features and frequencies of the various CYP1A2 haplotypes identified among Emiratis

Table 19: The classification and the results of the CYP1A2 enzyme activity based on the ratio of (17MX+17MU)/137MX

Table 20: (17MX + 17MU)/137MX molar ratios in subgroups assigned to CYP1A2 allele combinations

Table 21: (17MX + 17MU)/137MX ratios vs. genotypes. Suggested genotypic state: S/S homozygously slow, R/S heterozygotes intermediate, R/R homozygously rapid

Table 22: NAT2 diplotypes and genotypes in the study group. Comparison between genotypes reconstructed by PHASE and actual genotypes obtained by gene mapping

Table 23: The haplotype frequencies of NAT2

Table 24: Details of the results of the 6 samples that has been analysed using DNA sequencing

Table 25: The classification and the results of the acetylator status of NAT2 activity and its sub-classification based on (AFMU/1MX) ratio

Table 26: AFMU/1X molar ratios in subgroups assigned to NAT2 allele combinations

Table 27: AFMU/1X ratios vs. genotypes, based on the last update of NAT nomenclature scheme published by The Arylamine N-acetyltransferase Gene Nomenclature Committee according to consensus guidelines (Hein DW et al 2008). Genotypic acetylator state: S/S homozygously slow, R/S heterozygotes, R/R homozygously rapid
List of Figures

Figure 1: The process of drug metabolism through phase I, II and III................................. 2
Figure 2: Spectrum of P450........................................................................................................ 6
Figure 3: Structure of cytochrome P450.................................................................................... 7
Figure 4: Illustrates the process of acetylation of proteins....................................................... 12
Figure 5: CYP1A2 gene structure............................................................................................... 17
Figure 6: Non-synonymous SNPs of the CYP1A2 gene............................................................ 19
Figure 7: Schematic representation of caffeine metabolites. CYP1A2 catalyzes the 3-demethylation of caffeine to paraxanthine and NAT2 catalyzes the acetylation of paraxanthin metabolites to AMFU................................................................. 29
Figure 8: Chromosomal location and structure of N-acetyltransferase (NAT2) gene............................. 37
Figure 9: Structure of human NAT2 gene.................................................................................. 38
Figure 10: The metabolic pathways responsible for the activation and detoxification of the drug............................................................... 53
Figure 11: Instruction of Isohelix SK1 buccal swabs use............................................................ 63
Figure 12: Complementary TaqMan probe fluorescences after it anneals to the template and after cleavage by AmpliTaq Gold DNA Polymerase UP........................................................................................................ 70
Figure 13: Overview of a genotyping experiment using Real Time PCR......................................... 73
Figure 14: The thermal cycling condition specified for our TaqMan genotyping assays for CYP1A2........................................................................................................ 75
Figure 15: The Standard mode thermal cycling setting............................................................... 75
Figure 16: Variation in clustering due to the genotype of the target allele........................................ 76
Figure 17: The allelic discrimination plate read document........................................................ 77
Figure 18: RT-PCR plate layout during sample running in the system.  

Figure 19: The thermocycler program for the NAT2-PCR.  

Figure 20: NAT2 gene structure with the mutation sites.  

Figure 21: The age distribution of recruited subjects.  

Figure 22: Representative chromatogram curves that are specific for caffeine metabolites with their retention time.  

Figure 23: Standard Spectral conformation of 137MX, 17MX, 17MU, 1MU, 1MX and AFMU. The amount on X axis is in ug/ml.  

Figure 24: The genotype frequencies of the studied sample - 0.014 of our subjects were homozygote for mutant alleles, 0.161 were heterozygote and 0.825 were homozygous for the wild type genotyping.  

Figure 25: Allele frequencies- allele CYP1A2*1C frequency = 0.0636, allele CYP1A2*1k = 0.0287, allele CYP1A2*3 = 0.0027, allele CYP1A2*4 = 0 and allele CYP1A2*6 = 0. The wild type allele frequency was found to be 0.905.  

Figure 26: An example of allelic discrimination plot that demonstrates a variation in clustering due to the genotyping of the targeted allele (CYP1A2*1C).  

Figure 27: The frequency of CYP1A2 alleles among Emiratis in comparison with other populations.  

Figure 28: The distribution of subjects based on CYP1A2 activity (17MX+17MU)/137MX ratio.  

Figure 29: The frequency of CYP1A2 poor metabolizers among Emiratis in comparison with other populations.  

Figure 30: The genotype frequencies- 0.774 of our subjects were homozygote or heterozygote for 2 mutant alleles genotyping, 0.184 were heterozygote and 0.042 were homozygotes for the wild type genotyping.  

Figure 31: Allele frequencies- allele NAT2*5 frequency = 0.364, allele NAT2*6 = 0.129, allele NAT2*7 = 0.274 and allele NAT2*14 = 0.099. The wild type allele frequency was found to be 0.134.
Figure 32: Representative images of the gel documentation of the DNA fragments after amplification and restriction cleavage. bp = base pairs, w = wild type = reference sequence (RS) and m = mutant = sequence variation (SV).

Figure 33: The frequency of NAT2 slow acetylation alleles among Emiratis in comparison with other populations.

Figure 34: An example of DNA chromatogram of T-510C variant of sample 3 DNA.

Figure 35: The distribution of subjects based on NAT2 activity (AMFU/1MX) ratio.

Figure 36: Comparison of NAT2 phenotype and genotype in 555 subjects.

Figure 37: The frequency of NAT2 poor metabolizers among Emiratis in comparison with other populations.
<table>
<thead>
<tr>
<th>Symbol/Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>137MX or 137X:</td>
<td>Caffeine (1,3,7-trimethylxanthine)</td>
</tr>
<tr>
<td>17MU or 17U:</td>
<td>1,7-dimethyluric acid</td>
</tr>
<tr>
<td>17MX or 17X:</td>
<td>Paraxanthine (1,7-dimethylxanthine)</td>
</tr>
<tr>
<td>1MU or 1U:</td>
<td>1-Methyluric acid</td>
</tr>
<tr>
<td>1MX or 1X:</td>
<td>1-Methylxanthine</td>
</tr>
<tr>
<td>AAMU:</td>
<td>5-acetyl amino-6-amino-3-methyluracil</td>
</tr>
<tr>
<td>ADEC:</td>
<td>Abu Dhabi Education Council</td>
</tr>
<tr>
<td>ADRs:</td>
<td>Adverse Drug reactions</td>
</tr>
<tr>
<td>AFB1:</td>
<td>Aflatoxin B1</td>
</tr>
<tr>
<td>AFMU:</td>
<td>5-Acetyl amino-6-Formylamino-3-Methyluracil</td>
</tr>
<tr>
<td>AhR:</td>
<td>Aromatic Hydrocarbon Receptor</td>
</tr>
<tr>
<td>Anti-TB drugs:</td>
<td>Anti-Tuberculosis drugs</td>
</tr>
<tr>
<td>Arnt:</td>
<td>AhR nuclear translocator</td>
</tr>
<tr>
<td>BP:</td>
<td>Blood Pressure</td>
</tr>
<tr>
<td>CBT:</td>
<td>Caffeine Breath Test</td>
</tr>
<tr>
<td>CI:</td>
<td>Confidence Interval</td>
</tr>
<tr>
<td>CoA:</td>
<td>Coenzyme A</td>
</tr>
<tr>
<td>CT solution:</td>
<td>Capture buffer</td>
</tr>
<tr>
<td>CYP1A2:</td>
<td>Cytochrome P450 –family “1” subfamily “A” gene “2”</td>
</tr>
<tr>
<td>DILI:</td>
<td>Drug Induce Liver Injury</td>
</tr>
<tr>
<td>DNA:</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTPs:</td>
<td>Deoxyribonucleotide triphosphates</td>
</tr>
<tr>
<td>GSTT1:</td>
<td>Glutathione S-Transferase Theta 1</td>
</tr>
<tr>
<td>HPLC:</td>
<td>High Performance Liquid Chromatography</td>
</tr>
<tr>
<td>INH:</td>
<td>Isoniazid</td>
</tr>
</tbody>
</table>
LS solution: Lysis buffer
MAOB: Monoamine Oxidase B
NAPQI: N-acetyl-p-benzoquinone imine
NAT1: N-Acetyltransferase-1
NAT2: N-Acetyltransferase-2
NATP1: N-Acetyltransferase pseudogene
NFQ: Nonfluorescent quencher
ORF: Open Reading Frame
PABA: Para-Aminobenzoic Acid
PAHs: Polycyclic Aromatic Hydrocarbons
PAS: Para-Aminosalicylic Acid
PAS: Per-Arnt-Sim
PCR: Polymerase Chain Reaction
PCR-RFLP: Polymerase Chain Reaction - Restriction Fragment Length Polymorphism
PCT: Porphyria Cutanea Tarda
PD: Parkinson Disease
Per: Period (Drosophila period clock protein)
PGx: Pharmacogenetics
PK solution: Proteinase K
PK: Pharmacokinetics
SLE: Systemic Lupus Erythematosus
SMZ: Sulfamethazine
SNP: Single Nucleotide Polymorphism
SPSS: Statistical Package for the Social Sciences
RA: Rheumatoid Arthritis
TaqMan MGB: TaqMan Minor Groove Binder
TE solution: Tris and EDTA solution (Rehydration buffer)
Tm: Melting temperature
UAE: United Arab Emirates
UAEU: United Arab Emirates University
UP: Ultra-Pure
XO: Xanthine oxidase
List of Glossary

Alleles
One of a pair of genes that appear at a particular location on a particular chromosome and control the same characteristic, such as blood type or colour-blindness. Alleles are also called alleleomorphs.

BamHI
A restriction endonuclease enzyme that has specific restrictive cleavage sites and detect mutation at (857G>A) site used to identified \( NAT2^*7 \) (rs 1799931).

Ddel
A restriction endonuclease enzyme that has specific restrictive cleavage sites and detect mutations at (341T>C) and (803A>G) sites used to identified \( NAT2^*5 \) (rs 1801280) and \( NAT2^*12 \) (rs 1208) respectively.

FokI
A restriction endonuclease enzyme that has specific restrictive cleavage sites and detect mutation at (282C>T) site used to identified \( NAT2^*13 \) (rs 1041983).

IUPAC nomenclature
The IUPAC “International Union of Pure and Applied Chemistry” nomenclature system is a set of logical rules devised and used by organic chemists to circumvent problems caused by arbitrary nomenclature. Knowing these rules and given a structural formula, one should be able to write a unique name for every distinct compound.

KpnI
A restriction endonuclease enzyme that has specific restrictive cleavage sites and detect mutation at (481C>T) site used to identified \( NAT2^*11 \) (rs 1799929).

MspI
A restriction endonuclease enzyme that has specific restrictive cleavage sites and detect mutation at (191G>A) site used to identified \( NAT2^*14 \) (rs 1801279).

NADPH dependent
Category groups of enzymes that use nicotinamide adenine dinucleotide phosphate (NADP+ and its reduced form, NADPH) in redox reactions.

Polymorphism
Natural variations in a gene, DNA sequence, or chromosome that have no adverse effects on the individual and occur with fairly high frequency in the general population.
Rn  It is the fluorescence of the reporter dye divided by the fluorescence of a passive reference dye. It is used for fluorescence measurement.

SDS software plots  The SDS software plots showed the results of the allelic discrimination run on a scatter plot of Allele 1 versus Allele 2. Each well of the 96-well plate is represented as an individual point on the plot.

Taq-I  A restriction endonuclease enzyme that has specific restrictive cleavage sites and detect mutation at (590G>A) site used to identified NAT2*6 (rs 1799930).
Chapter 1: Introduction

1.1 Overview

1.1.1 Background

Drug metabolism is part of the xenobiotic metabolism processes that involve converting lipophilic drugs to more easily excretable hydrophilic products with each drug requiring a specific set of enzymes. The reactions included in these processes are clinically crucial in terms of optimal therapeutic dose, drug toxicity, cancer chemotherapy and some pathogens drug resistance (Mizuno N et al 2003).

Drug metabolism consists of three phases; phase I involves certain enzymes which include mainly the cytochrome P450 (CYP450) oxidases that act by adding reactive or polar groups to the xenobiotic. Usually the modified compound will be less active than the parent compound, although with some drugs this conversion forms the more pharmacologically active agents. In some cases the modified compound can be toxic, teratogenic or even carcinogenic (Meyer JM and Rodvold KA 1996). These modified compounds will further undergo conjugation process by transferase enzymes such as N-acetyltransferase to form polar compounds (phase II reaction). Finally, the conjugated xenobiotics will be pumped out of the cell by the phase III efflux xeno-transporters (Figure 1) (Homolya L et al 2003).
Figure 1: The processes of drug metabolism through phase I, II and III

Oxidation, reduction, hydrolysis, cyclization, decyclization, and addition of oxygen or removal of hydrogen, are carried out by mixed function oxidases "CYP450 like CYP1A2" in phase I reaction (Akagah B et al 2008). Acetylation, glucuronidation, glutathione conjugation, glycine conjugation, methylation and sulphation are example of the pathways in phase II reaction (Liston HL et al 2001).

1.1.2 Factors Influencing Drug Metabolism

1.1.2.1 Genetic Polymorphism

Genetic variation is an essential factor that contributes to interindividual variability in drug biotransformation. Genetics variations, termed genetic polymorphism, arise due to the presence of various combinations of alleles at the same chromosomal locus and in certain cases they are linked to inherited traits. In this context, the ability to metabolize a drug is a trait that is largely determined genetically. The definition of a polymorphism in pharmacogenetics is the availability
of at least genetic 2 groups within a population with distinctly different abilities to metabolize a drug (Poulsen HE and Loft S 1992, Hedrick P 2011).

Genetic differences in drug metabolism are the result of genetically based variation in genes that code for enzymes and proteins responsible for the metabolism or transport of the drug. In polymorphisms, the genes show variation (contain abnormal pairs or multiples or abnormal alleles) leading to altered enzyme activity. Differences in enzyme activity occur at different rates according to racial group. Genetic changes may inactivate or reduce enzyme activity leading to an increase in the level of the substrate drug. Genetic duplication may increase enzyme activity resulting in lower levels of the substrate drug concentration. (Poulsen HE and Loft S 1992).

Genetic polymorphisms of drug-metabolizing enzymes give rise to distinct subgroups in the population that differs in their ability to perform certain drug biotransformation reactions. Polymorphisms are generated by mutations in the genes for these enzymes, which cause decreased, increased, or absent enzyme expression or activity by multiple molecular mechanisms. Individuals can be classified into three main phenotypic groups: extensive, intermediate and poor metabolizers. Poor metabolizers often have higher risk of adverse effects including toxicity (Meyer JM and Rodvold KA 1996).

Generally, poor metabolizers have two defective alleles (e.g.: CYP2D6*4/*5 and CYP2D6*4/*4) or combination of alleles including one resulting in no enzyme (e.g.: CYP2D6*5 and CYP2D6*4 deletion).

Intermediate metabolizers on the other hand, are heterozygous “having only one wild type allele and one defective allele”. Normal metabolizers carry wild type
alleles (e.g.: CYP2D6*1/*1). Wild type alleles encode genes for normal enzyme function.

The combination of alleles encoding the gene determines the activity and effectiveness of the enzyme. The overall function of the enzyme is the phenotype of enzyme function. Phenotype is defined as the observable physical or biochemical characteristics determined by both genetic makeup and environmental influences (Meyer JM and Rodvold KA 1996).

1.1.2.2 Diseases

Infection, hepatitis, alcoholic liver disease, biliary cirrhosis, hepatocarcinoma and other forms of impaired liver functions can lead to decreased drug biotransformation and the degree of impairment is a function of the severity of the disease (Meyer J.M. and Rodvold K.A. 1996).

1.1.2.3 Age

The ability to metabolize a drug is also influenced by the age of the individual. For examples, very young infants do not have a mature enzyme system. In addition, decreases in liver mass, hepatic enzyme activity and hepatic blood flow are more concomitant with the elderly (Meyer JM and Rodvold KA 1996).

1.1.2.4 Concomitant Drugs

The most common causes that affect drug biotransformation reactions are induction and inhibition of cytochrome P450 enzymes and thus agents that promote these processes have impact on drug metabolism as briefed in the following sections (Meyer JM and Rodvold KA 1996).
A- Induction

Induction is an increase in enzyme activity that is associated with exposure to drugs. This can occur when a drug alters the biotransformation of co-administered drugs either by the same enzyme pathway or through an alternative pathway. However, inducers are usually specific for a given cytochrome P450 family. Sometimes a drug can induce its own biotransformation in addition to that of other agents (Meyer JM and Rodvold KA 1996).

The main factors that affect the time course of enzyme induction onset and offset are plasma concentration of the inducer and the half-life of enzyme production and degradation. The effect of induction can be seen within the first few days of therapy (Dossing M. et al 1983).

B- Inhibition

One of the most common mechanisms of inhibition is competitive inhibition and occurs when two or more drugs compete for the same enzyme, which becomes clinical significant based on the relative concentrations of the drugs, as well as, a variety of other patient specific factors (Meyer JM and Rodvold KA 1996).

The binding can be either irreversible or reversible with the heme-binding site of the enzyme and inhibits other drugs from binding, which leads to loss of function that activity can only be restored by synthesis of new enzymes, which may take several days (Meyer JM and Rodvold KA 1996).

More complex mechanisms of inhibition can occur where some drugs undergo metabolic activation by the cytochrome P450 system to inhibitory products that could result in relatively stable complexes with cytochrome P450. In this scenario the cytochrome is held in an inactive state that can be clinically significant.
Additionally, the interaction involves narrow therapeutic drugs which could increase the risk of toxicity (Meyer JM and Rodvold KA 1996).

Unlike induction enzyme, inhibition usually begins with the first dose of the inhibitor. The inhibition reaches the maximum level when the inhibitor reaches steady state (four to five half-lives). Similarly, the time needed for the interaction to resolve depends on the half-lives of the drugs involved (Dossing M et al 1983).

1.1.3 Cytochrome P450 (CYP450)

Cytochrome P450 enzymes are composed of a large superfamily of heme-thiolate proteins monooxygenases responsible for the metabolism of a wide variety of both exogenous and endogenous compounds (Degtyarenko KN 1995). In 1955, it was discovered in rat liver microsomes and they were characterized by an intense absorption band at 450 nm in the presence of carbon monoxide (Vercruysse A 1997) (Figure 2).

![Figure 2: Spectrum of P450](image)

The cytochrome P450 (CYP) mixed function monooxygenases are located on the smooth endoplasmic reticulum of cells throughout the body and are present
at highest concentrations in the liver (hepatocytes) and small intestine. These enzymes are largely responsible for phase I xenobiotic metabolism (Figure 3) (Buck ML 1997).

**Cytochrome P450_{sp}**

![PDB code: 3AWM](image)

Figure 3: Structure of cytochrome P450

### 1.1.3.1 Cytochrome P450 Family

Almost 12 cytochrome P450 gene families have been reported in the human genome. However, only 3 families, the cytochrome P450 1, 2 and 3 (CYP1, CYP2 and CYP3) are involved in the majority of drugs biotransformations (Meyer JM and Rodvold KA 1996). Based on amino acid sequence similarities the enzymes are divided into families, and each family can be further segregated into subfamilies, which are identified by capital letters following the family designation (for example CYP1 subfamily A is designated CYP1A). The most important and clinically relevant cytochromes are CYP3A4, CYP2D6, CYP1A2, CYP2C9, CYP2C19 and CYP2E1 (Martin J and Fay M 2001). In total, humans have 57 genes and more than 59 pseudogenes divided into 18 families of cytochrome P450 genes and 43 subfamilies as shown in table 1 (Nelson DR et al 2004).
Table 1: The summary of the cytochrome P450 protein families

<table>
<thead>
<tr>
<th>Family</th>
<th>Members</th>
<th>Gene Names</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP1</td>
<td>3 subfamilies, 3 genes, 1 pseudogene</td>
<td>CYP1A1, CYP1A2, CYP1B1</td>
</tr>
<tr>
<td>CYP2</td>
<td>13 subfamilies, 16 genes, 16 pseudogenes</td>
<td>CYP2A6, CYP2A7, CYP2A13, CYP2B6, CYP2C8, CYP2C9, CYP2C18, CYP2C19, CYP2D6, CYP2E1, CYP2F1, CYP2J2, CYP2R1, CYP2S1, CYP2U1, CYP2W1</td>
</tr>
<tr>
<td>CYP3</td>
<td>1 subfamily, 4 genes, 2 pseudogenes</td>
<td>CYP3A4, CYP3A5, CYP3A7, CYP3A43</td>
</tr>
<tr>
<td>CYP4</td>
<td>6 subfamilies, 12 genes, 10 pseudogenes</td>
<td>CYP4A11, CYP4A22, CYP4B1, CYP4F2, CYP4F3, CYP4F8, CYP4F11, CYP4F12, CYP4F22, CYP4V2, CYP4X1, CYP4Z1</td>
</tr>
<tr>
<td>CYP5</td>
<td>1 subfamily, 1 gene</td>
<td>CYP5A1</td>
</tr>
<tr>
<td>CYP7</td>
<td>2 subfamilies, 2 genes</td>
<td>CYP7A1, CYP7B1</td>
</tr>
<tr>
<td>CYP8</td>
<td>2 subfamilies, 2 genes</td>
<td>CYP8A1, CYP8B1</td>
</tr>
<tr>
<td>CYP11</td>
<td>2 subfamilies, 3 genes</td>
<td>CYP11A1, CYP11B1, CYP11B2</td>
</tr>
<tr>
<td>CYP17</td>
<td>1 subfamily, 1 gene</td>
<td>CYP17A1</td>
</tr>
<tr>
<td>CYP19</td>
<td>1 subfamily, 1 gene</td>
<td>CYP19A1</td>
</tr>
<tr>
<td>CYP20</td>
<td>1 subfamily, 1 gene</td>
<td>CYP20A1</td>
</tr>
<tr>
<td>CYP21</td>
<td>2 subfamilies, 1 gene, 1 pseudogene</td>
<td>CYP21A2</td>
</tr>
<tr>
<td>CYP24</td>
<td>1 subfamily, 1 gene</td>
<td>CYP24A1</td>
</tr>
<tr>
<td>CYP26</td>
<td>3 subfamilies, 3 genes</td>
<td>CYP26A1, CYP26B1, CYP26C1</td>
</tr>
<tr>
<td>CYP27</td>
<td>3 subfamilies, 3 genes</td>
<td>CYP27A1, CYP27B1, CYP27C1</td>
</tr>
<tr>
<td>CYP39</td>
<td>1 subfamily, 1 gene</td>
<td>CYP39A1</td>
</tr>
<tr>
<td>CYP46</td>
<td>1 subfamily, 1 gene</td>
<td>CYP46A1</td>
</tr>
<tr>
<td>CYP51</td>
<td>1 subfamily, 1 gene, 3 pseudogenes</td>
<td>CYP51A1</td>
</tr>
</tbody>
</table>
1.1.4 CYP1A2 Isoform

Each of the isoforms has wide substrate specificity, but each has its own specific substrate characteristics. These isoforms have differing regulatory mechanisms to control their activity. The regulatory mechanisms involve chemicals which induce or inhibit the enzyme. For example, CYP1A2 metabolises some carcinogenic tars in cigarette smoke and is induced by these chemicals. Members of other CYP gene families are induced by drugs such as barbiturates, anticonvulsants and rifampicin (Martin J and Fay M 2001).

As well as showing some degree of substrate selectivity, the individual isoforms also show selectivity for inhibitors. For example, sulfaphenazole is a specific inhibitor of CYP2C9 whereas quinidine is a potent and selective inhibitor of the isoform CYP2D6 (Martin J and Fay M 2001).

All isoforms will show some polymorphism. The frequency of these polymorphisms differs markedly between ethnic groups. These genetic differences mean some people have an enzyme with reduced or no activity. Patients who are 'slow metabolisers' may have an increased risk of adverse reactions to a drug metabolised by the affected enzyme (Martin J and Fay M 2001).

Variety of cytochrome P450 enzymes could be found in a single hepatocyte. An individual enzyme of cytochrome P450 may be able of metabolizing many different drugs; however, a given drug may be primarily metabolized by a single enzyme (Meyer JM and Rodvold KA 1996).

Drug interactions involving the cytochrome P450 system may not necessarily be clinically significant. There are several factors that help predict whether or not a drug interaction will be clinically significant. These include pharmacokinetic, pharmacodynamics, the wide variability of patient response to the same drug,
concomitant medical illness and factors relating to the route and timing of administration may also be important (Dresser GK et al 2000).

Cytochrome P450 1A2 (CYP1A2), is one of the cytochrome P450 mixed-function oxidase system and is responsible for the metabolism of xenobiotics in the body and involved in the synthesis of cholesterol, steroids and other lipids (Nelson DR et al 2004).

In humans, the CYP1A2 enzyme is encoded by the CYP1A2 gene. This cytochrome P450 enzyme is localized to the endoplasmic reticulum and its expression is induced by some polycyclic aromatic hydrocarbons (PAHs), some of which are found in cigarette smoke. It is of clinical importance due to the huge number of drug interactions associated with its induction and inhibition. Crucially, it is able to metabolize some PAHs to carcinogenic intermediates. Other xenobiotic substrates for this enzyme include caffeine, aflatoxin B1, and acetaminophen (Nelson DR et al 2004).

Caffeine is considered to be a model substrate for this enzyme due to its common consumption and safety. CYP1A2 enzyme activity may be altered by environmental factors including cigarette smoking, charbroiled meat, vegetables such as cabbages, cauliflower, broccoli and cruciferous and a number of drugs including rifampicin, phenobarbital and omeprazole (Runge D et al 2000).

1.1.5 Acetylation

In the IUPAC (International Union of Pure and Applied Chemistry) nomenclature, acetylation or ethanoylation is "a chemical reaction that describes the adding of an acetyl functional group into a chemical compound". The process of acetylation is essential in protein formation, drug biotransformation, regulation of deoxyribonucleic acid (DNA) and other genetic elements using histone acetylation.
In addition, acetylation mediated by acetyltransferase enzymes has been shown to be important in cancer and other human diseases (Robert K et al 2003 and Zhao SF et al. 2008).

**N-terminal Acetylation**

N-terminal acetylation is one of the most common co-translational covalent modifications of proteins in eukaryotes and it is crucial for the regulation and function of different proteins (Choudhary C et al 2009, Fritz KS et al 2012). N-terminal acetylation plays an important role in the synthesis, stability and localization of proteins (Robert K et al 2003 and Zhao SF et al 2008). About 85% of all human proteins and 68% in yeast are acetylated at their Nα-terminus (Van DP et al 2011).

Acetylation is an essential route of xenobiotics metabolism involving an aromatic amine (R-NH2) or a hydrazine group (R-NH-NH2), which are converted to aromatic amides (R-NH-COCH3) and hydrazides (R-NH-NH-COCH3), respectively as well as for a number of known carcinogens present in the diet, cigarette smoke and the environment (Starheim K et al 2012).

N-terminal Acetylation is catalyzed by a set of enzyme complexes, the N-terminal acetyltransferases (NATs). The reaction pathway is catalyzed by two cytoplasmic acetyltransferases, N-acetyltransferase Type I (NAT1) and N-acetyltransferase Type II (NAT2). These enzymes are located in the liver, lung, spleen, gastric mucosa, red blood cells (RBCs) and lymphocytes and the presence of co-factor acetyl coenzyme A is essential (Liston HL et al 2001). NATs transfer an acetyl group from acetyl-coenzyme A to the α-amino group of the first amino acid residue of the protein as shown in figure 4 (Starheim K et al 2012).
The genes encoding the two NAT proteins were first identified by Grant et al. (1991) who showed that each component consists of an intronless Open Reading Frame (ORF) of 870 base pairs. The two genes are 87% homologous and are located at 8p22, a chromosomal region commonly deleted in some human cancers (Grant DM et al 1991, Starheim K et al 2012).

The acetylation polymorphisms in NAT2 were observed in tuberculosis patients who received isoniazid and reported toxic side effects. This has been discovered over 60 years ago (Hughes HB 1954). In fact, the polymorphism was named the "isoniazid acetylation polymorphism" for several years until the significance of polymorphisms in the metabolism and disposition of other drugs and chemical carcinogens was fully appreciated (Weber WW and Hein DW 1985).

1.2 Statement of the Problem

Cytochrome P450s (CYPs) are responsible for the metabolism of a wide variety of clinically, physiologically and toxicologically important compounds (Guengerich FP. 2001). CYP families 1, 2 and 3 are responsible for 70–80% of phase I metabolism of clinically used drugs (Eichelbaum M, et al 2006) and are included in the biotransformation of a huge number of xenobiotics. Biotransformation can turn in bioactivation of pre-carcinogens to carcinogenic and drugs to toxic
metabolites or other toxic effects. The majority of CYP-mediated xenobiotic metabolism is carried out by polymorphic enzymes (Guengerich FP 2006).

Nevertheless, many reports showed that both genetic mutations and environmental factors influence the elimination of drugs that are metabolized by CYP1A2 (Ingelman-Sundberg M et al 2007; Zhou SF et al 2008, Gunes A and Dahl ML 2008 and Zhou SF et al 2009a).

Plasma levels of a drug can vary between individuals (Ingelman-Sundberg M 2001b and Cascorbi I 2006). In fact, genetic factors are one of the main factors that play important role in the differences between individuals in the response and susceptibility to drugs and xenobiotics (Ingelman-Sundberg M 2001a). Variability among individuals in drug metabolism can result in either reduce drug efficacies or increase risk of side effects, which are phenomena that may influence drug therapy and development.

Several adverse drug reactions reports related to specific drugs metabolic pathways that are catalysed by polymorphic CYP enzymes have been reported. For that, the complete understanding of the nature of CYP polymorphisms is crucial to gain a comprehensive view of the inter-individual differences in chemical exposure, adverse drug effects and toxicity (Ingelman-Sundberg M 2002).

CYP1A2 is an important enzyme that bioactivate a number of procarcinogens and some natural compounds such as aristolochic acids. In addition, this enzyme also metabolizes a number of essential endogenous compounds (Wang B and Zhou SF 2009). Several studies implicated CYP1A2 polymorphisms in cancer susceptibility (Suzuki H, et al 2008 and Altayli E, et al 2009) and other diseases (Pruiit KD et al 2012).
On the other hand, some studies have recorded that there is a strong association between the acetylation status and the risk of many diseases (Evan DA 1984, Clark DW 1985, Evan DAP 1993).

NAT enzymes polymorphisms have been suggested to be associated with risk of cancer, treatment responses and treatment resistance due to their roles in the activation or deactivation of xenobiotics. NAT2 enzymes have been implicated in chemical carcinogenesis pathways. For instance, several studies had examined the widespread occurrence of NAT2 genes polymorphisms in the population that have contributed to individual risk of having disease (Weber WW and Hein D W. 1985, Sim E et al 2008, Stanley LA and Sim E 2008, Sim E et al 2012, McDonagh EM et al 2014).

As mentioned earlier, NAT2 is a polymorphic enzyme that has essential roles in the deactivation or activation of numerous xenobiotics in humans. Based on the expression of its isoenzymes in the liver, the genetic variants of NAT2 have been correlated with drug metabolism, response and toxicity. In addition, drug metabolic rates and susceptibility to drug toxicity have varied according to acetylation phenotypes status which is conferred by NAT2 genotypes. More studies are crucial to help measure the clinical significant of NAT2 to determine the patient's dosage for maximum treatment efficacy and to avoid drug toxicity (Weber WW and Hein DW 1985, Sim E et al 2008, Stanley LA and Sim E 2008, Sim E et al 2012, McDonagh EM et al 2014).

**UAE Population**

Woolhouse et al (1997) who had conducted one study among Emiratis from the eastern region of Abu Dhabi Emirate determined the Polymorphic N-acetyltransferase (NAT2) genotypes in 106 unrelated Emiratis by PCR-RFLP analysis. This is the first study conducted to overview CYP1A2 and NAT2
polymorphism among nationwide Emiratis sample and to target the determination of
the alleles and genotype frequencies and their correlation with the phenotyping in
this population.

1.3 Literature Review

1.3.1 CYP1A2

1.3.1.1 Gene Localization, Protein Structure and Expression

A complete understanding of the gene location, structure, variation and
expression of CYP1A2 is crucial to gain a comprehensive view of the inter-individual
differences in chemical exposure, adverse drug effects and toxicity (Ingelman-
Sundberg M 2002).

1.3.1.1.1 The CYP1A2 Gene Family

CYP1A2 (Cytochrome P450, Family 1, Subfamily A, Polypeptide 2) is a
protein coding gene. This gene encodes one of the cytochrome P450 superfamily of
enzymes. The transcript from this gene includes four Alu sequences flanked by
direct repeats in the 3' untranslated region. In liver microsomes, CYP1A2 enzyme
plays a role in an NADPH-dependent electron transport pathway. It oxidizes a
number of structurally unrelated compounds, including steroids, fatty acids, and
xenobiotics. It is most active in catalyzing 2-hydroxylation. Caffeine is metabolized
primarily by cytochrome CYP1A2 in the liver through an initial N3-demethylation.
Moreover, it bioactivates the carcinogenic aromatic and heterocyclic amines. It
catalyzes the N-hydroxylation of heterocyclic amines and the O-deethylation of
phenacetin (Pruitt KD et al 2012).

The CYP1A2 gene is presented in a cluster with CYP1A1 and CYP1B1 on
chromosome 15 particularly, in the long arm region q24.1. (Zhou SF et al 2009a)
CYP1A2 is one of the major CYPs in human liver (~13–15%) compared to 2% with CYP2D6 (Shimada T et al 1994).

Moreover, 60-fold interindividual variation has been reported in CYP1A2 activity (Shimada T et al 1994 and Nakagawa K et al 2002). Also, more than 15 and 40-fold variation in mRNA and protein expression levels have been detected in the human liver (Ikeya K et al 1989 and Guengerich FP et al 1999).

The genetic component of variation in CYP1A2 activity is almost 75% and the remaining is related to environmental factors, such as smoking (induction), oral contraceptive use in women (inhibition), and others (Rasmussen BB et al 2002). However, in 2010 Klein et al provided a pathway-based analysis in human liver samples which demonstrated that the genetic variation of CYP1A2 activity accounted for 42, 38, and 33% of the catalytic activity, protein expression, and mRNA levels, respectively (Klein K et al 2010). In view of the predominant role of CYP1A2 in activation of toxic xenobiotics compared with its metabolism of prescription drugs, there are a lot of epidemiological reports testing the role of variant CYP1A2, metabolism of procarcinogens, and cancer risk (Zhou SF et al 2009a).

1.3.1.1.2 Reaction Mechanism, Structural Features and Functional Relevance of CYP1A2

CYP1A2 and CYP1A1 share a 5'-flanking region of approximately 23 kb, which includes shared regulatory elements, although the genes are positioned back to back and transcription occurs in opposite directions (Ueda R et al 2006). The CYP1A2 gene spans around 7.8 kb and comprising seven exons and six introns, the first exon 55-bp-long noncoding exon. CYP1A2 is a 515-residue protein with a molecular mass of 58 294 Da (Zhou SF et al 2009a).
The CYP1A1 and CYP1A2 genes are split by a 23-kb segment that includes no open-reading frames (Corchero J et al 2001). Between CYP1A2 and CYP1A1, exons 2, 4, 6, and particularly 5 are strikingly maintained in both nucleotides and total number of bases (Figure 5) (Ikeya K, et al 1989).

![Figure 5: CYP1A2 gene structure (Sachse et al 1999)](image)

Human CYP polymorphisms include mutations in both non-coding regions (promoters, introns) and protein-coding sequences (exons). Mutations in non-coding regions may result in changes in the expression levels, whereas mutations in protein-coding sequences may result in alteration in the structure and catalytic properties of the enzyme (Zhou H, et al 2004).

There are over 100 substrates reported for CYP1A2, including many clinically significant drugs; such as clozapine, tacrine, procancerogens (e.g. benzopyrene and aflatoxin b1), and endogenous substrates (e.g. steroids and arachidonic acid) (Zhou SF, et al 2009a). However, compared with other CYPs, there have been relatively few reports of pharmacogenetics (PGx) relationships. This can be explained by the small number of prescription drugs for which CYP1A2 is a metabolizing enzyme (9% compared with 37% for CYP3A4/5, 17% for CYP2C9, and 15% for CYP2D6). Indeed for several drugs, CYP1A2 is not the sole metabolizing enzyme, nor is it active at the rate-limiting step (Zanger UM, et al 2008). In addition, there are under-reported variants that impact CYP1A2 activity.
1.3.1.1.3 CYP1A2 Regulation

CYP1A2 is markedly expressed in liver and is inducible in the liver, lung, pancreas, gastrointestinal tract, and brain (Jorge-Nebert LF, et al 2010). Drug interactions have been reported many times in correlation with polymorphism of CYP1A2 that result in alteration of drug response (Zhou SF, et al 2009a). Smoking, dietary cruciferous vegetables, polyamine hydrocarbons from grilled meat, and omeprazole and other proton pump inhibitors have been shown to increase CYP1A2 activity (Gunes A and Dahl ML 2008). Oral contraceptives, fluvoxamine, and fluoroquinolone antibiotics inhibit CYP1A2 expression (Gunes A and Dahl ML 2008).

Similar to CYP1A1 and 1B1, CYP1A2 is primarily regulated by the aromatic hydrocarbon receptor (AhR), a ligand-activated transcription factor and a basic helix-loop-helix protein belonging to the Per-Arnt-Sim family of transcription factors (Wang B and Zhou SF et al 2009).

The expression regulation of CYP1A1/1A2 is complex because gene transcription not only includes the AhR but also a number of transcription factors, and is potentially influenced by the actions of transcriptional coactivators and corepressors. AhR-mediated signaling pathways show a first line of defense against potentially toxic environmental contaminants. While, metabolic induction procedures by the AhR may produce highly carcinogenic metabolites, generating a link between AhR activation and chemical carcinogenesis (Wang B and Zhou SF 2009).

1.3.1.1.4 CYP1A2 Variants and its Molecular Epidemiology

CYP1A2 variants have been reported several times, with some implications and alteration in drug metabolism (Browning SL et al 2010).
More than, 177 single-nucleotide polymorphisms (SNPs) of human CYP1A2 upstream sequence have been determined (from which 41 variant alleles (*1B to *21) were recorded), illustrating the alterations on DNA sequence levels and in some cases their epidemiological frequencies (Browning SL, et al 2010). Among the SNPs located in seven exons, there are 22 non-synonymous that alter amino acid sequence in exons 1-7 of CYP1A2 (Figure 6) (Zhou SF et al 2009a).

![Figure 6: Nonsynonymous SNPs of the CYP1A2 gene](image)

The variability in liver expression of the CYP1A2 gene reach almost 40% and 60% variability in caffeine metabolism; the most often probe drug used for CYP1A2 (Gunes A and Dahl ML. 2008). Only few variants could clearly explain the phenotypic variability in CYP1A2 gene expression contrarily it is much more with other drug-metabolizing CYPs (Ghotbi R et al 2007). Ghotbi et al in 2007 observed a very low frequency of the coding sequence variants in Caucasian and Asian populations (Ghotbi R et al 2007). Browning et al in 2010 examined CYP1A2 variation in Ethiopians, suggested that because of the overall greater incidence of variation, including some novel presumable deleterious variants, there could be some individuals freed from any CYP1A2 activity in this population (Browning SL, et al 2010).

Based on epidemiological studies, it has been clearly observed that there is significant ethnic variability in the distribution of common and rare CYP1A2 SNPs and haplotypes (Zhou SF et al 2009a). The G-3860A (*1C) SNP has less frequency in Caucasians than in Asians (0.21–0.25) (Chida M et al 1999). The frequency of *1C is significantly lower in the Turkish population (0.04) (Bilgen T et al 2008) than in Japanese (0.21) (Chida M et al 1999) and Chinese (0.25), whereas it is almost equal in the Turkish and Egyptian populations (0.07) (Hamdy SI et al 2003).

The frequency of the -2467delT (CYP1A2*1D) allele is lower in Caucasians compared with Asians and Africans. The allelic frequency of the *1D allele is 4.1–7.9% in Caucasians (Sachse C et al 2003, Skarke C et al 2005). However, the -2467T deletion appeared at higher level frequencies of 42.0–43.8% in Japanese (Soyama A et al 2005, Chida M et al 1999) and of 40% in Egyptians (Hamdy SI et al 2003). In Turkish, the frequency of CYP1A2*1D is very high (92%) (Bilgen T et al 2008). The -739T>G SNP (allele CYP1A2*1E and CYP1A2*1G) is frequent in Ethiopians (0.10) (Aklillu E et al 2003), Saudi Arabsians (0.096) (Aklillu E et al 2003),
and Japanese (0.082) (Chida M et al 1999). However, British (Sachse C et al 2003), German (Skarke C et al 2005), Spaniard (Aklillu E et al 2003), Turkish (Bilgen T et al 2008), and Egyptian (Hamdy SI et al 2003) populations have a low frequency for this allele (0.0044, 0.016, 0.017, 0.01, and 0.03, respectively) (Table 2).


<table>
<thead>
<tr>
<th>Ethnic group</th>
<th>No. of subjects</th>
<th>G-383AA (°1C)</th>
<th>A-3113G</th>
<th>-2487delT</th>
<th>T-739G (°1E, °1G, °1J, and °1&lt;)</th>
<th>C-729T (°1K)</th>
<th>C-163A (°1F, °1J, and °1&lt;)</th>
<th>T-684TC (°1B, °1N, °1Q, °2, °8, °15, and °16)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caucasian</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>British</td>
<td>114</td>
<td>0.009</td>
<td>0.048</td>
<td>0.004</td>
<td></td>
<td>0.333</td>
<td>0.382</td>
<td></td>
</tr>
<tr>
<td>German</td>
<td>236</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italian</td>
<td>95</td>
<td>0.04</td>
<td>0.24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spanish</td>
<td>117</td>
<td>0.008</td>
<td>0.023</td>
<td>0.193</td>
<td></td>
<td>0.003</td>
<td>0.286</td>
<td></td>
</tr>
<tr>
<td>Swedish</td>
<td>194</td>
<td>0.008</td>
<td>0.023</td>
<td>0.193</td>
<td></td>
<td>0.003</td>
<td>0.286</td>
<td></td>
</tr>
<tr>
<td>Italian</td>
<td>1,170</td>
<td>0.04</td>
<td>0.92</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turkish</td>
<td>110</td>
<td>0.04</td>
<td>0.92</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese</td>
<td>27</td>
<td>0.22</td>
<td>0.09</td>
<td>0.50</td>
<td></td>
<td>0.30</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td>Japanese</td>
<td>1,150</td>
<td>0.21</td>
<td>0.42</td>
<td>0.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japanese</td>
<td>250</td>
<td>0.24</td>
<td>0.03</td>
<td>0.44</td>
<td></td>
<td>0.03</td>
<td>0.32</td>
<td>0.19</td>
</tr>
<tr>
<td>Korean</td>
<td>150</td>
<td>0.267</td>
<td>0.027</td>
<td>0.707</td>
<td></td>
<td>0.027</td>
<td>0.373</td>
<td></td>
</tr>
<tr>
<td>Saudi Arabian</td>
<td>173</td>
<td>0.10</td>
<td>0.41</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>African</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Egyptian</td>
<td>212</td>
<td>0.07</td>
<td>0.40</td>
<td>0.03</td>
<td></td>
<td>0.68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethiopian</td>
<td>173</td>
<td>0.10</td>
<td>0.10</td>
<td>0.03</td>
<td></td>
<td>0.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tanzanian</td>
<td>71</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tunisian</td>
<td>98</td>
<td>0.07</td>
<td>0.075</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zimbabwean</td>
<td>143</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.43</td>
</tr>
</tbody>
</table>

The existence of allelic imbalance in CYP1A2 expression and the importance of epigenetic genetic variation in impacting CYP1A2 mRNA expression and enzyme activity in vitro using human liver were also reported (Ghotbi R et al 2007). Epigenetic and environmental factors play a role in identifying interindividual and interethnic variability in CYP1A2 expression and enzyme activity, apart from genetic variation.
Compared to other CYP enzymes, the understanding of the pharmacogenomic effects of CYP1A2 variation is still at an early stage. There is also the need to fully evaluate all other variants before assigning haplotypes (Soyama A et al 2005, Ghotbi R et al 2007, Browning SL et al 2010).

**Important Variants in CYP1A2**

The most frequently studied polymorphisms are G-3860A (allele *1C), -2467delT (*1D), T-739G (*1E), and C-163A located in intron 1 (*1F), which have been initially reported in a Japanese population (Zhou SF et al 2009a). These polymorphic CYP1A2 alleles present differences in the promoter region and are associated with changed enzyme activity particularly, allele *1C, *1D, *1F, and *1K (Zhou SF et al 2009a).

**rs762551, CYP1A2: C-163A**

This SNP is the most well-studied genetic variant in CYP1A2. It is the sole variant of the CYP1A2*1F haplotype and found with other variants in several haplotypes (CYP1A2*1J, CYP1A2*1K, CYP1A2*21, and others that have not been confirmed). It is located in the intron between the noncoding exon 1 and exon 2, where the coding sequence begins and causes increase in enzyme activity (Gunes A and Dahl ML 2008). The frequency of rs762551:C>A varies widely with populations: the A allele (slow) frequencies range from 0.3 to 0.39 in Asians, from 0.4 to 0.51 in Blacks or African Americans, and from 0.29 to 0.33 in Whites (Gunes A and Dahl ML. 2008).

**rs2069514, CYP1A2: G-3860A**

The *1C haplotype contains G-3860A in the 5'-flanking region of CYP1A2, originally called G-2964A (Obase Y et al 2003).
It was initially reported in Japanese with a frequency of 23.3% (Nakajima M et al 1999). This allele showed decrease in enzyme activity based on the measurement of the rate of caffeine 3-demethylation in Japanese smokers, that could be due to the decreased inducibility and expression of the enzyme (Nakajima M et al 1999).

The CYP1A2*1C haplotype was found at frequencies of 7% in Blacks or African Americans, at 6–25% in Asians, and at 0.4–1% in Whites (Nakajima M et al 1999, Hamdy SI et al 2003, Sachse C et al 2003, Tiwari AK et al 2005, Takata K et al 2006, Djordjevic N et al 2010). This variant was associated with a reduced rate of caffeine demethylation in Japanese smokers but not in non-smokers. The variant has changed a transcription factor binding site in the gene promoter; however the factor was not defined (Obase Y et al 2003).

**rs12720461, CYP1A2: C-729T**

CYP1A2: C-729T is also called CYP1A2: C-730T. This variant was shown in vitro to affect binding of an E-twenty six transcription factor (Aklillu E et al 2003). The *1K haplotype was associated with 40% lower inducibility in vitro, and non-smokers heterozygous for *1K had significantly lower CYP1A2 activity compared with the wild type (Aklillu E et al 2003). Alkillu et al (2003) stated that this variant had lower inducibility with 2,3,7,8-tetrachlorodibenzo-p-dioxin in human hepatoma cells and may affect bioactivation and sensitivity to carcinogens. This could illustrate the lower caffeine metabolism observed for the *1K haplotype, which is the only haplotype that contains this variant.

CYP1A2*1K was reported to be rare in Swedes (0.3%) and absent in Koreans (Ghotbi R et al 2007). However, the effect of allele CYP1A2*1K could not be found in these populations.
Other important variants includes rs56276455, 5347T>C (exon 4, causing Asp348Asn and the synonymous 5347T>G (Asn516Asn)), rs72547516, 2499A>T (exon 5, leading to Ile386Phe) and rs28399424, 5090G>T (exon 7, leading to Arg431Trp) which represent allele CYP1A2*3, CYP1A2*4 and CYP1A2*6, respectively. All of them are associated with decreased gene expression and enzyme activity that has been identified in the French population with frequency of 0.5-1% (Chevalier D et al 2001, Allorge D et al 2003, Zhou H et al 2004).

Other CYP1A2 Variants

CYP1A2*1J (C-163A; T-739G) and *1K (C-163A; T-739G; C-729T, all located in intron 1) have been initially identified in Ethiopian non-smokers (Aklillu E et al 2003). The T-739G had a frequency of 3.2% in Japanese; however, the C-729T SNP was not reported in the Japanese population (Soyama A et al 2005).

CYP1A2*2 (63C>G), mutation that causes a Phe21Leu substitution, which was initially identified in one subject out of 157 Chinese subjects with an allele frequency of 0.32% (Huang JD et al 1999). Its functional impact is unknown. The CYP1A2*2 allele was not found in British (n = 114) (Sachse C et al 2003) and Italian populations (n = 500) (Pucci L et al 2007). The *5 haplotype entails 3496G>A in exon 6 resulting in Cys406Tyr (Zhou SF et al 2009a).

CYP1A2*7 (3533G>A) mutation at the splice donor site of intron 6 which has been reported in a 70-year-old patient who had very high plasma concentrations of clozapine when administered at normal dose. This SNP caused RNA splicing defect and lead to a loss of CYP1A2 activity (Zhou SF et al 2009a).

1.3.1.1.5 CYP1A2 Important Haplotypes

Almost 41 CYP1A2 haplotypes have been published on the Human Cytochrome P450 Allele Nomenclature Committee home page. Twenty four
haplotypes are related to nucleotides changes in the non-coding region and 17 haplotypes are related to nucleotides changes in the coding region of the CYP1A2 protein.

There are several studies that characterize gene polymorphisms in the non-coding region, however, only few reports describe non-synonymous protein polymorphisms. Murayama N et al (2004), Saito Y et al (2005) and Zhou SF et al (2004) characterized most of the non-synonymous polymorphisms with a limited number of substrates.

A haplotype can refer to a combination of alleles or a set of single nucleotide polymorphisms (SNPs). This set of alleles is located on one chromosome. Thus, the alleles making up a haplotype can be located in different places on the same chromosome but they are inherited together.

An allele is one of two or more forms of the DNA sequence of a particular gene. Sometimes, different DNA sequences (alleles) may result in different gene expression or have the same result in the expression of the gene.

Many studies have focused on certain alleles as they alter the gene expression. This alteration results in changing the catalytic activity of the enzyme encoded by this gene which in turn reflects inter-individual differences in the phenotype status. The most relevant alleles are briefly described in the following section and listed in table 3.

CYP1A2*1A

This is the wild-type allele to which all variants are compared to and it is associated with normal gene expression.
CYP1A2*1F

The *1F haplotype has been reported in many studies with a changed phenotype (Ghotbi R et al 2007). This haplotype associated with increased gene expression which in turn results in increased enzyme activity. However, the phenotype effect is seen only in the availability of an inducer, such as smoking or heavy coffee consumption (Djordjevic N et al 2008, Djordjevic N et al 2010).

There are some conflicts in the literature regarding the designation of this haplotype (Ingelman-Sundberg M et al 2007). The variant that defines this haplotype, CYP1A2: C-163A (rs762551), has different frequencies in different populations and the assignment of major and minor alleles therefore varies. Most publications have listed *1F as C>A (Ingelman-Sundberg M et al 2007), but others have A>C (Cornelis MC et al 2006).

CYP1A2*1C

The only single-nucleotide polymorphism (SNP) for this haplotype is G-3860A (rs2069514) has been well-defined and also seen in the literature as G-2964A (Obase Y et al 2003). This haplotype is associated with decreased gene expression.

CYP1A2*1K

There are three variants that define this haplotype: T-739G (rs2069526), C-729T (rs12720461) and C-163A (rs762551). This allele was shown firstly in Ethiopians at a frequency of 3% (n = 173), in Spaniards at a frequency of 0.5% (n = 117), and in Saudi Arabians at 3.6% (n = 136) (Aklillu E et al 2003). CYP1A2*1K allele has significantly reduced CYP1A2 enzyme activity in non-smokers compared with *1A or *1F, using caffeine as a probe substrate (Aklillu E et al 2003).
This haplotype has not been reported in Japanese \( (n = 350) \) (Takata K et al 2006) and Koreans \( (n = 50) \) and occurs at a very low frequency in Swedes \( (n = 193, 0.3\%) \) (Ghotbi R et al 2007).

**CYP1A2*3**

There are two variants that define this haplotype 2116G>A and 5347T>C (rs56276455). This haplotype is associated with decreased enzyme activity (Chevalier D et al 2001, Zhou H et al 2004).

**CYP1A2*4**

There is one SNP for this haplotype 2499A>T (rs72547516). This haplotype is associated with decreased enzyme activity (Chevalier D et al 2001, Zhou H et al 2004).

**CYP1A2*6**

There is one SNP for this haplotype 5090C>T (rs28399424). This haplotype is associated with decreased enzyme activity (Chevalier D et al 2001, Zhou H et al 2004).

### 1.3.1.16 CYP1A2 Alleles Nomenclature/Variant Annotations

**Table 3:** CYP1A2 alleles, their references SNPs and catalytic activity based on the Human Cytochrome P450 Allele Nomenclature Committee

<table>
<thead>
<tr>
<th>CYP1A2 Allele</th>
<th>Location</th>
<th>Nucleotide Change(s) and rs #</th>
<th>Amino Acid Change(s)</th>
<th>Enzyme activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>*1A</td>
<td>Reference</td>
<td></td>
<td>Reference</td>
<td>Normal</td>
</tr>
<tr>
<td>*1C</td>
<td>5'-flanking region</td>
<td>G-3860A (rs2069514)</td>
<td>Gua2964Ade</td>
<td>Decreased</td>
</tr>
<tr>
<td>*1F</td>
<td>Intron between the noncoding exon 1 and exon 2</td>
<td>C-163A (rs762551)</td>
<td>Ile386Phe</td>
<td>Higher inducibility</td>
</tr>
<tr>
<td>*1K</td>
<td>Intron 1 region</td>
<td>C-729T (rs12720461)</td>
<td>Cyt873Thy</td>
<td>Decreased</td>
</tr>
<tr>
<td>*3</td>
<td>Exon 4</td>
<td>2116G&gt;A (rs56276455)</td>
<td>Asp348Asn</td>
<td>Decreased expression</td>
</tr>
</tbody>
</table>
1.3.1.1.7 CYP1A2 Genotype/Phenotype Relationship

Caffeine is a methylxanthine naturally available in some beverages and also considered as a pharmacological agent. Its main pharmacological effect is as a central nervous system stimulant. Caffeine is naturally occurring xanthine derivative like theobromine and the bronchodilator theophylline. Caffeine biotransformation occurs in several steps involving hepatic cytochrome P4501A2 (CYP1A2) (figure 7). Caffeine is almost completely metabolized with 3% or less being excreted unchanged in urine (Begas E et al 2007, Marta K and Wadysawa DA 2008a).

The main route of metabolism in humans (70-80%) is via N-3 demethylation to paraxanthine (1,7-dimethylxanthine or 17X) (Begas E et al 2007, Marta K and Wadysawa DA 2008a and Benowitz N L et al 1995). This reaction is carried out by CYP1A2 in the liver (Begas E et al 2007). Experimentally, human liver microsomes estimate that N-1 demethylation to theobromine (3,7-dimethylxanthine) and N-7 demethylation to theophylline (1,3-dimethylxanthine) accounts for approximately 7 to 8% of caffeine metabolism for each (Marta K and Wadysawa DA 2008b). The remaining 15% or less of caffeine undergoes C-8 hydroxylation to form 1,3,7-trimethyluric acid (Figure 7) (Marta K and Wadysawa DA 2008b).
Figure 7: Schematic representation of caffeine metabolites. CYP1A2 catalyzes the 3-demethylation of caffeine to paraxanthine and NAT2 catalyzes the acetylation of paraxanthin metabolites to AMFU.

As discussed earlier, CYP1A2 is responsible for more than 95% of the primary metabolism of caffeine (Kalow W and Tang B K 1993). Therefore caffeine is used as a probe drug for CYP1A2 activity with the relative ratios of urinary metabolites used as an indicator of the flux through different parts of the pathway (Begas E et al 2007). As well as paraxanthine, the major metabolites of caffeine in urine are 1-methylxanthine (1MX), 1-methyluric acid (1MU), 5-acetylamino-6-formylamino-3-methyluracil (AFMU) and 1,7-dimethyluric acid (17MU) (Begas E et al 2007).

Based on relative ratios of urinary metabolites, the ratio of paraxanthine/caffeine (17MX/137MX) has been measured from plasma after 4 hours of caffeine consumption to compare CYP1A2 genetic polymorphism, enzyme activity
and the genotype-phenotype relationship in Sweden and Korean (Ghotbi R et al 2007).

The urinary molar ratio of \((17\text{MX} + 17\text{MU}) / 137\text{MX}\), taken at 4-5 h after caffeine ingestion, was identified from pharmacokinetic analyses as being better correlated for both NAT2 and CYP1A2 phenotyping (Butler MA et al 1992).

Evaluation of caffeine as an in vivo probe for CYP1A2 using measurements in plasma, saliva, and urine have been reported by Carrillo JA et al (2000) where 100 mg of oral caffeine has been consumed by subject to estimates CYP1A2 activity. They reported that both plasma and saliva total clearances of caffeine were highly correlated with each other (Carrillo JA et al 2000).

The ratio \(17\text{MX}/137\text{MX}\) restricted to one sampling point taken 4 hours after dose, where highly correlated in plasma and saliva with the standard reference (Carrillo JA et al 2000). Moreover, the ratio \((\text{AFMU} + 1\text{MU} + 1\text{MX} + 17\text{MU} + 17\text{MX})/137\text{MX}\) in a 0–24 hours urine sampling showed the highest correlation with the standard reference. (Carrillo JA et al 2000)

In conclusion they suggest the use of \(17\text{X}/137\text{X}\) ratio for plasma or saliva, whereas \((\text{AFMU} + 1\text{U} + 1\text{X} + 17\text{U} + 17\text{X})/137\text{X}\) ratio is the recommended one for urine samples through a sampling interval of at least 8 hours, starting at time zero since caffeine intake (Carrillo JA et al 2000).

1.3.1.1.8 Clinical Impact, Drug Metabolism and Substrate Specificity of CYP1A2 Polymorphism

CYP1A2 is an important enzyme together with CYP1A1 and 1B1 that bioactivates a number of procarcinogens including polycyclic aromatic hydrocarbons (e.g., benzo(a)pyrene), heterocyclic aromatic amines/amides (e.g. 2-amino-1-methyl-6-phenylimidazo (4,5-b)pyridine), mycotoxins (e.g. aflatoxin B(1)) and some
natural compounds such as aristolochic acids present in several Chinese herbal medicines. This enzyme also metabolizes a lot of essential endogenous compounds including retinols, melatonin, steroids, uroporphyrinogen and arachidonic acids. (Wang B and Zhou SF 2009)

Several studies have indicated implications of CYP1A2 polymorphisms in cancer susceptibility, namely for pancreatic cancer (Suzuki H et al 2008), bladder cancer (Altayli E et al 2009) and colorectal cancer (He X et al 2014).

Diseases associated with CYP1A2 include toxicity or absent response to clozapine and porphyria cutanea tarda (Pruitt KD et al 2012).

Huge interindividual differences in the elimination of drugs that are metabolized by CYP1A2 have been reported, and referred to both genetic mutations and environmental factors (Ingelman-Sundberg M et al 2007, Zhou SF et al 2008, Gunes A and Dahl ML 2008 and Zhou SF et al 2009b).

Xenobiotic metabolism is varied widely between individuals. Plasma levels of a drug can vary more than 1000-fold between individuals upon the same drug dosage (Ingelman-Sundberg M 2001b, Cascorbi I 2006). These differences can result from induction or inhibition and genetic polymorphisms of metabolizing enzymes, or from physiological, pathophysiological and environmental factors (Cascorbi I 2006). In fact, genetic factors are indicated to be responsible for 20–40% of the differences between individuals in the response and susceptibility to drugs and xenobiotics (Ingelman-Sundberg M 2001a). Variability among individuals in drug metabolism can result in decrease drug efficacies or undesirable toxic side effects.

There are wide interindividual differences (10- to 200-fold) in CYP1A2 (also called phenacetin O-deethylase) expression and activity (Gunes A and Dahl ML 2008). Approximately 15- and 40-fold interindividual variations in CYP1A2 mRNA
and protein expression levels have been reported in human livers (Ikeya K et al 1989). These observations indicate a genetically measured difference in constitutive and/or inducible CYP1A2 gene expression. Unimodal, bimodal, and trimodal distributions of CYP1A2 activity when determined by caffeine urinary metabolic ratios have been reported in different studies populations (Shimizu T et al 1991).

Pharmacogenetic studies on the impact of common CYP1A2 polymorphisms on drug clearance and response are still under investigation. More well-designed studies are important to explore the genotype-phenotype relationships of CYP1A2 in terms of drug clearance and response. Personalized pharmacotherapy and individualized dosing of drugs may require incorporation of both genetic and environmental factors.

Many clinical studies have been established to test the impact of CYP1A2 polymorphisms on drug clearance and drug response. For example, Obase Y et al (2003) examined the effect of genetic polymorphisms in the 5'-flanking region to intron 1 of the CYP1A2 gene on theophylline metabolism in Japanese asthmatic patients (Obase Y et al 2003).

Among asthmatic patients, theophylline clearance was significantly lower in patients with the polymorphism at site G-2964A whose genotype was G/A (intermediate) or A/A (slow) than in those whose genotype was G/G (rapid). Therapeutic drug monitoring may be required in patients with the A allele at site 2964 in the CYP1A2 gene, due to low clearance level in those patients. Resistance to clozapine therapy has been reported in smoking schizophrenic patients due to low plasma drug levels harboring the 163 C/C (CYP1A2*1F) genotype (Ozdemir V et al 2001, Eap CB et al 2004).

Eap et al. (2004) recorded four smoking patients who did not respond to clozapine therapy at usual dosage and they observed that those individuals were
ultra-rapid metabolizers of CYP1A2 carrying the *1F allele (Eap CB et al 2004). On the other hand, higher plasma concentrations of clozapine and its metabolite N-desmethylclozapine have been found in patients carrying two CYP1A2 variants associated with reduced enzyme activity (-3860A, -2467del, -163C, -739G, and/or -729T) compared with those with one or none (Melkersson KI et al 2007).

Subjects with C-163C or C-163A genotype had lower plasma melatonin concentrations after administration of 6 mg melatonin compared with A-163A carriers (Hartter S, et al 2006). Another study in patients with rheumatoid arthritis (n = 105) found that the *1F haplotype (C-163A) affected the risk for leflunomide-induced organ toxicities (Bohanec GP et al 2008). Patients with A-163A genotype had a 9.7-fold higher risk for overall leflunomide-induced toxicity compared with patients with the CYP1A2*1F C/A or C/C genotype (Bohanec GP et al 2008).

Porphyria cutanea tarda (PCT) is considered to be genetically predisposed to development of clinically overt disease through mutations and polymorphisms in genes associated with known precipitating factors. Christiansen et al. (2000) elucidated that the frequency of the highly inducible C/C (rapid) genotype at site C-163A was increased in both familial and sporadic PCT. The authors presented that the C/C genotype was a susceptible factor for PCT (Christiansen et al 2000).

Paolini M et al. (1999) demonstrated that corresponding high levels of CYPs in humans would predispose an individual to cancer risk from the widely bioactivated tobacco-smoke procarcinogens. No association has been reported between CYP1A2 activity and sex and age based on Bebia Z et al (2004) in bioequivalence study when they determined to which extent age and sex influence CYP activity (Bebia Z et al 2004).

Acetaminophen as an effective analgesic and antipyretic agent, but causes serious hepatic and renal toxicity at high doses by conversion of acetaminophen
to the toxic intermediate N-acetyl-p-benzoquinone imine (NAPQI) through CYP1A2, CYP2E1, and CYP3A4 (Sarich T et al 1997). In addition, using CYP1A2 enzyme inhibitors, it may increase risk of toxicity of lidocaine due to clearance inhibition (Olkkola KT et al 2005).

The correlation was reported between caffeine breath test (CBT) results and tacrine blood levels as the first evidence supporting a critical role for CYP1A2 activity in the disposition of this drug in vivo (Fontana RJ et al 1996).

Mexiletine is mainly catalyzed by CYP2D6 and partially catalyzed by CYP1A2 (Kusumoto M et al 2001). Xenobiotics inducing CYP1A1 mRNA expression have been found to induce CYP1A2 and CYP1B1, varied only based on the level and time course of induction (Krusekopf S et al 2003).

Generally, the 15q24.1 locus, including CYP1A2, is associated with blood pressure (BP). Higher CYP1A2 activity on the basis of positivity of the CYP1A2*1F allele was linearly associated with lower BP after quitting smoking, but not while smoking. In non-smokers, the CYP1A2 variants were associated with higher reported caffeine intake, which in turn was associated with lower BP (Guessous I et al 2012).

Some studies have also used caffeine as a modulator of disease etiology, looking at intake of caffeine containing foods and beverages with respect to disease risk. Two independent genome-wide association studies have examined the effect of variants on caffeine intake. They found associated variants in the regulatory region of CYP1A2 (rs2472304 and rs2472297) and in its regulator AhR (rs4410790 and rs6968865) (Cornelis MC et al 2011, Sulem P et al 2011).

CYP1A2 has been found to be essential for dosing of several antipsychotics and for assessing both drug efficacy and adverse drug reactions. CYP1A2 is the
main CYP isoform responsible for clozapine metabolism (Eap CB at el 2004). Case studies presented ultrarapid metabolizers of clozapine that showed resistant to treatment. Increased clozapine doses and co-administration with the CYP1A2 inhibitor fluvoxamine were considered an alternative way to improve patient outcomes (Ozdemir V et al 2001, Eap CB et al 2004).

CYP1A2 also alters the antithrombotic drug clopidogrel (Wang Z et al 2015). Smoking elevates clopidogrel-mediated platelet inhibition (Bliden KP et al 2008, Gremmel T et al 2009). This is related to induction of CYP1A2, which can play a significant role in clopidogrel activation.

Like many of other CYPs, CYP1A2 is subjected to induction and inhibition by a number of compounds. Particularly, antofloxacine, carbamazepine, dihydralazine, furafylline, isoniazid, rofecoxib, clorgyline, thiabendazole, and zileuton are mechanism-based inhibitors of CYP1A2. Reversible and irreversible inhibition of CYP1A2 that may give an illustration of some clinical drug-drug interactions.

The process of drug interactions is a complex process because the medication interacts with one substrate of a specific cytochrome P450 pathway, does not mean it influences all substrates of that isozyme. Genetics, age, nutrition, stress, liver disease, hormones, and other endogenous chemicals also influence drug metabolism.

**Substrate Specificity of Human CYP1A2**

Phenacetin O-deethylation, caffeine 3-demethylation and ethoxyresorufin O-deethylation are a commonly used marker reactions for CYP1A2. Particularly, phenacetin O-deethylation is the most common marker reaction for CYP1A2 activity in the *in vitro* study. (Yuan R et al 2002).
To a less extent, tacrine (1,2,3,4-tetrahydro-9-aminoacridine, 1-hydroxylation), melatonin (6-hydroxylation) and tizanidine (oxidation) are also used as probes for CYP1A2 in vivo. For in vitro studies, phenacetin (O-deethylation), 7-ethoxycoumarin (O-deethylation), and 7-ethoxyresorufin (O-deethylation) are also used probes for determining CYP1A2 activity (Waxman DJ and Chang TK 2006).

1.3.2 N-acetyltransferase-2 (NAT2)

1.3.2.1 Gene Location, Structure and Expression

Almost 22 NAT-like genes have been reported in 14 different prokaryotic and eukaryotic species (Butcher NJ et al 2002). The genes invariably have an intronless coding sequence that encodes for a protein between 254 and 332 amino acids in length (Butcher NJ et al 2002).

1.3.2.1.1 N-acetyltransferase Gene Family

Human NAT1, NAT2 as well as NATP1 (pseudogene) were localized to the short arm of chromosome 8, more specifically in region 8p22. Both NAT1 and NAT2 genes show pronounced allelic variation (Hickman D et al 1994, Blume M et al 1990, Butcher NJ et al 2002, Stanley LA and Sim E 2008, Hein DW 2009). The NAT loci are segregated by 170–360 kb only and are in the orientation NAT1 → NATP1 → NAT2, with NAT1 being on the centromeric side of marker D8S261 and NAT2 coinciding with marker D8S21 (Figure 8) (Matas N et al 1997).
The two functional NAT genes possess an 87.5% nucleotide identity, which reach up to 81% homology at the amino acid level and around 80% with the corresponding sequence in NATP (Blume M et al 1990). The whole transcript of NAT1 is derived from a single exon; whereas NAT2 is derived from the protein encoding exon together with a second noncoding exon of 100 bp located about 8 kb upstream of the translation start site (Blume M et al 1990, Ebisawa T and Deguchi 1991).

1.3.2.1.2 Reaction Mechanism, Structural Features, Variation and Functional Relevance of NAT2

The reaction is composed of two steps. Initially, acetyl coenzyme A binds to the enzyme and the acetyl moiety is shifted from the cofactor to a cysteine (Cys68 for the human isoforms) of the protein. The second step occurs when the binding of substrate to the acetylated enzyme where the acetyl moiety is transferred to the
substrate. Finally, the acetylated product is separated from the enzyme (see aforementioned Figure 4) (Minchin RF et al 1992, Delomenie et al 2001).

Sinclair (2000) reported the first crystal structure of an arylamine N-acetyltransferase. A cysteine-histidine-aspartate catalytic triad was recorded in the N-terminus of the protein (Sinclair JC 2000). In this regard, the protein has been divided into three domains. The first consists of a helical bundle, located from amino acid 1 to approximately 90, which forms one side of a cleft in which the cysteine involved in acetyl transfer resides.

The second domain consists of residues from approximately 90 to 210 amino acids and is located on the other side of the cleft. It mostly consists of β-sheet structures. The last domain at the carboxyl terminus is a combination of β-sheets and α-helices, which represents the biggest variation between species (Sinclair JC 2000).

Nine missense mutations where the alteration is in an amino acid (C-190T, 191G>A, 341T>C, 434A>C, 499G>A, 590G>A, 803A>G, 845A>C, and 857G>A), and four silent where it occurs in the noncoding region (111T>C, 282C>T, 481C>T, and 759C>T) have been recorded in different studies (Grant DM et al 1991, Vatsis KP 1995, Hein DW et al 2000). The NAT2 gene has one non-coding exon around 8.6kb upstream of the intronless ORF (Figure 9) (Ebisawa T and Deguchi T 1991, Boukouvala S and Sim E 2005, Husain A et al 2007).

![Figure 9: Structure of human NAT2 gene](image)

Leff et al (1999) studied several different human NAT2 alleles in a yeast expression system. They observed that three novel alleles, namely NAT2*5D
haplotype (341T>C), NAT2*14G haplotype (191G>A, 282C>T, and 803A>G), and NAT2*6D haplotype (111C>T, 282C>T, and 590G>A), expressed proteins that had N- and O-acetylation capacities similar to the expressed protein of the commonly occurring slow NAT2*5B allele, and significantly less than that of the wild-type NAT2*4 allele.

The expression of NAT2*5B and NAT2*5D haplotypes was reported to be significantly lower than that of the wild-type protein, suggesting that the base substitution at position 341, which is common to the NAT2*5 cluster, is sufficient for reduction in NAT2 protein expression. This was not observed to be the case for NAT2*6D and NAT2*14G haplotypes, which were expressed at levels comparable to wild-type. Contrarily, NAT2*6D and NAT2*14G haplotypes were observed to be significantly less stable than the wild-type.

1.3.2.1.3 NAT2 Regulation

N-acetyltransferases (NATs) are xenobiotic metabolizing enzymes for which three distinct enzymatic activities have been described. One is includes the acetyl coenzyme A (CoA) dependent N-acetylation of arylamines and arylhydrazines, a reaction usually associated with xenobiotic detoxification (Hein DW 2009).

The second one is also acetyl-CoA dependent and involves O-acetylation of N-hydroxyarylamines (Hein DW 2009), typically generated through N-oxidation of arylamines by cytochrome P450 enzymes.

The third one is an acetyl-CoA independent N,O-acetyltransfer performed on N-arylhydroxamic acids, generating highly reactive mutagenic compounds that bind to DNA. NATs have essential roles in the metabolism and detoxification of xenobiotics and therapeutic drugs, and are implicated in cancer risk because of their role in the activation or detoxification of carcinogens and their interaction with

1.3.2.1.4 NAT2 Variation and its Epidemiology

NAT2 is a polymorphic gene with almost 88 NAT2 alleles which have been assigned official symbols by the Arylamine N-acetyltransferase Gene Nomenclature Committee, according to consensus guidelines (Hein DW et al 2008). NAT2*4 is the reference (or "wild type") allele for the respective gene, and the other variant alleles differ from this by one or more single nucleotide polymorphisms (SNPs).

The NAT2 gene has a high frequency of functional variation, which varies amongst populations that are ethnically different, and has high levels of haplotype variation (Patin E et al 2006b, Mortensen HM et al 2011). SNPs within the NAT2 gene cause several changes in NAT2 function by altering enzyme stability, altering affinity for substrate, or a protein that is targeted for proteasome degradation (Walraven JM et al 2008, Hein DW 2009).

NAT2 genotypes have been classified into three different phenotypes: "slow acetylator" (two slow alleles), "intermediate acetylator" (1 slow and 1 rapid allele), and "rapid" acetylator (2 rapid alleles) (Stanley LA and Sim E 2008). Some papers report rapid as any genotype containing NAT2*4 and slow as any non-carriers of NAT2*4 acetylators (Soejima M et al 2007). However, other rapid alleles that have been found apart from NAT2*4 are NAT2*11A, NAT2*12A-C, NAT2*13A, NAT2*18. In addition, within the slow acetylator genotype group there is heterogeneity in phenotype due to variation in enzyme activity conferred by different alleles (Hein DW et al 1995, Cascorbi I et al 1999, Hein DW 2009, Ruiz JD et al 2012), which may affect the ability to identify significant associations (Selinski S et al 2013b).
The frequency of the slow acetylator phenotype differs markedly among ethnic groups (Evans DAP 1984), and this is referred to the differing frequencies of the polymorphisms that correspond to the slow acetylator alleles. In Caucasian and African populations, the frequency of the slow acetylation phenotype varies between 40 and 70%, whereas it ranges from 10 to 30% in Asian populations, such as Japanese, Chinese, Korean, and Thai (Meyer UA and Zanger UM 1997). Caucasian and African populations have high frequencies of \textit{NAT2*5} haplotypes (>28%) and low frequencies of \textit{NAT2*7} haplotypes (<5%), while Asian populations have low incidences of \textit{NAT2*5} haplotypes (<7%) and higher incidences of \textit{NAT2*7} haplotypes (>10%). Moreover, \textit{NAT2*14} haplotypes are almost absent from Caucasian and Asian populations (<1%), but are present in African populations at comparably higher frequencies (>8%).

Several studies examining the variation of \textit{NAT2} haplotypes between different populations and ethnicities support the hypothesis suggesting the \textit{NAT2} slow acetylator phenotype was positively selected for the transition to an agricultural/pastoral lifestyle from a hunter-gatherer/nomadic lifestyle, leading to changes in diet and thus exposure to different xenobiotics (Patin E et al 2006a, Magalon H et al 2008, Sabbagh A et al 2008, Luca F et al 2008).

For example, slow acetylator status is higher among Tajik populations (agriculturists) compared to Kirghiz populations (nomads) in Central Asia (Magalon H et al 2008), and a high frequency of rapid or intermediate status is observed in hunter-gatherer populations in Western/Southern Africa (Kung San, Bakola Pygmy, Biaka Pygmy populations) (Patin E et al 2006b, Mortensen HM et al 2011). In India, the frequency of slow acetylators (based on genotype) was higher than rapid acetylators in areas where a vegetarian diet dominates, and the opposite was found in areas where non-vegetarian diet is more frequent (Khan N et al 2013).
Yuliwulandari et al (2008) detected 23 different variants in the promoter and coding regions of the \( \text{NAT2} \) gene among 212 Indonesian individuals and they found that phenotypes for rapid, intermediate and slow acetylators were 13.6%, 50.8%, and 35.6%, respectively. The frequency of slow acetylators was similar to that recorded for other Southeast Asian populations.

Magalon et al. (2008) genotyped 138 individuals from 6 populations in central Asia, they found that the Tajiks and Kazakhs showed the highest frequency of the slow acetylator haplotype \(*5B\), ranging from 22 to 26%. The \(*6A\) haplotype was exhibit at high frequencies in the Tajik populations (39 to 45%), whereas it was at the lowest frequency in Kazakhs (13%). The Kazakhs exhibited the highest frequency (23%) of \(*7B\), which is mainly restricted to East Eurasian populations. The 'rapid' haplotype \(*4\) was observed at high frequency in the Kirghizs (46 to 48%) and in the Uzbeks (42%), but at nearly 2-fold lower frequencies in the Tajiks (23 to 26%), and intermediate frequency in the Kazakhs (38%). Overall, the Tajiks showed significantly higher proportions of slow acetylators (55% to 63%) as compared to the Uzbeks, Kirghizs, and Kazakhs, where the proportions of slow acetylators ranging from 26 to 35% (Table 4 and 5) (Magalon H et al 2008).

Table 4: Frequency of important SNPs in human \( \text{NAT2} \) gene (Guaoua S et al 2014)

<table>
<thead>
<tr>
<th>Ethnic groups</th>
<th>( \text{NAT2}^*5 )</th>
<th>( \text{NAT2}^*6 )</th>
<th>( \text{NAT2}^*7 )</th>
<th>( \text{NAT2}^*14 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caucasians</td>
<td>0.46</td>
<td>0.285</td>
<td>0.029</td>
<td>0.00</td>
</tr>
<tr>
<td>Germans</td>
<td>0.425</td>
<td>0.278</td>
<td>0.013</td>
<td>0.001</td>
</tr>
<tr>
<td>Americans</td>
<td>0.437</td>
<td>0.266</td>
<td>0.019</td>
<td>0.001</td>
</tr>
<tr>
<td>Southern Korean</td>
<td>0.01</td>
<td>0.224</td>
<td>0.132</td>
<td>0.00</td>
</tr>
<tr>
<td>Spanish</td>
<td>0.47</td>
<td>0.25</td>
<td>0.006</td>
<td>0.004</td>
</tr>
<tr>
<td>Southern Brazil</td>
<td>0.289</td>
<td>0.104</td>
<td>0.021</td>
<td>0.014</td>
</tr>
<tr>
<td>Egyptian</td>
<td>0.497</td>
<td>0.26</td>
<td>0.028</td>
<td>-</td>
</tr>
<tr>
<td>Argentine</td>
<td>0.37</td>
<td>0.256</td>
<td>0.08</td>
<td>0.013</td>
</tr>
<tr>
<td>Omani</td>
<td>0.44</td>
<td>0.27</td>
<td>0.04</td>
<td>0.00</td>
</tr>
<tr>
<td>Senegalese</td>
<td>0.322</td>
<td>0.188</td>
<td>0.00</td>
<td>0.084</td>
</tr>
<tr>
<td>Tunisians</td>
<td>0.315</td>
<td>0.175</td>
<td>0.15</td>
<td>0.05</td>
</tr>
<tr>
<td>South Africa</td>
<td>0.361</td>
<td>0.17</td>
<td>0.067</td>
<td>0.103</td>
</tr>
</tbody>
</table>
Japanese
Indians
Moroccans

0.019 0.23 0.011
0.33 0.38 0.03
0.53 0.25 0.02

Table 5: Frequency of other SNPs in human NAT2 gene (Luca F et al 2008)

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Dendi</th>
<th>Amhara</th>
<th>Oromo</th>
<th>Egyphians</th>
<th>Italian</th>
<th>Greeks</th>
<th>Czechs</th>
<th>Mordvin</th>
<th>Russian</th>
<th>Khanty</th>
<th>Mansi</th>
<th>Yakuts</th>
<th>Chukchi</th>
</tr>
</thead>
<tbody>
<tr>
<td>*4</td>
<td>0</td>
<td>0.06</td>
<td>0.08</td>
<td>0.12</td>
<td>0.24</td>
<td>0.21</td>
<td>0.21</td>
<td>0.34</td>
<td>0.11</td>
<td>0.08</td>
<td>0.17</td>
<td>0.17</td>
<td>0.27</td>
</tr>
<tr>
<td>*5</td>
<td>0</td>
<td>0</td>
<td>0.04</td>
<td>0.01</td>
<td>0</td>
<td>0.01</td>
<td>0.04</td>
<td>0.04</td>
<td>0.07</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.09</td>
</tr>
<tr>
<td>*5B</td>
<td>0.23</td>
<td>0.53</td>
<td>0.33</td>
<td>0.49</td>
<td>0.32</td>
<td>0.38</td>
<td>0.39</td>
<td>0.50</td>
<td>0.33</td>
<td>0.29</td>
<td>0.19</td>
<td>0.19</td>
<td>0.27</td>
</tr>
<tr>
<td>*5C</td>
<td>0.05</td>
<td>0</td>
<td>0.04</td>
<td>0.01</td>
<td>0.03</td>
<td>0.04</td>
<td>0.06</td>
<td>0</td>
<td>0</td>
<td>0.07</td>
<td>0</td>
<td>0</td>
<td>0.09</td>
</tr>
<tr>
<td>*5D</td>
<td>0.05</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.06</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.04</td>
</tr>
<tr>
<td>*5G</td>
<td>0.05</td>
<td>0.03</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.04</td>
</tr>
<tr>
<td>*5J</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.04</td>
</tr>
<tr>
<td>*6</td>
<td>0.27</td>
<td>0.27</td>
<td>0.21</td>
<td>0.27</td>
<td>0.34</td>
<td>0.29</td>
<td>0.26</td>
<td>0.11</td>
<td>0.08</td>
<td>0.11</td>
<td>0.25</td>
<td>0.27</td>
<td>0.09</td>
</tr>
<tr>
<td>*6B</td>
<td>0.09</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.05</td>
</tr>
<tr>
<td>*6C</td>
<td>0.05</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.05</td>
</tr>
<tr>
<td>*7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.07</td>
</tr>
<tr>
<td>*7B</td>
<td>0</td>
<td>0.03</td>
<td>0.17</td>
<td>0.01</td>
<td>0.03</td>
<td>0.03</td>
<td>0</td>
<td>0.07</td>
<td>0.04</td>
<td>0.18</td>
<td>0.13</td>
<td>0.08</td>
<td>0.09</td>
</tr>
<tr>
<td>*11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.11</td>
<td>0</td>
<td>0</td>
<td>0.11</td>
</tr>
<tr>
<td>*12</td>
<td>0.05</td>
<td>0.03</td>
<td>0.08</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.08</td>
<td>0.04</td>
<td>0</td>
<td>0.04</td>
<td>0</td>
<td>0.05</td>
</tr>
<tr>
<td>*12B</td>
<td>0.05</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
<td>0</td>
<td>0.04</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.05</td>
</tr>
<tr>
<td>*12C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0.04</td>
<td>0.04</td>
<td>0.03</td>
<td>0</td>
<td>0</td>
<td>0.04</td>
</tr>
<tr>
<td>*13</td>
<td>0.14</td>
<td>0.03</td>
<td>0.04</td>
<td>0.07</td>
<td>0.01</td>
<td>0.01</td>
<td>0</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0</td>
<td>0</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Important Variants

NAT2 alleles containing the 191G>A (rs 1801279), 341T>C (rs1801280), 590G>A (rs 1799930), and/or 857G>A (rs 1799931) missense substitutions are associated with slow acetylator phenotype(s). Striking ethnic differences in the frequencies of these missense substitutions (Grant DM et al 1997) play an important role in ethnic differences in frequency of slow acetylator alleles (Cascorbi I et al 1999) and phenotype(s) (Weber WW and Hein DW 1985). For example, the 191G>A substitution commonly represents the NAT2*14 gene cluster that is available in greater extent in African Americans and native Africans, but it is virtually absent in Caucasian populations.

Cascorbi et al. (1995) found 7 different alleles of the NAT2 gene coding for the 'slow' phenotype. They observed a slow acetylation genotype in 58.9% of the 844 unrelated German subjects studied (Cascorbi et al 1996). In vivo acetylation capacity of homozygous wildtype subjects was significantly higher than in heterozygous genotypes. All mutant alleles showed low in vivo acetylation capacities, and the term 'slow' genotypes differed significantly among each other, as reflected in lower acetylation capacity of some alleles as compared with others.

**Other Important Variants**

Other NAT2 alleles that have been reported to be associated with slow acetylator phenotype are those containing the A-434C (rs 72554616) and C-190T (rs 1805158) (Dai Z et al 2008). However, till now there are no reports which detect SNP rs72554616 influence on drug metabolism or any disease so it may not have any clinical impact. On the other hand, only one report of isoniazid side effect has been detected with SNP rs1805158 (Fuselli S et al 2007).

**Other Variants**

Other variants include G-499A (rs 72554617) which represent haplotype *10, shows slow catalytic activity but substrate dependent. Variant 803A>G (rs 1208) presented in allele *12, variant 282C>T (rs 1041983) presented in allele *13 and variant A-845C (rs 56054745) presented in allele *18 have shown to have rapid catalytic activities that is also included in the metabolism of certain drugs like 1,7-dimethylxanthine, antivirals for treatment of HIV infections, drugs for treatment of

### 1.3.2.1.5 NAT2 Important Haplotypes

NAT2*4 is considered the wild-type allele because of the absence of any of these substitutions and based on many studies in many ethnic groups, NAT2*4 is not the most common allele (Hein DW et al 1994, Hein DW et al 1995, O’Neil WM et al 2000).

NAT2* 5, NAT2* 6, NAT2* 14, and NAT2* 7 alleles have low catalytic activity associated with the slow acetylator phenotype, whereas NAT2*12 and NAT2*13 alleles associated with high catalytic activities that show rapid acetylator phenotype similar to NAT2*4 (Hein DW 1994, Hein DW et al 1995, O’Neil WM et al 2000).

### 1.3.2.1.6 NAT2 Alleles Nomenclature/Variant Annotations

Since 1995 many alleles and their nomenclature have been identified and published (Vatsis KP et al 1995). The Human Gene Nomenclature Committee has agreed that the symbol NAT be assigned to the arylamine N-acetyltransferase genes. NAT2 alleles *1, *2 and *3, have been found but these are non-human NAT2 alleles. Based on this, human NAT2 alleles start with allele *4 (wild type) (table 6) (Hein DW et al 2008).

<table>
<thead>
<tr>
<th>NAT2 Allele (Haplotype)</th>
<th>Nucleotide Change(s) and rs #</th>
<th>Amino Acid Change(s)</th>
<th>Phenotyped</th>
</tr>
</thead>
<tbody>
<tr>
<td>*4</td>
<td>Reference</td>
<td>Reference</td>
<td>Rapid</td>
</tr>
<tr>
<td>*5A</td>
<td>341T&gt;C (rs1801280)</td>
<td>I114T</td>
<td>Slow</td>
</tr>
<tr>
<td></td>
<td>481C&gt;T (rs1799929)</td>
<td>L161L (synonymous)</td>
<td></td>
</tr>
<tr>
<td>*5B</td>
<td>341T&gt;C (rs1801280)</td>
<td>I114T</td>
<td>Slow</td>
</tr>
<tr>
<td></td>
<td>481C&gt;T (rs1799929)</td>
<td>L161L (synonymous)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>803A&gt;G (rs1208)</td>
<td>K268R</td>
<td></td>
</tr>
<tr>
<td>*5C</td>
<td>341T&gt;C (rs1801280)</td>
<td>I114T</td>
<td>Slow</td>
</tr>
</tbody>
</table>

Table 6: NAT2 alleles and its references SNPs and catalytic activity
1.3.2.1.7 NAT2 Genotype/Phenotype Relationship

The caffeine metabolite Paraxanthine can undergo acetylation by NAT2 to form 5-acetylamino-6-formylamino-3-methyluracil (AFMU) (Thorn CF et al 2012). Caffeine can be utilized as a non-toxic probe drug in vivo for examining the acetylation phenotype status; by determining metabolite ratio A FMU/1-methyl xanthine (1MX) in urine after caffeine consumption, a bi- or tri-modal pattern in a given population is determined (Grant DM et al 1984, Grant DM et al 1990, Kaufmann GR et al 1996). AFMU/AFMU+1X+1-methyluric acid (1U), AFMU+5-acetylamino-6-amino-3-methyluracil (AAMU)/AFMU+AAMU+1MX+1MU or AAMU / AAMU+1X+1U metabolite ratios can also be used to measure acetylation phenotype status (Jetter A et al 2004, Begas E et al 2007, Djordjevic N et al 2011, Djordjevic N et al 2012).

Therefore, most commonly used probe drug for NAT2 phenotype measurement is caffeine (Grant DM et al 1984, Lorenzo B and Reidenberg MM
1989, Leyland-Jones B et al 1999) and excellent NAT2 genotype/phenotype correlations have been recorded (Cascorbi I et al 1995, Grant DM et al 1997). Although administration of caffeine is a relatively noninvasive method for measuring NAT2 acetylator phenotype, the method used involves determination of the quantity of other metabolites. Thus, genetic and environmental factors may influence the levels of the metabolites used in the phenotype determination. For example, NAT1 phenotype affects the urinary caffeine ratio utilized for determination of NAT2 phenotype (Cribb AE et al 1996).

No association has been reported between NAT2 acetylation activity and sex, race, age, education, smoking, physical activity, weight, vitamin E supplements, alcohol and dietary intakes of red meat, processed meat or cruciferous vegetables or use of menopausal estrogens (Le Marchand I et al 1996).

In one study, up to 54% of the differences in acetylation activity measured by caffeine test could be demonstrated by NAT2 genotype (homozygous wildtype, homozygous variant or heterozygous determined by PCR-RFLP), although it was found that use of phenotyping had some advantages in sensitivity over genotyping (Le Marchand L et al 1996).

The original method for determining the NAT2 genotype used Southern blot analysis of genomic DNA from leukocytes but soon afterwards, a PCR-based RFLP method was described (Mashimo M et al 1992, Abe M et al 1993). In addition, use of the urinary molar ratio of AFMU/1-methylxanthin to determine NAT2 phenotype it was described and validated at around the same time (Butler MA et al 1992).
1.3.2.1.8 Clinical Impact, Drug Metabolism and Substrate Specificity of NAT2 Polymorphism

The association between acetylator status and the risk of various diseases has been extensively recorded, and reviewed in detail (Evans DAP 1984, Clark DW 1985, Evans DAP 1993).

Altered risk with either the slow or rapid phenotype has been reported for several diseases including bladder, colon and breast cancer, systemic lupus erythematosi, diabetes, Gilbert's disease, Parkinson's disease (PD) and Alzheimer's disease. Humans are exposed to many toxic NAT substrates including the food-derived heterocyclics present in the diet as well as arylamines such as 4-aminobiphenyl and β-naphthylamine present in tobacco smoke (Nagao M et al 1996, Smith CJ et al 1997, Besarati NA et al 2000, Sram RJ and Binkova B 2000). Their influences cause several adverse effects with a lot of drugs like dipyrone, drugs for treatment of tuberculosis, hydralazine, isoniazid, pyrazinamide, sulfamethoxazole, sulfapyridine, sulfasalazine and caffeine (Weber WW and Hein DW 1985, Stanley LA and Sim E 2008, Sim E et al 2012, McDonagh EM et al 2014).

In cancer, NAT2 phenotype/genotype may be a risk factor. NAT2 has been implicated in chemical carcinogenesis pathways. Unlike the relatively rare but highly penetrant genes included in familial cancers, those genes responsible for metabolic polymorphisms have low penetrance and cause only a moderate increase in cancer risk. Nevertheless, their widespread occurrence in the general population suggests that they are a significant contributor to individual risk (Weber WW and Hein DW 1985, Sim E et al 2008, Stanley LA and Sim E 2008, Sim E et al 2012, McDonagh EM et al 2014).

For that, polymorphisms in the NAT2 genes have been examined for an association with cancer risk, however the findings are inconsistent; this may be due

In 1986, Lang et al first illustrated an association between the slow acetylator phenotype and bladder cancer while the rapid acetylator phenotype was suggested to have an increased risk factor for colon cancer (Lang et al 1986, Ilett et al 1987, Wohlleb JC et al 1990). Other studies have been unable to confirm these findings (Welfare MR et al 1997, Chen J et al 1998, Brockton N et al 2000).

Cartwright et al (1982) divided the ratio that used to measure the phenotype status into 4 groups: 0.3 and greater correspond to rapid acetylators, and the other three (0.01–0.09; 0.1–0.19; and 0.2–0.29) correspond to different levels of slow acetylator phenotype. Urinary bladder cancer risk is markedly elevated in NAT2 slow acetylator phenotype, particularly with the slowest NAT2 phenotype (Cartwright RA et al 1982). This same conclusion has been reported by different studies including some recent large genome-wide association studies (Brockmoller J et al 1996, Okkels H et al 1997, Filiadis IF et al 1999, Hamdy SI et al 2003, Lubin JH et al 2007, Garcia-Closas M et al 2011, Figueroa JD et al 2014).

Another study showed an enhanced effect for smoking intensity and bladder cancer in NAT2 slow acetylators which increases with intensity (Lubin JH et al 2007).

Several studies have been reported that rapid N-acetylation phenotype was associated with an increased risk of colorectal cancer, particularly, in patients with right-sided cancer. However, the frequency of the *7A haplotype associated with the slow acetylator status, was increased among colorectal cancer patients on the left side and along the sigmoid/rectal region (Lang NP et al 1986, Ilett KF et al 1987, Lee EJD et al 1998).
Women with very slow acetylator genotype (homozygous for NAT2*5) who smoked for 20 years showed an increased breast cancer risk when they were compared to individuals with rapid NAT2 genotype (Vander HL et al 2003).


In yet other studies, the association between NAT2 acetylator genotype and breast cancer has been examined in relation to diet. It was observed that red meat consumption and NAT2 genotype were not associated with breast cancer risk (Ambrosone CB et al 1998, Gertig DM et al 1999). However, in another study, rapid / intermediate NAT2 genotypes were associated with breast cancer risk in women who consistently consume very well-done meat (Deitz AC et al 1999).


Consistent with that, rapid NAT2 acetylators who consumed pan-fried meats had higher levels of urinary mutagenicity than slow acetylators (Marini OM et al 1997). Whereas it is only limited to homozygous rapid (*4/*4) acetylators (Gil JP and Lechner MC 1998), this finding was also observed for lung (Cascorbi I et al 1996) and laryngeal (Henning S et al 1999) cancers. Bell DA et al 1995 recorded an association between the NAT1*10 allele and colorectal cancer, and the risk were highest among NAT2 rapid acetylators. Another study also found a higher risk for colorectal cancer in individuals who consumed well-done meat and possessed both the NAT1*10 allele and rapid acetylator NAT2 genotype (Chen J et al 1998).

In another study, the homozygous slow acetylator NAT2 genotype was correlated with an increased risk of oral/pharyngeal cancer, but not laryngeal cancer (Jourenkova-Mironova N et al 1999). However, the homozygous rapid acetylator (*4/*4) genotype was strongly correlated with laryngeal cancer in a German study (Smelt VA et al 1998).
Several studies conducted to examine the associations between acetylator genotypes and prostate cancer found no relationship between NAT2 genotype and prostate cancer (Agundez JAG et al 1998, Wadelius M et al 1999). Additionally, aromatic amine N-acetyltransferase activity levels in human prostates were independent of NAT2 genotype (Agundez JAG et al 1998).

Head and neck cancers are strongly correlated with smoking, and several studies have investigated the role of NAT1 and NAT2 polymorphisms and the risk of head and neck cancer in smokers. The slow NAT2 acetylator phenotype was observed to be associated with the development of head and neck cancer in Caucasians (Drózdz M et al 1987, Gonzalez MV et al 1998) and with the development of oesophageal cancer in Japanese (Morita S et al 1998).

Another study showed that NAT2 slow acetylator genotype can be a small, low penetrance risk factor for head and neck cancer (Zhang L et al 2014). Seeking further studies is important to have clear picture about the associations between NAT2 variants and susceptibility to other cancers (Boukouvala S and Fakis G 2005, Sim E et al 2008, Butcher NJ and Minchin RF 2012).

Chan et al. (2003) found that the frequency of the slow acetylator genotype for NAT2 was significantly higher in Parkinson's disease (PD) patients than in a control group after adjusting age, sex, and smoking history in Hong Kong Chinese. A significant associations with PD for NAT2 was also reported in a meta analysis (Tan EK et al 2000). Similarly, Bandmann O et al (2000) stated that "the slow phenotype appears to increase risk of PD".

Increased risk of the late-onset Alzheimer's disease has been observed with the rapid NAT2 phenotype in non-apoE epsilon 4 carriers (Ogawa M 1999).
NAT2 polymorphisms and acetylator phenotype may affect risk of other complex multifactorial diseases (including asthma and diabetes), however, the results are conflicting and further studies are needed (Ladero JM 2008, Batra J and Ghosh B 2008).

The development of molecular epidemiological studies exploring the relationship between the acetylation polymorphisms and cancer are still incomplete but there has been considerable progress recently due to the availability of genome-wide association studies (Chung CC et al 2010). The inconsistency in the human epidemiological studies may be related to differences in carcinogen exposures, genotype and/or phenotype methods, insufficient sample sizes, and/or other susceptibility genes and factors.

The drug adverse effects related to NAT2 phenotype mostly occurred as the result of a shift in the metabolic pathways responsible for the activation and detoxification of the drug. This is best demonstrated by the amine-containing sulphonamides, such as sulphamethoxazole, that undergoes hydroxylation to a reactive N-hydroxy metabolite capable of covalently binding to macromolecules and giving rise to idiosyncratic adverse reactions (Figure 10) (Spielberg SP 1996).

![Figure 10: The metabolic pathways responsible for the activation and detoxification of the drug](image-url)
These drugs can also be acetylated by NAT2 to non-reactive N-acetyl metabolites. In slow acetylators, a higher proportion of the drug is N-hydroxylated and based on that individuals with slow acetylation status are at a greater risk of sulphonamide-induced toxicity (Das KM and Dubain R 1976, Shear NH et al 1986, Cribb AE et al 1996).

The major focus of NAT2 pharmacogenetic studies are on the association of anti-tuberculosis (anti-TB) drugs and drug-induced hepatotoxicity, liver injury (DILI), or hepatitis.

Sulfamethoxazole is acetylated to N-acetyl sulfamethoxazole, or oxidized to sulfamethoxazole hydroxylamine (a reactive metabolite that is responsible in toxicity) by CYP450 enzymes (Davis CM and Shearer WT 2008). The association between NAT2 genotypes and sulfamethoxazole pharmacokinetics (PK) have also been supported in renal transplant patients treated with an immunosuppressive regimen, significantly higher sulfamethoxazole concentrations in slow acetylators (defined as homozygotes or compound heterozygotes for *5, *6, or *7 variants) are seen compared to rapid acetylators (homozygous *4/*4) (Kagaya H et al 2012).

NAT2 is important pathway in the metabolism of isoniazid (INH), mediating its biotransformation to the metabolite acetyl-INH, which is hydrolyzed to isonicotinic acid or acetyl-hydrazine and subsequently to the non-toxic diacetylhydrazine, or hydrolyzed to hydrazine (Preziosi P 2007, Tostmann A et al 2008, Metushi IG et al 2011, Mahapatra S et al 2012, Daly AK and Day CP 2012). Liver toxicity of INH treatment derives from INH itself (a hydrazine derivative) and its metabolites, including acetyl-hydrazine, hydrazine and ammonia, and is thought to include the formation of reactive oxygen species that can cause necrosis and autoimmunity (Mitchell JR et al 1975, Preziosi P 2007, Tostmann A et al 2008, Fukino K et al
2008, Metushi IG et al 2011) and may also involve epigenetic effects (Murata K et al 2007).

Consequently, NAT2 slow acetylators have been associated with an increased risk of hepatotoxicity / liver injury / hepatitis induced by anti-TB drug treatment when compared to rapid acetylators (Bose PD et al 2011, Chamorro JG et al 2013 and Gupta VH et al 2013).

NAT2 variants 590G>A and 857G>A were independently observed at a significantly higher frequency in children with co-trimoxazole induced adverse drug reactions (ADRs) compared to those without (Zielińska E et al 1998).

In another study on Japanese patients who were treated with co-trimoxazole, it was found to have higher risk of adverse events with slow acetylator status (determined in this study by NAT2 genotypes *6A/*6A, *6A/*7B, *7B/*7B) compared to rapid acetylators (genotypes *4/*4, *4/*5B, *4/*5E, *4/*6A, *4/*7B) (Soejima M et al 2007).

The increased risk of developing side effects such as neurotoxicity or haemolytic anaemia, to dapsone therapy is very similar to that elucidated for the sulphonamides (May DG et al 1990). The most severe incidence of toxicity found in individuals with a slow acetylator phenotype who are rapid hydroxylators (Bluhm et al 1999).

While slow acetylators are at a greater risk of toxicity from sulphonamides and dapsone, amonafide which is drug used for treatment of various cancers it is an arylamine that show elevated incidence of adverse reactions in rapid acetylators after being given standard dose compared to the slow acetylators. The drug undergoes N-acetylation to an active metabolite that contributes to systemic toxicity (myelosuppression). Based on that, different doses for the two groups were
recommended. However, the drug displayed variable and unpredictable toxic effects
failed to reach phase III clinical trial primary end points (Ratain MJ et al 1991, 1993,

One study found a reduction in the antihypertensive activity of hydralazine in
the rapid acetylators which required a 40% higher dose to show similar therapeutic
effect compared with slow acetylators. This variation appeared to be due to
alteration in the bioavailability of the drug which reduced from 33% in slow
acetylators to less than 10% in rapid acetylators (Shepherd AM et al 1980).

Studies revealed that rapid acetylators show lower hydralazine plasma
concentrations and area under the concentration-time curve compared to slow
acetylators (Israel ZH and Dayton PG 1977, Weber WW and Hein DW

Sulfasalazine consists of sulfapyridine that metabolizes to N-acetyl-
sulfapyridine which is significantly diminished in slow acetylators (carriers of two
variant alleles \textit{NAT2*5B, NAT2*6A, NAT2*7B or NAT2*5, NAT2*6 and NAT2*7})
compared to both intermediate (one variant and one \textit{NAT2*4} haplotype) and rapid
acetylators (\textit{NAT2*4/*4}) (Yamasaki Y et al 2008, Ma JJ et al 2009). Slow acetylators
have higher concentrations and elimination half-life of sulfapyridine (based on
genotyping \textit{NAT2} SNPs rs1041983, rs1801280, rs1799929, rs1799930 and
rs1799931) (Kuhn UD et al 2010).

A supported prospective study of female rheumatoid arthritis (RA) patients
treated with sulfasalazine detected 4 patients who had reported adverse effects to
sulfasalazine in a one year period - none had the \textit{NAT2*4} haplotype, each carrying
two variant alleles in Japan (Soejima M et al 2008).
In summary, the examples discussed above have highlighted the essentials of NAT2 as a genetic influence on drug pharmacokinetics and pharmacodynamics (Table 7) (Ratain MJ et al 1996, Innocenti F et al 2001).

Table 7: Effect of acetylator status on drug response and toxicity (Butcher NJ et al 2002)

<table>
<thead>
<tr>
<th>Drug</th>
<th>Phenotype</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dapsone</td>
<td>Slow</td>
<td>Neurotoxicity</td>
</tr>
<tr>
<td>Sulphamethoxazole</td>
<td>Slow</td>
<td>Hypersensitivity</td>
</tr>
<tr>
<td>Hydralazine</td>
<td>Slow</td>
<td>Systemic lupus erythematosus</td>
</tr>
<tr>
<td>Isoniazid</td>
<td>Rapid</td>
<td>Decreased therapeutic effect</td>
</tr>
<tr>
<td>Cotrimoxazole</td>
<td>Slow</td>
<td>Interaction with phenytoin</td>
</tr>
<tr>
<td>Sulphasalazine</td>
<td>Slow</td>
<td>Interaction with rifampicin</td>
</tr>
<tr>
<td>Amonafide</td>
<td>Rapid</td>
<td>Various adverse reactions</td>
</tr>
<tr>
<td>Procainamide</td>
<td>Slow</td>
<td>Various toxicities</td>
</tr>
<tr>
<td>Phenelzine</td>
<td>Rapid</td>
<td>Hepatotoxicity, nausea/vomiting</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Leukopenia</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Systemic lupus erythematosus</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Decrease therapeutic effect</td>
</tr>
</tbody>
</table>

Some studies demonstrated that bacterial NAT from the gut flora has a role in the activation and detoxification of xenobiotics in the host organism and may play an essential role in the metabolism of anti-inflammatory drugs, such as 5-aminosalicylic acid (Trepanier LA et al 1997, Delomenie C 2001).

**Substrate Specificity of Human NAT2**

There is no specific substrate structure that is uniform for the different isoforms of NAT, although, p-aminobenzoic acid (PABA), p-aminobenzoyl glutamate and p-aminosalicylic acid (PAS) are identified as specific substrates for human NAT1. These substrates can be featured by the availability of relatively small hydrophilic substitutions in the para position of the aromatic ring. On the other hand, sulfamethazine, procainamide and dapsone are acetylated primarily by human NAT2. Other compounds such as 2-aminofluorene are excellent substrates for both human NAT1 and NAT2 (Minchin RF et al 1992, Delomenie C et al 2001).
Endogenous substrates for human NAT2 are not well known, although, in rodents N-acetyltransferases metabolize a number of aromatic and heterocyclic amine carcinogens that form tumors (Layton DW et al 1995). Aromatic amines and hydrazines (N-acetylation) and N-hydroxy-aromatic and heterocyclic amines (O-acetylation) are both examples of acceptor substrates that are deactivated (N-acetylation) or activated (O-acetylation) by NAT2 (Hein DW 1988).

1.4 Potential Contribution and Limitation of the Study

Time constraints have been the main challenge that we have faced during this study. Molecular and epidemiological research involving human subjects requires obtaining the appropriate ethical approvals which involved numerous communications with managers and administrators in several institutions. This was an absolute pre-request before sample collection.

Moreover, our study required volunteers who had to drink caffeinated drinks and come back after 2 hours for sample collection which was inconvenient for some individuals who refused to participate. Additionally, not all the places were considered appropriate for sample collection which limited the number of centers that we could use for subjects recruitment.

Interestingly, the level of consanguinity in the UAE varies with more than 50% of marriage are consanguineous. This clearly may influences the genotypes frequencies and may result in their deviation from Hardy-Weinberg.
Chapter 2: Aims and Objectives

2.1 Aims

This study has been conducted to determine CYP1A2 and NAT2 genotypes and their frequencies and correlations with the corresponding phenotypes in Emiratis.

2.2 Specific Objectives

✓ Literature evaluation of CYP1A2 and NAT2 polymorphism and its association with diseases, drug response and toxicity.

✓ Recruitment of a sample of volunteers that is genetically representative of Emiratis and willing to go through phenotyping and genotyping analysis.

✓ Measurement of CYP1A2 phenotyping status of the recruited samples.

✓ Determining the polymorphic CYP1A2 alleles and genotype frequencies of the recruited samples and examine if they follow or deviate from the Hardy-Weinberg equilibrium.

✓ Defining the correlation between CYP1A2 genotyping and phenotyping.

✓ Measuring the NAT2 phenotyping status of UAE population samples.

✓ Determining the NAT2 genotype frequencies of UAE population, and their deviations from Hardy-Weinberg equilibrium.

✓ Defining the correlation between NAT2 genotyping and phenotyping.

✓ Discovering new mutations, identifying and classifying them if present.
Chapter 3: Methods

This study was approved by the Research Ethics Committee of the Faculty of Medicine and Health Sciences of UAE University (Al-Ain Medical District Human Research Ethical Committee, AAMD/HREC No: 21M059). No deviation from the protocol was implemented without the prior review and approval.

3.1 Research Design

3.1.1 Type of Study
Randomized, open labeled and multi-center study of the CYP1A2 and NAT2 polymorphism and their phenotyping and genotyping correlation in Emirati population, from the period of September 2012 till September 2016.

3.1.2 Setting
All the 7 emirates of UAE were involved; subjects were UAE National volunteers.

3.1.3 Including Criteria
Healthy subjects of more than 12 years of age.

3.1.4 Exclusion Criteria
Subjects who were smokers, not Emirati or aged less than 12 years were excluded.

3.1.5 Place of Study
The study was conducted in the United Arab Emirates.
3.1.6 Study Duration / Period of Study

The period of this study started from September 2012 and ended in September 2016 during the PhD study period.

3.1.7 Ethics and Ethical Approvals

Ethical approval for the research was taken by the Research Ethics Committee of the Faculty of Medicine and Health Sciences of UAE University (Appendix 1 and 2: Al-Ain Medical District Human Research Ethical Committee, AAMD/HREC No: 21M059).

Communication letters were prepared and submitted for all the places that we were planning to visit for samples collection. Approval letters were collected from those places that had been visited (Appendix 3: UAEU Approval, Appendix 4: Abu Dhabi Education council (ADEC) Approval, Appendix 5: Dubai Education zone Approval and Appendix 6: Ras Al Khaimah Education Zone Approval). The volunteers signed a written consent form that their participation is voluntary (Appendix 7: Consent Form).

3.1.8 Sampling

Individuals with United Arab Emirates nationality had been chosen who were above 12 years old, non-smokers and had been assigned randomly in a systemic way, from multicenters including all of the Emirates of UAE.

The assigned number of individuals who participated from each Emirate was defined based on UAE National Bureau of statistics of UAE (Emiratis) population (Table 8).
Sample Size Calculation

The sample size calculation was carried out to ensure the precision of confidence interval to be 5%. Although, the frequencies of alleles were different from region to region we used sample size calculation assuming the maximum variability in data. By doing so, the sample size required for this prevalence study was around 500 subjects. The volunteers signed a written consent form that their participation was voluntary. To ensure that this number was achieved we recruited 581 subjects (16% more subjects) to make up for incomplete data and withdrawals.

Sample Collection

Based on UAE population (Emiratis) distribution between the Emirates, the expected distribution of our samples was shown in following table (Table 8):

Table 8: The expected distribution of our samples based on the distribution of UAE population (Emiratis) between the Emirates

<table>
<thead>
<tr>
<th>UAE emirates</th>
<th>Population</th>
<th>Samples expected to be collect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abu Dhabi (AD)</td>
<td>400000 = 42%</td>
<td>210 samples</td>
</tr>
<tr>
<td>Dubai</td>
<td>171000 = 18%</td>
<td>90 samples</td>
</tr>
<tr>
<td>Sharjah</td>
<td>152000 = 16%</td>
<td>80 samples</td>
</tr>
<tr>
<td>Ras AlKhaimah</td>
<td>95000 = 10%</td>
<td>50 samples</td>
</tr>
<tr>
<td>Fujairah</td>
<td>66500 = 7%</td>
<td>35 samples</td>
</tr>
<tr>
<td>Ajman</td>
<td>47500 = 5%</td>
<td>25 samples</td>
</tr>
<tr>
<td>Um Al Quwain</td>
<td>19000 = 2%</td>
<td>10 samples</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>950000 = 100%</strong></td>
<td><strong>500 samples</strong></td>
</tr>
</tbody>
</table>

3.2 Data Collection

Data collection form was filled by each volunteer who agreed to participate and signed the consent form only (Appendix 8: Data Collection Form). This was followed by samples collection and treated as anonymous since no names were disseminated. Finally, codes were assigned for each volunteer that had been used in tracing the data in data collection and analysis.
3.2.1 Sample Collection

Five hundred and eighty one subjects were enrolled in this study. The subjects were asked simply to consume 300ml of a caffeinated cola drink or a cup of strong tea or coffee and then to provide a spot urine sample 2 hours later. No restrictions whatsoever regarding food intake or caffeine consumption are required of our subjects prior to their taking part in the study. However, volunteers were asked not to consume any of these beverages during the two hours period.

In addition, each volunteer had to provide buccal swab at the time of urine sample collection using Isohelix SK1 buccal swabs with Isohelix dri-capsules for sample collection (Figure 11). This buccal swab had a unique swab matrix that greatly improves DNA yields, it is an alternative to blood collection method, suitable for human, easy to handle and quick to use.

![Figure 11: Instruction of Isohelix SK1 buccal swabs use](image)

3.2.2 Storage and Arrangement for Urine Samples and Buccal Swabs

Urine samples were placed on ice immediately after voiding and the pH of each sample was adjusted to <3.5 within 1 to 2 hours of collection to prevent loss of
5-acetylamino-6-formylamino-3-methyluracil (AFMU), a key metabolite of caffeine and other caffeine metabolites. After adjustment of pH, aliquots of urine were stored at -20°C until analysis. Sample analysis was done as described by Grant et al 1984 and Woolhouse et al 1997 as detailed in the following sections.

The collected buccal swab which consisted of Isohelix SK1 buccal swabs and Isohelix dry-capsule for each subject was placed immediately on ice after collection and then was stored at -20°C until analysis.

3.2.3 Determination of Phenotypes

3.2.3.1 The Extraction Protocol of Caffeine Metabolites from Urine

The urine samples were thawed to room temperature and vortexed thoroughly for 1 minute, ultrasonicated for 15 minutes and vortexed again for 5 seconds. Then, 0.2ml of urine and 120mg ammonium sulfate were added into a labeled clean 15ml centrifuge tube and vortexed for 30 seconds. This was followed by the addition of 6ml of chloroform: isopropanol (95:5) and vortexed for 1 minute.

The solution was centrifuged at 3000g for 5 minutes; the supernatant aqueous phase was carefully removed and dried in N2 flow at 42 °C. Finally, when the dryness of the solution was assured 0.5 ml of acetic acid (0.05%) was added to dissolve the contents.

3.2.3.2 Using HPLC Method

The mixture was filtered through 0.22u filter, 100 ul of the final product was injected into the column of HPLC. Each sample took one hour to complete the cycle on the HPLC.

High performance liquid chromatography (HPLC) was performed by separating the metabolites using a reversed-phase 4.6x250 mm column (ultra-
sphere ODS-35uM, Hichrom, UK) and eluted with mobile phases A (0.045% acetic acid) and B (100% methanol) using a gradient profile starting 1ml/min, 18% B at 0 minute and increasing to 50% B at 27 minute; increasing to 90% B at 32 minute; 90% B at 37 minute; reduced to 18% B at 42 minute till the end of 50 minute for re-equilibration.

The standards included AFMU, 1MU, 1MX, 17MU, 17MX and 137MX. Spectral conformation of caffeine metabolite peaks was performed routinely, using a programmable multiple wavelength detector (Waters 2998PDA, Waters Corp., Milford MA). Using this methodology, recovery of caffeine, paraxanthine, AFMU and 1-methylxanthine ranged from 91 to 103%.

Subsequently, the levels of caffeine metabolites including caffeine, AFMU, 17MX, 17MU, 1MU and 1MX were determined by HPLC.

Based on Butler et al (1992), caffeine 3-demethylation activity (CYP1A2) was assessed by determining the ratio (17MX + 17MU)/137MX. This molar ratio correlates well with the rate constant for hepatic 3-demethylation of caffeine based on a pharmacokinetic study (r= 0.73, P value = 0.007) compared to other urinary metabolism ratios 17MX/137MX, (AFMU+1MX+1MU)/17MU and (AAMU+1MX+1MU)/17MU (Butler MA et al 1992). Since other studies had used this molar ratio, this measure was calculated and compared in this study, the frequency distribution of (17MX + 17MU)/137MX in our data indicated a trimodal distribution with specific cut points that represented slow, intermediate and rapid acetylators.

The ratio of (17MX+ 17MU)/137MX was calculated to determine the phenotype status of CYP1A2 activity in each subject as described by Tang B et al 1994 and Joshua EM et al 2008; the following table was used to identify the CYP1A2 status (Table 9).
Table 9: Classification of CYP1A2 enzyme activity based on the ratio of \((17\text{MX} + 17\text{MU})/13\text{MX}\)

<table>
<thead>
<tr>
<th>((17\text{MX} + 17\text{MU})/13\text{MX}) ratio</th>
<th>Enzyme activity status</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7 - 2.5</td>
<td>Slow CYP1A2 enzyme activity</td>
</tr>
<tr>
<td>2.51 - 6.5</td>
<td>Intermediate CYP1A2 enzyme activity</td>
</tr>
<tr>
<td>6.51 - above</td>
<td>Rapid CYP1A2 enzyme activity</td>
</tr>
</tbody>
</table>

Caffeine 3-demethylation activity (NAT2) was assessed by determining the ratio of AFMU/1MX. This molar ratio correlates well with the rate constant for hepatic 3-demethylation of caffeine based on a pharmacokinetic study (Butler MA et al 1992). This ratio typically falls into three phenotypes (slow, intermediate or rapid) (Butler MA et al 1992). The ratio of AFMU/1MX was calculated to determine the phenotype status of NAT2 activity in each subject as described by Grant DM et al 1984 and Woolhouse NM et al 1997. Table 10 was used to identify the acetylator status by calculation of AFMU/1MX ratio.

Table 10: The acetylator status of NAT2 activity was classified based on the ratio of AFMU/1MX

<table>
<thead>
<tr>
<th>AFMU/1MX ratio</th>
<th>Acetylator status</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 0.5</td>
<td>Slow NAT2 acetylator</td>
</tr>
<tr>
<td>0.5 - 2</td>
<td>Intermediate NAT2 acetylator</td>
</tr>
<tr>
<td>2 - above</td>
<td>Rapid NAT2 acetylator</td>
</tr>
</tbody>
</table>

3.2.4 Determination of Genotypes

3.2.4.1 The Extraction Protocol of DNA from the Isohelix SK1 Buccal Swabs

DNA isolation protocol (ref: http://www.isohelix.com/documentation) was carried out using Isohelix Buccal DNA Isolation Kit. Isohelix Buccal DNA Isolation Kits were specifically formulated to produce high DNA yield and purity from buccal swabs. The kits were fully optimised at cell projects LTD company (UK) for use on buccal cell samples and offer reduced handling times, increased DNA yields and
many other important technical benefits for their use in manual, 96-well or other high throughput formats.

**Part 1: DNA Stabilization**

Five hundred microliter Lysis buffer (LS solution) and 20μl Proteinase K (PK solution) from kit were added to the tube containing the buccal swab and vortexed briefly. At this point the DNA was stabilized to proceed for DNA isolation or store the stabilized swab in a sealed tube at room temperature which can be stored for at least 3 ½ years.

**Part 2: DNA Isolation**

After DNA stabilization, the tube containing the swab, LS solution and PK solution was placed in a 60 °C water bath for 1 hour and vortex briefly. Then, the liquid in the tube (approximately 400μl) was transferred into a 1.5ml centrifuge tube using a sterile pipette tip. Then, 400μl of Capture buffer (CT solution) was added to the tube and vortexed briefly and the tube was then placed in a micro-centrifuge (with hinge positioned outwards so the liquid can be removed from the opposite side) and spun at maximum speed (13.4K rpm/12,000 x g) for 7 minutes to pellet the DNA. Subsequently, all the supernatant were carefully removed with a pipette tip taking care not to disturb the DNA pellet. The tube was re-spun briefly and any remaining liquid removed as it was important to do this. Then, a small volume of re-hydration buffer (TE solution) "~70μl" was added to the tube to dissolve the DNA. The solution was left for at least 10 minutes at room temperature and vortexed briefly for the DNA to re-hydrate. Finally, the tube was re-spun for 15 minutes at maximum speed (13.4K rpm/12,000 x g) to remove un-dissolved debris. The supernatant was transferred to a sterile 1.5ml tube, being careful not to disturb the pellet.
The DNA sample at this stage was ready for use in amplification. The storage of the DNA sample was at 4 °C for short term storage or at -20 °C for long term storage. The expected yield from a buccal swab was on average 1 to 10μg DNA (5 to 70ng/μl) from an adult.

The purity and concentration of the DNA was ascertained by measuring the optical density of the DNA samples using Nano Drop Spectrophotometer.

At this stage and before starting PCR and during PCR, care was taken to avoid contamination as PCR assays require special laboratory practices to avoid false positive amplifications.

3.2.4.2 TaqMan Real Time PCR (Polymerase Chain Reaction)

TaqMan genotyping assays genotype single nucleotide polymorphisms (SNPs) using the 5′ nuclease assay for amplifying and detecting specific SNP alleles in purified genomic DNA samples. Each assay allows to genotype individuals for a specific SNP. Each TaqMan genotyping assay contains two primers for amplifying the sequence of interest and two TaqMan Minor Groove Binder (TaqMan MGB) probes for detecting alleles. The presence of two probe pairs in each reaction allows genotyping of the two possible variant alleles at the SNP site in a DNA target sequence. The genotyping assay determines the presence or absence of a SNP based on the change in fluorescence of the dyes associated with the probes.

The TaqMan® MGB Probes consist of target-specific oligonucleotides with: first a reporter dye at the 5′ end of each probe (VIC dye is linked to the 5′ end of the Allele 1 probe and FAM dye is linked to the 5′ end of the Allele 2 probe) (Table 11). Second, a MGB increased the melting temperature (Tm) without increasing probe length (Afonina L et al 1997), thereby allowing the design of shorter probes; shorter probes resulted in greater differences in Tm values between matched and
mismatched probes, resulting in accurate allelic discrimination. Third, a non-fluorescent quencher (NFQ) was at the 3' end of the probe. Because the quencher did not fluoresce, Applied Biosystems Real Time PCR system could measure reporter dye contributions with great sensitivity.

The assays are supplied by Life Technologies "Applied Biosystems", USA. The detailed procedure can be checked by TaqMan® SNP genotyping assays protocol booklet provided by Applied Biosystems.

Table 11: Fluorescence signal correlations

<table>
<thead>
<tr>
<th>Fluorescence Increase</th>
<th>Indication</th>
</tr>
</thead>
<tbody>
<tr>
<td>VIC dye Fluorescence only</td>
<td>Homozygosity for allele 1</td>
</tr>
<tr>
<td>6FAM dye Fluorescence only</td>
<td>Homozygosity for allele 2</td>
</tr>
<tr>
<td>Fluorescence signals for both dyes</td>
<td>Heterozygosity for allele 1 and allele 2</td>
</tr>
</tbody>
</table>

Basics of the 5' Nuclease Assay and during PCR

Genomic DNA was introduced into a reaction mixture consisting of TaqMan Genotyping Master Mix, forward and reverse primers and two TaqMan MGB Probes. Each TaqMan MGB Probe annealed specifically to its complementary sequence, if present, between the forward and reverse primer sites. When the probe was intact, the proximity of the reporter dye to the quencher dye resulted in quenching of the reporter fluorescence primarily by Forster-type energy transfer (Lakowicz JR 1983). The primers bounded to the template were extended by AmpliTaq Gold DNA Polymerase which in turn cleaved the probes only that were hybridized to the target. The cleavage separated the reporter dye from the quencher dye, which resulted in increased fluorescence signal by the reporter. Thus, the fluorescence signal generated by PCR amplification indicated which alleles were presented in the sample (Figure 12).
CYP1A2 genotyping of all subjects was determined by TaqMan Real Time PCR analysis to detect the five major observed CYP1A2 mutations responsible for the phenotype (slow) of the corresponding allele described as small subset of G-3860A, C-729T, 2116G>A, 2499A>T and 5090C>T substitutions.

The G-3860A substitution was found in the CYP1A2*1C gene. It was located in the 5'-flanking region of human CYP1A2 gene altering guanine (wild type) to adenine (mutated type) at position 2964 in the gene. The C-729T substitution was found in CYP1A2*1K gene. It was located in the intron 1 region of the gene.
changing cytosine (wild type) to thymine (mutated type) at position 873 in the gene as mentioned earlier in table 3.

The 2116G>A missense substitution was found on CYP1A2*3 gene cluster. It was located in exon 4 causing Asp348Asn. The 2499A>T missense substitution was found on the CYP1A2*4 gene cluster. It was located in exon 5 leading to Ile386Phe. The 5090C>T missense substitution was found on the CYP1A2*6 gene cluster. It was located in exon 7 leading to Arg431Trp. The WT allele was formerly defined as the absence of G-3860A, C-729T, 2116G>A, 2499A>T and 5090C>T nucleotide substitutions as mentioned earlier in table 3.

The presence of these mutations "G-3860A, C-729T, 2116G>A, 2499A>T and 5090C>T" were determined by specific designed assays. Each assay was able to genotype individuals for a specific SNP. Each TaqMan genotyping assay contained two primers for amplifying the sequence of interest and two TaqMan Minor Groove Binder (TaqMan MGB) probes for detecting alleles. The presence of two probe pairs in each reaction defined the genotyping of the two possible variant alleles at the SNP site in a DNA target sequence. The genotyping assay determined the presence or absence of a SNP based on the change in fluorescence of the dyes associated with the probes.

In our assay, TaqMan genotyping assays were used to identify the important SNPs that we investigated in CYP1A2 gene (rs2069514, rs12720461, rs56276455, rs72547516 and rs28399424). The assays with their IDs had been defined by the manufacturer company (ThermoFisher Scientific) for each dbSNP with their specific primers that would detect a specific single nucleotide changes. Particularly, G-3860A, C-729T, 2116G>A, 2499A>T and 5090C>T that represented the major SNPs/alterations responsible for the slow phenotype of the corresponding *1C, *1K, *3, *4 and *6 haplotypes, respectively (Table 12).
Table 12: Assay ID or design that has been used for each SNP

<table>
<thead>
<tr>
<th>Assay ID</th>
<th>dbSNP</th>
<th>Haplotype</th>
<th>Forward and Reverse Primers Context Sequence (VIC/FAM)*</th>
<th>Forward fluorescence signal generated</th>
<th>Reverse fluorescence signal generated</th>
<th>Allele Nomenclature</th>
</tr>
</thead>
<tbody>
<tr>
<td>C_1585</td>
<td>rs20695</td>
<td>At position 3860</td>
<td>TGGCTCACCAGCAACCTCC GCCTCTC[G/A]GATTCAAG CAAATTGTGATGCCCCAG</td>
<td>G</td>
<td>A</td>
<td>Wild type allele</td>
</tr>
<tr>
<td>9191_30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C_3063</td>
<td>rs12720</td>
<td>At position 729</td>
<td>GGCTAGGTGGTAGGGGTC CTGAGTT[C/T]GGGCTTT GCTACCCACGCTCTTGACT</td>
<td>C</td>
<td>T</td>
<td>Wild type allele</td>
</tr>
<tr>
<td>4146_10</td>
<td>461</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C_3063</td>
<td>rs56276</td>
<td>At position 2116</td>
<td>GTACATGGGGGGCCTCCCA ACCCTATA[G/A]ACAGAGG AAGATCCAGAAGGAGCTG</td>
<td>G</td>
<td>A</td>
<td>Mutant allele</td>
</tr>
<tr>
<td>4247_20</td>
<td>455</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C_3063</td>
<td>rs72547</td>
<td>At position 2499</td>
<td>ACCTCCTCTTTCTTGCC CTCACC[A/T]TCCTCCTCCACT AGGTGAGGCTTGCCGT</td>
<td>A</td>
<td>T</td>
<td>Wild type allele</td>
</tr>
<tr>
<td>4246_10</td>
<td>516</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C_3063</td>
<td>rs28399</td>
<td>At position 5090</td>
<td>GGACCCCTCTGAGATCTCG GCCCTGAG[C/T]GGTCTTCT CACCGCGGTAGGCCACTG</td>
<td>C</td>
<td>T</td>
<td>Mutant allele</td>
</tr>
<tr>
<td>4244_20</td>
<td>424</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Dote represent no signal has been generated

Storage and Stability

TaqMan Genotyping Master Mix is stable when stored at 2 to 8 °C.

Specific Guidelines

Particularly for TaqMan real time PCR, defining the method for adding DNA is important. Wet DNA delivery was selected where the aliquot genotyping reaction mixed to an optical reaction plate and delivered genomic DNA to the final reaction mix.

Nevertheless, proper selection of the instrument and reaction plate is also essential in order to detect and record the fluorescent signals generated by the cleavage of TaqMan probes. For that, Applied Biosystems Thermal Cycler QuantStudio 7 Flex (278870068) with MicroAmp™ Optical 96-well reaction plate;
reaction volume 20uL was used throughout the experiments. Figure 13 summarizes the procedures for performing a genotyping experiment.

Finally, quantifying genomic DNA by adding 1 to 10 ng of DNA template per reaction well to have a uniformed DNA concentration with all the samples and along the experiment. Quantitate genomic DNA has been done using Nano Drop Spectrophotometer machine for DNA quantification.

![Diagram of genotyping experiment]

Figure 13: Overview of a genotyping experiment using Real Time PCR
Preparing the Reaction Mix Guidelines

For optimal PCR performance: first, all TaqMan reagents were kept protected from light until they were ready to use as excessive exposure to light may affect the fluorescent probes. Second, freeze-thaw cycles were minimized. Third, prior to use: the TaqMan Genotyping Master Mix was mixed thoroughly by swirling the bottle, thawing any frozen TaqMan reagents by placing them on ice and when thawed, the samples were re-suspended by vortexing then centrifuging the tubes briefly. Finally, the PCR reaction mix was prepared for each assay before transferring it to the optical reaction plate for thermal cycling and fluorescence analysis.

The number of reactions were calculated to be performed for each assay with one extra reaction for each ten to provide excess volume for the loss that occur during reagent transfers. Then, the volume was calculated of each reaction mix component needed for each assay “20ul reaction for each sample using 96-well standard plate”.

The reaction plate was prepared using wet DNA delivery method by pipetting the reaction mix into each well of a reaction plate, inspecting each well for volume uniformity, noting which wells do not contain the proper volume, covering the plate with MicroAmp Optical Adhesive Film, centrifuging the plate briefly to spin down the contents and eliminating any air bubbles from the solutions, diluting 1 to 10 ng of each purified genomic DNA sample into DNAase-free water, pipetting samples into the plate, covering the plate with MicroAmp Optical Adhesive Film or MicroAmp Optical Caps and centrifuging the plate briefly to spin down the contents and eliminate air bubbles from the solutions.

Later, amplification of genomic DNA was performed by the Sequence Detection System (SDS) using the thermal cycling condition that specified and
optimized to be used with TaqMan genotyping assay only (Figure 14) where 1 cycle
denaturation at 60°C initially for 30 sec, then held at 95°C for 10 min, then 40 cycle
(denature at 95°C for 15 sec and anneal/extend at 60°C for 1 min), then 1 cycle at
60°C for 30. Standard mode thermal cycling setting was selected in the plate
document (Figure 15), the reaction volume was settled accordingly and the reaction
plate was loaded the thermal cycler to start running.

Figure 14: The thermal cycling condition specified for our TaqMan genotyping
assays for CYP1A2

Figure 15: The Standard mode thermal cycling setting
Allelic Discrimination Plate Read and Analysis

After PCR amplification, an endpoint plate read was performed on a Real Time PCR instrument (Applied Biosystems Thermal Cycler QuantStudio 7 Flex model no. 278870068) using the fluorescence measurements (Rn) made during the plate read and the SDS software plots Rn values based on the fluorescence signals from each well.

The plotted fluorescence signals indicated which alleles were in each sample "The SDS software plots showed the results of the allelic discrimination run on a scatter plot of Allele 1 versus Allele 2. Each well of the 96-well plate was represented as an individual point on the plot" (Figure 16). The alleles were determined in each sample as following: first, the allelic discrimination plate read documents were created and set up (Figure 17). Second, post-PCR allelic discrimination plate read was performed on a real time PCR instrument (Figure 18). Third, the plate read documents were analyzed. Finally, an automatic or manual allele calls and allele types verified were made.

![Figure 16: Variation in clustering due to the genotype of the target allele](image-url)
Figure 17: The allelic discrimination plate read document

Figure 18: RT-PCR plate layout during sample running in the system
The CYP1A2*1A haplotype represented the wild type allele whereas CYP1A2*1C, CYP1A2*1K, CYP1A2*3, CYP1A2*4 and CYP1A2*6 genotypes represented the mutant alleles. Individuals who were homozygote for CYP1A2 polymorphisms (*1C, *1K, *3, *4 and *6) were classified as slow phenotype. Individuals who were heterozygote for CYP1A2 polymorphisms had an intermediate phenotype. Individuals who lacked CYP1A2 polymorphisms had a extensive or normal phenotype.

3.2.4.3 Polymerase Chain Reaction - Restriction Fragment Length Polymorphism (PCR – RFLP)

3.2.4.3.1 Polymerase Chain Reaction (PCR)

Amplification of genomic DNA was performed by the polymerase chain reaction (PCR) using the pair of forward and reverse oligonucleotide primers of NAT2, master mix that is specified for NAT2 amplification and to be amplified at certain conditions as described by Hickman & Sim (1991).

Reagents for PCR

Tenfold PCR buffer concentrate, including magnesium chloride (15mM), was supplied by the manufacturer Bioline of the DNA Taq polymerase, the Q solution, Deoxynucleoside triphosphates (dNTPs), restriction enzymes: MspI, Ddel, FokI, kpnI, TaqI and BamHI, restriction enzyme buffer 10-fold concentrate supplied by the manufacturer of the restriction enzymes, sterile water for the amplification and themostable DNA Taq polymerase supplied by the manufacturer.

The reagents were stored at -20 °C and the work was done on ice through all the procedures, the PCR tubes were kept on ice until placed in the thermal cycler. The Q solution facilitates amplification of templates that have a high degree of secondary structure or that are GC-rich by modifying the melting behavior of DNA.
The PCR buffer provides a final concentration of 1.5mM MgCl2 in the final reaction mix which provided satisfactory results.

**Amplification of 442bp and 559bp Segments of NAT2 Gene Sequence**

Genomic DNA was obtained from the buccal swab samples of subjects by specific procedure for extraction. DNA amplification was performed by thermocycler GeneAmp® PCR system 9700 version 3.12 using 0.1-0.5 µM of oligonucleotide primers (forward primers; 5'-GTC ACA CGA GGA M T CM ATG C-3' and 5'-ACA CAA GGG TTT ATT TTG TTC C-3') 18.6nmol/100ul and (reverse primers; 5'-ACC CAG CAT CGA CAA TGT AAT TCC TGC CCT CA-3' and 5'-AAT TAC ATT GTC GAT GCT GGG T-3') 11.03nmol/100ul described by Hickman and Sim (1991) in a 20 µl mixture of 2ul of 10X solution (PCR buffer contain 15mM MgCl2, 4ul of Q solution (5X), 0.2ul of dNTP solution 200uM, 11ul H2O (sterile water for amplification), 0.2ul of Taq DNA polymerase 2.5 units/reaction, 0.8ul of each primer and 1ug of the sample DNA. Reagents and mixture were kept on ice during preparation. Negative control (DNA not added) was used for quality control.

The 10X buffer, dNTP mix, primer solutions, Q-solution, sterile water for amplification and Taq DNA polymerase were thawed on ice and the master mix was kept on ice after complete thawing and mixed thoroughly before use to avoid localized differences in salt concentration. The reaction mix was prepared accordingly and mixed gently but thoroughly by pipetting up and down a few times. Then, the template DNA (≤ 1ug/reaction) was added to the individual PCR tubes containing the reaction mix. Finally, amplification was done using GeneAmp® PCR system 9700 version 3.12. Amplification conditions were denaturation 1 cycle at 94°C for 5 min, then 35 cycles (94°C 30 sec, 56°C 1 min, 72°C 2 min), then 1 cycle 72°C 7 min, then cooling at 4°C (Figure 19).
Figure 19: The thermocycler program for the NAT2-PCR

Reagents for Agarose Gel Electrophoresis

Agarose, bromophenol blue, ethidium bromide “store in dark place”, Boric acid (H3BO3), ethylenediaminetetraacetic acid disodium salt dehydrate (EDTA) and tris-hydroxymethylaminomethane (TRIS). The last three were used to prepare the TBE buffer concentrate (Table 13). The agarose gel was prepared by dissolving agarose in TBE solution in different concentrations based on the needs. One percent gel concentration was prepared to run the control PCR before digestion. One percent, 1.5% and 2.5% gel concentration were prepared to run PCR after digestion with certain enzymes. Then, the gel was placed in electrophoresis machine with addition of Ethidium bromide (0.5ug/ml) 5uL.

The 100bp DNA ladder (supplied by Promega Corporation, USA) was composed of bands where the fragments of DNA samples would be compared to. It consisted of eleven double-stranded DNA fragments with sizes of 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 and 1500bp. The intensity of these fragments appeared to be equal on the gel with an exception of the 500bp band which showed triple the intensity of other fragments. Accordingly, the DNA ladder and DNA samples were added into specific wells to be run in the gel for certain time (60-90 minutes) under 80-100 V cm⁻¹.
Table 13: Preparation of TBE solution

<table>
<thead>
<tr>
<th>Reagent</th>
<th>2L</th>
<th>1L</th>
<th>500mL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boric acid</td>
<td>110g</td>
<td>55g</td>
<td>27.5g</td>
</tr>
<tr>
<td>TRIS</td>
<td>216g</td>
<td>108g</td>
<td>54g</td>
</tr>
<tr>
<td>EDTA</td>
<td>14.8g</td>
<td>7.4g</td>
<td>3.7g</td>
</tr>
</tbody>
</table>

Dissolve TRIS and EDTA first then add Boric acid. Make up to volume with distilled water.

3.2.4.3.2 Restriction Fragment Length Polymorphism (RFLP)

In molecular biology, restriction fragment length polymorphism (RFLP) is a technique that exploits variations in homologous DNA sequences. It refers to a difference between samples of homologous DNA molecules from different locations of restriction enzyme sites. In RFLP analysis, the DNA sample was broken into pieces (digested) by restriction enzymes and the resulting restriction fragments were separated according to their lengths by gel electrophoresis. **NAT2** genotyping of all subjects was carried out by PCR-RFLP analysis as described by Vatsis KP et al (1995), Woolhouse NM et al (1997) and Chida M et al (1999).

**NAT2** genotype of all subjects was determined by PCR-RFLP analysis to detect the four commonly observed **NAT2** mutation associated with slow acetylation usually described as minor allele of 191G>A (rs1801279), 341T>C (rs1801280), 590G>A (rs1799930) and 857G>A (rs1799931). The 341T>C (rs1801280) substitution is found in **NAT2***5* gene cluster. The 590G>A missense substitution is found in **NAT2***6* gene cluster. The 857G>A missense substitution is found on **NAT2***7* gene cluster. The 191G>A missense substitution is found on the **NAT2***14* gene cluster (Vatsis K P et al 1995). The WT allele was formerly defined as the absence of 341T>C, 590G>A, 857G>A and 191G>A nucleotide substitutions.

The presence of these mutations "341T>C, 590G>A, 857G>A and 191G>A" were determined by restriction endonuclease digestion using Ddel, TaqI, BamHI and MspI, respectively. Then, the presence of 282C>T, 481C>T and 803A>G were
determined by restriction endonuclease digestion using FokI, KpnI and Ddel respectively to identify the gene cluster or subtype. After amplification of DNA, digestion of the PCR product (1ug - quantitate genomic DNA has been done using Nano Drop Spectrophotometer machine for DNA quantification.) was carried out in a total volume of 50ul using the appropriate digestion buffer (Table 14). The PCR product obtained using aforementioned primers was digested in separate experiments for 6hrs at 37° with KpnI, Ddel, FokI, BamHI and Mspl and for 8hrs at 65° with TaqI. Digested DNA (10ul) then was separated by electrophoresis on 1% for Mspl restriction enzyme, 1.5% for Ddel, FokI, KpnI and BamHI restriction enzyme and 2.5% for TaqI restriction enzyme agarose gel under 80-100 V cm-1 (Table 14). DNA bands were visualized with ethidium bromide and UV transillumination.

Table 14: The restriction cleavage for the amplificates of the PCR product

<table>
<thead>
<tr>
<th>Volumes for one sample</th>
<th>KpnI</th>
<th>TaqI</th>
<th>BamHI</th>
<th>Mspl</th>
<th>Ddel</th>
<th>FokI</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR product</td>
<td>1ug</td>
<td>1ug</td>
<td>1ug</td>
<td>1ug</td>
<td>1ug</td>
<td>1ug</td>
</tr>
<tr>
<td>Master mix</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Restriction enzyme</td>
<td>1ul</td>
<td>1ul</td>
<td>1ul</td>
<td>1ul</td>
<td>1ul</td>
<td>1ul</td>
</tr>
<tr>
<td>10X NE Buffer</td>
<td>5ul</td>
<td>5ul</td>
<td>5ul</td>
<td>5ul</td>
<td>5ul</td>
<td>5ul</td>
</tr>
<tr>
<td>Total Rxn Volume</td>
<td>50ul</td>
<td>50ul</td>
<td>50ul</td>
<td>50ul</td>
<td>50ul</td>
<td>50ul</td>
</tr>
<tr>
<td>Incubation</td>
<td>37° C</td>
<td>65° C</td>
<td>37° C</td>
<td>37° C</td>
<td>37° C</td>
<td>37° C</td>
</tr>
<tr>
<td>Temperature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incubation Time</td>
<td>6hrs</td>
<td>8 hrs.</td>
<td>6hrs.</td>
<td>6 hrs.</td>
<td>6 hrs.</td>
<td>6 hrs.</td>
</tr>
<tr>
<td>Can also be used overnight</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>with no star activity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agarose gel to run the samples</td>
<td>1.5%</td>
<td>2.5%</td>
<td>1.5%</td>
<td>1%</td>
<td>1.5%</td>
<td>1.5%</td>
</tr>
</tbody>
</table>

Primers

The pairs of forward and reverse oligonucleotide primers were (forward primers; 5'-GTC ACA CGA GGA AAT CAA ATG C-3' and 5'-ACA CAA GGG TTT ATT TTG TTC C-3') and (reverse primers; 5'-ACC CAG CAT CGA CAA TGT AAT TCC TGC CCT CA-3' and 5'-AAT TAC ATT GTC GAT GCT GGG T-3') as described
by Hickman & Sim (1991) respectively were used to amplifying 442bp and 559bp segment of NAT2 gene sequence. The 442bp fragment was used for 191G>A, 282C>T, 341T>C and 481C>T mutations detection, whereas the 559bp fragment was used for 590G>A, 803A>G and 857G>A mutations detection.

The primers were dissolved in sterile water so that concentrations of 100pmol/ul were available. They were stored at -20°C. The primers were stable for approximately 1 to 2 years under these conditions.

**DNA Sequencing**

DNA sequencing was done using Applied Biosystems 3130xl Genetic Analyzer machine for certain alleles that were not analyzed based on the designed evaluation sheet described below (Table 16). DNA sequencing was done for certain samples randomly as a confirmation of the accuracy of our technique. The gene accession number is NG_012246.1 and the transcript accession number is NM_000015.2, CDS 13760 – 14632.

**Nomenclature**

Using RFLP analysis, the presence/absence of polymorphism(s) in each DNA sample was identified. Accordingly, the genotype was determined as per the last update of NAT nomenclature scheme published by The Arylamine N-acetyltransferase Gene Nomenclature Committee according to consensus guidelines (http://nat.mbg.duth.gr/Human%20NAT2%20alleles_2013.htm).

**Allelic Linkage Analysis - Restriction Map of NAT2 and Restriction Cleavage Site**

On completion of electrophoresis, the wet gel was placed on the UV trans-illuminator and the ethidium bromide intercalated in the DNA was excited to cause fluorescence emission at 312 nm. The band pattern was photographed with the
camera of the evaluation device and printed out. The results for the investigated sample were recorded by marking the pertinent field in an evaluation sheet (Table 16) for each allele variation detected in the DNA fragments (Table 15). The allele combination in the samples can be determined by superimposing the evaluation scheme (Table 16) of the same scale. The evaluation is complete, when the variations specified on the scheme can be assigned to all points (reference sequence = wild type or sequence variation = mutant). If only one row on the evaluation scheme showed a match with the entries in the evaluation row, the person was a homozygous allele carrier. If two rows of the evaluation sheet matched with the evaluation scheme rows, the person was a heterozygous allele carrier. If it was impossible to determine which of the alleles concerned was really present, DNA sequencing was done to identify these alleles.

DNA fragment lengths from the restriction cleavage has been illustrated by the following table (Table 15 and Figure 20) where the reference sequence (RS) = wild type, sequence variation (SV) = mutant and the control PCR = 442bp and 559bp.

Table 15: The restriction cleavage site and its DNA fragment lengths from the restriction cleavage

<table>
<thead>
<tr>
<th>Restriction cleavage</th>
<th>Mutation site</th>
<th>Sequence - fragment lengths (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mspl</td>
<td>191</td>
<td>RS = 181, 168, 93</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SV = 274, 168</td>
</tr>
<tr>
<td>Fokl</td>
<td>282</td>
<td>RS = 337, 105</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SV = 442</td>
</tr>
<tr>
<td>Ddel</td>
<td>341</td>
<td>RS = 221, 163, 58</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SV = 189, 163, 58, 32</td>
</tr>
<tr>
<td>Kpnl</td>
<td>481</td>
<td>RS = 424, 135</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SV = 559</td>
</tr>
<tr>
<td>Taql</td>
<td>590</td>
<td>RS = 226, 170, 142, 21</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SV = 396, 142, 21</td>
</tr>
<tr>
<td>Ddel</td>
<td>803</td>
<td>RS = 345, 124, 90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SV = 345, 97, 90, 27</td>
</tr>
<tr>
<td>BamHI</td>
<td>857</td>
<td>RS = 515, 44</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SV = 559</td>
</tr>
</tbody>
</table>

First PCR = 442       Second PCR = 559
The *4 allele represents the wild type allele whereas *5, *6, *7 and *14 haplotypes represent the mutant alleles. Individuals who were homozygous for NAT2 polymorphisms (NAT2*5, NAT2*6, NAT2*7 and NAT2*14) were classified as slow acetylator phenotype. Individuals who were heterozygous for NAT2 polymorphisms had an intermediate acetylator phenotype. Individuals who lacked NAT2 polymorphisms had a rapid acetylator phenotype.

**Allelic Linkage Analysis - Restriction Enzymes**

Evaluation sheet (✓ = fragment length combination found) and restriction enzymes with mutation sites and fragment length for NAT2 have been documented in the following table (Table 16) where the control PCR = 442 bp and 559 bp to show how we assigned NAT2 alleles. Nevertheless, examples of fragment lengths found for 15 samples have been clearly illustrated as guidance.
Table 16: Evaluation sheet (✓ or √ = fragment length combination found) of the restriction enzymes with mutation sites and fragment length for NAT2

<table>
<thead>
<tr>
<th>First PCR = 442</th>
<th>Second PCR = 559</th>
</tr>
</thead>
<tbody>
<tr>
<td>MspI (191)</td>
<td>FokI (282)</td>
</tr>
<tr>
<td>WT</td>
<td>Mu</td>
</tr>
<tr>
<td>Fragment lengths (bp) after restriction</td>
<td></td>
</tr>
<tr>
<td>181</td>
<td>274</td>
</tr>
<tr>
<td>168</td>
<td>168</td>
</tr>
<tr>
<td>93</td>
<td>58</td>
</tr>
<tr>
<td>32</td>
<td>21</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Examples of fragment lengths</th>
</tr>
</thead>
<tbody>
<tr>
<td>'4'</td>
<td>Wild type allele</td>
</tr>
<tr>
<td>*5A</td>
<td>Mutant allele</td>
</tr>
<tr>
<td>*5B</td>
<td>Mutant allele</td>
</tr>
<tr>
<td>*5C</td>
<td>Mutant allele</td>
</tr>
<tr>
<td>*5D</td>
<td>Mutant allele</td>
</tr>
<tr>
<td>*6A</td>
<td>Mutant allele</td>
</tr>
<tr>
<td>*6B</td>
<td>Mutant allele</td>
</tr>
<tr>
<td>*7B</td>
<td>Mutant allele</td>
</tr>
<tr>
<td>*14A</td>
<td>Mutant allele</td>
</tr>
<tr>
<td>*14B</td>
<td>Mutant allele</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample code</th>
<th>Examples of fragment lengths found for 15 samples</th>
<th>Assigned NAT2 allele combinations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>2</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>3</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>4</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>5</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>6</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>7</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>8</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>9</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>10</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>11</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>12</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>13</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>14</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>15</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
3.3 Data Analysis

The CYP1A2 haplotype construction and analysis was performed using QuantStudio™ 6 and 7 Flex Real Time PCR System Software v1.0 through Thermo Fisher Cloud Scientific Analysis, instrument type: QuantStudio 7 Real-Time PCR System. Slow and rapid CYP1A2 genotypes were reported according to the Human CYP-Allele Nomenclature Committee.

The NAT2 haplotype construction and analysis was performed with the PHASE v2.1.1 program (Stephens M et al 2001, Agundez JA et al 2008 and Selinski S et al 2013a) using the default model for recombination rate variation (Li N and Stephens M 2003). Seven independent runs with 1000 iterations, 500 burn-in iterations, and a thinning interval of 1. The run that showed the maximum consistency of results across all runs was selected to be the best run, as described by Agundez et al. 2008. Slow and rapid NAT2 genotypes were reported according to the consensus NAT2 gene nomenclature (Hein DW et al 2008).

Data entry analysis SPSS (Statistical Package for the Social Sciences) version 19 was used for data entry and analysis. Kruskal-Wallis test, Spearman's rho and other statistical tests were used for analysis like Paired sample T test, Pearson Chi-Square, Likelihood Ratio, Regression, Phi, Cramer's V, Odds Ratio and Relative Risk tests. A p value of 0.05 was considered statistically significant. The deviation from Hardy-Weinberg Equilibrium was also calculated for each genotype to determine whether observed genotype frequencies are consistent with Hardy-Weinberg equilibrium using Chi-Square test.
Chapter 4: Results

4.1 Population Data

The study was carried out on 556 non-smoker Emirati males ("95% – mean age 18 years") and 25 females ("5% – mean age 25 years"). Five of the recruited (~1%; 3 males and 2 females) had been assigned to standardize the procedures and protocols before proceeding with the study samples and 576 subjects (99%; 553 males and 23 females) were assigned as study samples.

The age ranged from 16 – 51 years old; the vast majority of the recruited subjects were between 16 and 19 years of age. (Figure 21).

![Age Distribution](image)

Figure 21: The age distribution of the recruited subjects

All of our subjects were from UAE population who were originally from this region. Out of all subjects, 269 (46.3% of the total recruits) were from Abu Dhabi Emirate (96.6% of them were males and 3.4% females), 113 (19.4%) were from Dubai Emirate (95.5% males and 4.5% females) and 199 (34.3%) were from Northern Emirates (94.4% males and 5.6% females).

The distribution of subjects who were enrolled in this study and the areas where samples were collected represented almost the same distribution of UAE population (Emiratis) between the Emirates. There is no significant difference in
sample distribution between the Emirates compared to the data provided by UAE National Bureau of statistics of UAE (Emiratis) population (P = 0.262; using 95% confidence interval of differences) using Paired Samples T test.

4.2 Descriptive Analysis of the Recruited Sample

Out of all sample subjects 290 (50.3%) were children of consanguineous marriages reflecting the high of consanguinity among Emiratis. Out of those 290 subjects, 185 subjects (63.7% “181 males and 4 females”) had third degree related parents and 105 (36.3% “101 males and 4 females”) had the same tribe or family name. On the other hand, 286 subjects (49.7% “271 male and 15 females) were children of non-relative parents.

Out of all sample subjects 520 (90% “499 males and 21 females”) were healthy whereas 56 (10% “54 males and 2 females”) had medical illness. Out of those with history of illness 3 subjects (0.6%) had allergic conditions, 4 subjects (0.8%) had cardiovascular disease, 23 subjects (4%) were with asthma, 2 subjects (0.4%) had epilepsy, one subject (0.2%) had hypercholesterolemia, one case (0.2%) had vertebral disc disease, 4 subjects (0.8%) had diabetes mellitus, 16 cases (2.9%) had Sickle-cell disorder, one (0.2%) was with migraine and (0.2%) had thyroid disease.

Out of all sample subjects, 555 (96.3%; 532 males and 23 females) reported no sensitivity to medications whereas 21 (3.7%; 21 males) subjects had reported sensitivity to medications. Out of them 2 cases (0.3%) had antibiotics allergy (not specified by the participants), 6 subjects (1%) had sensitivity to G6PD deficiency related drugs, 2 cases (0.3%) were sensitive to ibuprofen and 11 cases (1.9%) had allergies of unknown cause.
4.3 HPLC Analysis

High performance liquid chromatography (HPLC) was performed by separating the metabolites included AFMU, 1MU, 1MX, 17MU, 17MX and 137MX, and their levels were determined based on the area under the curve that has been specified through specific retention time for each metabolites peak (AFMU = 3.182, 1MU = 4.868, 1MX = 5.812, 17MU = 8.279, 17MX = 8.588 and 137MX = 12.440) measured by specific standards (Figure 22).

![Figure 22](image)

Figure 22: Representative chromatogram curves that are specific for caffeine metabolites with their retention time

The standards included AFMU, 1MU, 1MX, 17MU, 17MX and 137MX. Spectral conformation of caffeine metabolite peaks was performed routinely, using a programmable multiple wavelength detector (Waters 2998PDA, Waters Corp., Milford MA). Using this methodology, recovery of caffeine, paraxanthine, AFMU and 1-methylxanthine ranged from 91 to 103% (Figure 23, 24, 25, 26, 27 and 28).
Figure 23: Standard Spectral conformation of 137MX, 17MX, 17MU, 1MU, 1MX and AFMU. The amount on X axis is in ug/ml
4.4 Statistical Analysis

4.4.1 CYP1A2

4.4.1.1 CYP1A2 Genotype

Genetically, we were able to genotype all cases for the selected polymorphism using TaqMan Real Time PCR analysis. We found out of 576 subjects 8 (1.4%) subjects were homozygote for mutant alleles, 93 (16.1%) were heterozygote and 475 (82.5%) homozygous for the wild type allele genotyping (Figure 24).

![Figure 24: The genotype frequencies of the studied sample.](image)

The overall mutant alleles' frequency was 0.095 with allele CYP1A2*1C frequency = 0.0636, allele CYP1A2*1K = 0.0287, allele CYP1A2*3= 0.0027, allele CYP1A2*4= 0 and allele CYP1A2*6 = 0 while the wild type allele frequency was found to be 0.905 (Figure 25).
Figure 25: Allele frequencies- allele CYP1A2*1C frequency = 0.0636, allele CYP1A2*1K = 0.0287, allele CYP1A2*3= 0.0027, allele CYP1A2*4= 0 and allele CYP1A2*6= 0. The wild type allele frequency was found to be 0.905

The following figure (26) shows an example of allelic discrimination plot that demonstrates a variation in clustering due to the genotyping of the targeted allele (CYP1A2*1C) using QuantStudio™ 6 and 7 Flex Real Time PCR System Software v1.0.

Figure 26: An example of allelic discrimination plot that demonstrates a variation in clustering due to the genotyping of the targeted allele (CYP1A2*1C)

We are able to identify six different CYP1A2 diplotypes. Out of the 576 subjects, 568 (98.6%) subjects were homozygote and heterozygote for the wild type allele". Of whom, 475 (82.5%) subjects were homozygote for the wild type allele, 59
(10.2%) subjects were heterozygote for CYP1A2*1A*1C genotype, 33 (5.8%) were heterozygote for CYP1A2*1A*1K genotype and 1 (0.2%) was heterozygote for CYP1A2*1A*3 genotype. Whereas, 8 (1.4%) subjects were homozygote for the mutant alleles. Of whom, 7 (1.2%) subjects were homozygote for allele CYP1A2*1C and 1 (0.2%) subjects was homozygote for allele CYP1A2*3 (Table 17).

Table 17: The CYP1A2 genotype frequencies and distribution

<table>
<thead>
<tr>
<th>CYP1A2 genotype</th>
<th>n</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP1A2<em>1A</em>1A</td>
<td>475</td>
<td>0.825</td>
</tr>
<tr>
<td>CYP1A2<em>1A</em>1C</td>
<td>59</td>
<td>0.102</td>
</tr>
<tr>
<td>CYP1A2<em>1A</em>1K</td>
<td>33</td>
<td>0.058</td>
</tr>
<tr>
<td>CYP1A2<em>1A</em>3</td>
<td>1</td>
<td>0.002</td>
</tr>
<tr>
<td>CYP1A2<em>1C</em>1C</td>
<td>7</td>
<td>0.012</td>
</tr>
<tr>
<td>CYP1A2<em>3</em>3</td>
<td>1</td>
<td>0.002</td>
</tr>
</tbody>
</table>

The characteristic features and frequencies of the various CYP1A2 haplotype identified among Emiratis, determined from TaqMan Real Time PCR analysis of genomic DNA is shown in table 18. The most common haplotypes by far among our population sample belonged to the CYP1A2*1A (rs2069514) gene and occurred with a frequency of 0.905. Three different CYP1A2 alleles associated with slow activity were found. The frequencies of these slow alleles CYP1A2*1C, CYP1A2*1K and CYP1A2*3 were less common than wild type allele, the allele belonged to the CYP1A2*1C (rs12720461) gene occurring with a frequency of 0.0636. Whereas, the frequency of CYP1A2*1K (rs56276455) and CYP1A2*3 (rs72547516) genes were negligible 0.0287 and 0.0027, respectively. The frequency of CYP1A2*4 (rs28399424) and CYP1A2*6 (rs2069514) alleles were absent in this population (Table 18).

While examining the deviation from Hardy-Weinberg equilibrium, it was also found that the observed genotype frequencies were different from the expected one. nonetheless, the data for CYP1A2*1K were in accordance with the Hardy-Weinberg equilibrium since P>0.05. For CYP1A2*1C, the data did not obey the equilibrium
since \( P = 0.001 \). \( CYP1A2*3, CYP1A2*4 \) and \( CYP1A2*6 \) were too rare for the equilibrium to be determined (Table 18).

Table 18: The characteristic features and frequencies of the various \( CYP1A2 \) haplotypes identified among Emiratis

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>dbSNP</th>
<th>Enzyme activity</th>
<th>n</th>
<th>Percentage</th>
<th>Hardy-Weinberg Equilibrium</th>
</tr>
</thead>
<tbody>
<tr>
<td>*1A</td>
<td>rs2069514</td>
<td>Normal</td>
<td>1043</td>
<td>90.5%</td>
<td></td>
</tr>
<tr>
<td>*1C</td>
<td>rs12720461</td>
<td>Decreased</td>
<td>73</td>
<td>6.36%</td>
<td>Var. allele freq. = 0.07, ( P=0.001 )</td>
</tr>
<tr>
<td>*1K</td>
<td>rs56276455</td>
<td>Decreased</td>
<td>33</td>
<td>2.87%</td>
<td>Var. allele freq. = 0.03, ( P=0.44 )</td>
</tr>
<tr>
<td>*3</td>
<td>rs72547516</td>
<td>Decreased</td>
<td>3</td>
<td>0.27%</td>
<td>Can't be counted*</td>
</tr>
<tr>
<td>*4</td>
<td>rs28399424</td>
<td>Decreased</td>
<td>0</td>
<td>0%</td>
<td>Can't be counted*</td>
</tr>
<tr>
<td>*6</td>
<td>rs2069514</td>
<td>Decreased</td>
<td>0</td>
<td>0%</td>
<td>Can't be counted*</td>
</tr>
</tbody>
</table>

Var. allele freq. = Variant allele frequency

*Not accurate if <5 individuals in any genotype group.

Based on epidemiological studies, it has been clearly observed that there is a significant ethnic variability in the distribution of common and rare \( CYP1A2 \) SNPs and haplotypes (Zhou SF et al 2009b). The available data suggested that the frequency of \( CYP1A2*1C \) allele was 0.009 in British, 0.008 in Swedish, 0.23 in Japanese, 0.26 in Korean, 0.22 in Chinese, 0.07 in African American, 0.07 in Tunisian, 0.04 in Turkish and 0.07 in Egyptian populations (Nakajima M et al 1999, Chida M et al 1999, Hamdy SI et al 2003, Tiwari AK et al 2005, Bilgen T et al 2008). In the contrary, based on our study; the Emirati population was one of the lowest frequencies with 0.06 when compared to most of the mentioned populations. Taking in consideration, the number of the subjects used in our study was the highest in contrast to other studies (Figure 27A).

Furthermore, the frequency of \( CYP1A2*1K \) allele was 0.003 in Swedish, 0.005 in Spanish, 0.04 in Saudi Arabian and 0.03 in Ethiopian populations, whereas it was absent in Korean and Japanese populations (Aklillu E et al 2003, Ghotbi R et al 2007). Likewise, the genotype frequency of \( CYP1A2*1K \) haplotype in Emiratis
population was slightly lower than Saudi Arabian and Ethiopian population with frequency of 0.02 (Figure 27B).

Lastly, the genotype frequencies of CYP1A2*3, CYP1A2*4 and CYP1A2*6 alleles have been identified in the French population with frequency of (0.005-0.01) (Chevalier D et al 2001, Allorge D et al 2003, Zhou H et al 2004). On the contrary, the frequencies of these haplotypes were lower in the Emirati population specifically the haplotype CYP1A2*3 frequency, that was found to be 0.002. Similarly, the frequencies of CYP1A2*4 and CYP1A2*6 haplotypes were absent in Emiratis population (Figure 27C).
Figure 27: The frequency of CYP1A2 alleles among Emiratis in comparison with other populations, figure (A) shows the CYP1A2*1C haplotype frequencies, figure (B) shows the CYP1A2*1K haplotype frequencies, figure (C) shows the CYP1A2*3, CYP1A2*4 and CYP1A2*6 haplotypes frequencies, $n =$ sample size, Emiratis ($n = 576$), Japanese ($n = 250$), Koreans ($n = 150$), Chinese ($n = 168$), African Americans ($n = 112$), Tunisians ($n = 98$), Turkish ($n = 110$), Egyptians ($n = 212$), Swedish ($n = 194$), French ($n = 100$), British ($n = 114$), Spaniards ($n = 117$), Saudi Arabians ($n = 136$), Ethiopians ($n = 173$)

4.4.1.2 Caffeine Metabolites Ratio and CYP1A2 Phenotype Status

Since other studies had used "$(17MX + 17MU)/137MX$" molar ratio, this measure was calculated and compared in this study (Butler MA et al 1992 and, Lang NP and Kadlubar FF 1991). The frequency distribution of $(17MX + 17MU)/137MX$ in our data indicated a trimodal distribution with specific cut points 2.5 and 6.5 that represented slow, intermediate and rapid acetylators (Joshua EM et al 2008).

Out of 576 subjects, 560 showed positive results whereas the remaining 16 did not complete the study. From those 560 subjects, 8 (1.4%) were slow CYP1A2 enzyme activity "ratio < 2.5", 91 (16.3%) were intermediate activity "ratio between 2.5 -2.6" and 461 (82.3%) were rapid activity for this enzyme "ratio > 6.5" (Table 19).
Table 19: The classification and the results of the CYP1A2 enzyme activity based on the ratio of (17MX+ 17MU)/137MX

<table>
<thead>
<tr>
<th>(17MX+ 17MU)/137MX ratio</th>
<th>phenotype status</th>
<th>No. of subjects</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7 - 2.5</td>
<td>Slow CYP1A2 activity</td>
<td>8</td>
<td>1.4%</td>
</tr>
<tr>
<td>2.51-6.5</td>
<td>Intermediate CYP1A2 activity</td>
<td>91</td>
<td>16.3%</td>
</tr>
<tr>
<td>&gt; 6.51</td>
<td>Rapid CYP1A2 activity</td>
<td>461</td>
<td>82.3%</td>
</tr>
</tbody>
</table>

The ratio of (17MX + 17MU)/137MX ratio was calculated to determine the phenotype status of CYP1A2 activity in each subject. The frequency distribution histogram of urinary (17MX + 17MU)/137MX molar concentration ratio for 560 UAE nationals shows non-normally distributed result "Median = 5.05, Skewness = 1.215 with Std error = 0.103, Kutosis = 1.239 with Std error = 0.206" (Figure 28).

![Figure 28: The distribution of subjects based on CYP1A2 activity (17MX+ 17MU)/137MX ratio](image)

There was clear evidence of a tri-modal distribution of (17MX + 17MU)/137MX ratio with an apparent anti-mode in the region of 2.5 and 6.5 which was in close agreement to that observed by several studies (Afonina I et al 1997). The mode corresponding to the rapid enzyme activity was not uniformly distributed and may include both heterozygous and homozygous rapid phenotype. The frequency distribution histogram suggests the possible existence of a second anti-
mode (in the region of 6.5) separating the heterozygous and homozygous rapid phenotype.

A total of six different genotypes were found, four associated with rapid phenotype and two with slow phenotype. The most common genotype found was CYP1A2*1A/*1A with frequencies of 0.825. Followed by alleles CYP1A2*1A/*1C and CYP1A2*1A/*1K with frequency of 0.102 and 0.058 respectively. The CYP1A2*1C/*1C and CYP1A2*3/*3 genotypes showed significantly the lowest median values 2.06 and 2.00 respectively (Table 20).

Table 20: (17MX + 17MU)/137MX molar ratios in subgroups assigned to CYP1A2 allele combinations

<table>
<thead>
<tr>
<th>CYP1A2 phenotype</th>
<th>CYP1A2 alleles</th>
<th>N</th>
<th>Frequency</th>
<th>Median</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rapid</td>
<td>*1A/*1A</td>
<td>461</td>
<td>0.825</td>
<td>7.58</td>
<td>6.08</td>
<td>25.25</td>
</tr>
<tr>
<td>Intermediate</td>
<td>*1A/*1C</td>
<td>57</td>
<td>0.102</td>
<td>4.98</td>
<td>3.13</td>
<td>6.96</td>
</tr>
<tr>
<td></td>
<td>*1A/*1K</td>
<td>33</td>
<td>0.058</td>
<td>4.87</td>
<td>2.53</td>
<td>6.87</td>
</tr>
<tr>
<td></td>
<td>*1A/*3</td>
<td>1</td>
<td>0.002</td>
<td>6.85</td>
<td>6.85</td>
<td>6.85</td>
</tr>
<tr>
<td>Slow</td>
<td>*1C/*1C</td>
<td>7</td>
<td>0.012</td>
<td>2.06</td>
<td>0.1</td>
<td>2.39</td>
</tr>
<tr>
<td></td>
<td>*3/*3</td>
<td>1</td>
<td>0.002</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>560</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The resulting urinary (17MX + 17MU)/137MX ratio and genotypic assignments of slow (S/S homozygotes), intermediate (R/S heterozygotes) and rapid (R/R homozygotes) are presented in table 21. Using Kruskal-Wallis test, the mean rank of these groups were 4.5, 67.71 and 327.29 respectively, which represented strong association between CYP1A2 genotype and phenotype (X2 = 219.23, df = 2 and P value < 0.0001; 95% CI of differences).

Table 21: (17MX + 17MU)/137MX ratios vs. genotypes. Suggested genotypic state: S/S homozygously slow, R/S heterozygotes intermediate, R/R homozygously rapid

<table>
<thead>
<tr>
<th></th>
<th>S/S homozygotes</th>
<th>R/S heterozygotes</th>
<th>R/R homozygotes</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>8</td>
<td>91</td>
<td>461</td>
</tr>
<tr>
<td>Mean</td>
<td>1.8400</td>
<td>5.0830</td>
<td>9.3085</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>Std. Deviation</td>
<td>Range</td>
</tr>
<tr>
<td>----------------</td>
<td>---------</td>
<td>----------------</td>
<td>-------</td>
</tr>
<tr>
<td></td>
<td>2.0300</td>
<td>0.74310</td>
<td>2.29</td>
</tr>
<tr>
<td></td>
<td>4.9800</td>
<td>1.36216</td>
<td>3.95</td>
</tr>
<tr>
<td></td>
<td>7.5800</td>
<td>3.37762</td>
<td>18.75</td>
</tr>
</tbody>
</table>

S/S homozygotes = slow metabolizers, R/S heterozygotes = intermediate metabolizers, R/R homozygotes = rapid metabolizers

In this study, the phenotype status of CYP1A2 activity in the Emirati population shows the lowest frequency of poor metabolizers with only 1.4% slow phenotype while the percentage in other populations such as Australians, Japanese, Chinese, Americans and Italian were 5%, 14%, 5%, 12% and 13% respectively (Figure 29) (Butler et al 1992, Nakajima M et al 1994, Zhou SF et al 2009b).

![Figure 29: The frequency of CYP1A2 poor metabolizers among Emiratis in comparison with other populations, n = sample size, Emiratis (n = 576), Australians (n = 171), Japanese (n = 250), Chinese (n = 78), Americans (n = 101), Italians (n = 95)](image)

The degree of phenotype/genotype concordance was equal to 81.6% based on the current phenotype classification (i.e. rapid/intermediate/slow phenotypes), where those who were homozygotes for the allele CYP1A2*1A genotypes are considered to be a rapid metaboliser, those who were heterozygotes for the wild type allele are considered to be intermediated metaboliser and those who had no CYP1A2*1A genotype are considered to be a slow metaboliser.
CYP1A2 and Degree of Parent’s Relationship

Out of 8 subjects “homozygote for the mutant alleles”, 2 (25%) subjects were children of third degree relationship, 1 (12.5%) subject was their parents from the same tribe and 5 (62.5%) subjects were not related. On the other hand, out of 568 subjects “homozygote or heterozygote for the wild type allele”, 183 (32.2%) subjects were children of third degree relationship, 104 (18.3%) subjects were parents from the same tribe and 281 (49.%) subjects were not related (Likelihood Ratio = 0.552, degree of freedom (df) = 2 and P value = 0.759; 95% CI of differences) which was more than 0.05 representing a weak evidence of a relationship or association between genotype and degree of relationship. There was no significant correlation between degree of relationship and genotypes with correlation coefficient 0.028 value (P value = 0.508; 95% CI of differences using Spearman’s rho statistical test).

CYP1A2 and “Medical Illness”

With regard to medical illness, it was found that out of 56 cases 1 (1.8%) subject who had medical illness was from those who were homozygote for the mutant alleles compared to 55 (98.2%) subjects from those who were homozygote or heterozygote for the wild type allele (Likelihood Ratio = 7.296, degree of freedom (df) = 10 and P value = 0.697; 95% CI of differences) which was more than 0.05 represented a weak evidence of a relationship or association between genotype and medical illness. There was no significant correlation between medical illness and genotypes with correlation coefficient 0.010 value (P value = 0.804; 95% CI of differences using Spearman’s rho statistical test).

Out of 8 subjects “homozygote for the mutant alleles”, 1 (12.5%) subject had medical illness compared to 7 (87.5%) subjects who had no medical illness. Out of 568 subjects “homozygote or heterozygote for the wild type allele” 55 (9.9%) subjects had medical illness compared to 513 (90.1%) subjects that had no medical
illness with OR = 1.306, relative risk for those with medical illness 1.268 and relative risk for those with no medical illness 0.971 (Likelihood Ratio of $X^2 = 0.058$, df = 1 and P value = 0.81; 95% CI of differences).

**CYP1A2 and Drug Sensitivity**

With regard to drug sensitivity, it was found that out of 21 cases 1 (4.8%) who reported drug sensitivity was from those who were homozygote for the mutant alleles compared to 20 (95.2%) subjects from those who were homozygote or heterozygote for the wild type allele (Likelihood Ratio of $X^2 = 1.339$, degree of freedom (df) = 3 and P value = 0.72; 95% CI of differences) which was more than 0.05 represented a weak evidence of a relationship or association between CYP1A2 genotype and drug sensitivity. There was no significant correlation between drug sensitivity and genotypes with correlation coefficient 0.056 value (P value = 0.179; 95% CI of differences using Spearman's rho statistical test).

Out of 8 subjects "homozygote for the mutant alleles", 1 (12.5%) subject was having drug sensitivity compared to 7 (87.5%) subjects with no drug sensitivity. Out of 568 subjects "homozygote or heterozygote for the wild type allele" 20 (3.5%) subjects had drug sensitivity compared to 548 (96.5%) subjects who had no drug sensitivity with OR = 3.914, relative risk for those with drug sensitivity 3.55 and relative risk for those with no drug sensitivity 0.907 (Likelihood ratio of $X^2 = 1.14$, df = 1 and P value = 0.286; 95% CI of differences).

Using Kruskal-Wallis Test, it was found that there was no difference in the Mean Rank between the groups which represented a lack of effect of CYP1A2 genotype on the Drug sensitivity ($X^2 = 0.909$, df = 3 and P value = 0.823; 95% CI of differences).

**CYP1A2 and Genotype Distribution within the UAE Emirates**
Out of the 8 subjects "homozygote for the mutant alleles" 5 (62.5%) were from AD Emirate, 3 subjects (37.5%) were from Dubai Emirate and no one was found from Northern Emirates. On the other hand, out of 568 subjects "homozygote and heterozygote for the wild type allele" 261 (46%) were from AD Emirate, 108 (19%) were from Dubai Emirate and 199 (35%) were from Northern Emirate. However, this resulted in significant statistical difference between the cities (Likelihood ratio of $X^2 = 7.085$, df = 2 and $P$ value = 0.029; 95% CI of differences) with a positive small effect based on the symmetric measures of Phi and Cramer's $V$ values (Phi = 0.09 and Cramer's $V$ = 0.09) with insignificant $P$ value ($P$ value = 0.097; 95% CI of differences).

### 4.4.2 NAT2

#### 4.4.2.1 NAT2 Genotype

Genetically, we were able to define 570 subjects using PCR-RFLP analysis whereas DNA sequencing has been done for the remaining 6 samples. It was found that out of 576 subjects 446 (77.4%) subjects were carrier of 2 mutant alleles either homozygote or heterozygote, 106 (18.4%) were heterozygote for one mutant allele and 24 (4.2%) homozygous for the wild type allele (Figure 30).
Figure 30: The genotype frequencies- 0.774 of our subjects were homozygote or heterozygote for 2 mutant alleles genotyping, 0.184 were heterozygote and 0.042 were homozygotes for the wild type genotyping

The overall mutant alleles’ frequency was 0.866 with allele NAT2*5 frequency = 0.364, allele NAT2*6 = 0.129, allele NAT2*7 = 0.274 and allele NAT2*14 = 0.099 while the wild type allele frequency was found to be 0.134 (Figure 31).

Figure 31: Allele frequencies- allele NAT2*5 frequency = 0.364, allele NAT2*6 = 0.129, allele NAT2*7 = 0.274 and allele NAT2*14 = 0.099. The wild type allele frequency was found to be 0.134

The different patterns of RFLPs obtained after separate restriction enzyme digestion of the PCR product with Mspl, FokI, Ddel, Kpnl, TaqI and BamHI are shown in the following figure 31. The presence of a Kpnl site is indicated by fragments of size 424bp and 135bp. Similarly, the presence of a BamHI site is indicated by the fragments size 515bp and 44bp. In each instance, the presence of 559bp, indicates loss of the cutting site, and is diagnostic for 481C>T (Kpnl) or 857G>A (BamHI).

After digestion with TaqI, site was indicated by fragments of size 226bp and 170bp. Fragments size 142bp and 21bp also appeared but were uninformative. The presence of 396bp fragment indicates the loss of the polymorphic TaqI site and is
diagnostic for the 590G>A mutation. After digestion with Mspl, site was indicated by fragments of size 181bp and 93bp. Fragments size 168bp also appeared but was uninformative. The presence of 274bp fragment indicates the loss of the polymorphic Mspl site and is diagnostic for the 191G>A mutation.

The presence of Fokl site is indicated by fragments of size 337bp and 105bp. The presence of 442bp fragment indicates the loss of polymorphic Fokl site and is diagnostic for the 282C>T mutation. After digestion with Ddel from the first PCR (442bp), site was indicated by fragment size 221bp. Fragments size 163bp and 58bp also appeared but were uninformative. The presence of 189bp and 32bp fragments indicates the loss of the polymorphic Ddel site and is diagnostic for the 341T>C mutation. Whereas, the digestion with Ddel from the second PCR (559bp), site was indicated by fragment size 124bp. Fragments size 345bp and 90bp also appeared but were uninformative. The presence of 97bp and 27bp fragments indicates the loss of the polymorphic Ddel site and is diagnostic for the 803A>G mutation. The presence of wild type (NAT2*4) allele in subjects phenotyped as rapid acetylators was inferred by exclusion of each of the 191G>A, 341T>C, 590G>A and 857G>A mutations.

The following figure (32) shows the gel documentation of the DNA fragments after amplification and restriction cleavage.
Figure 32: Representative images of the gel documentation of the DNA fragments after amplification and restriction cleavage. bp = base pairs, w = wild type = reference sequence (RS) and m = mutant = sequence variation (SV)

We were able to identify 27 NAT2 diplotypes with the actual genotypes identified in the study group shown in Table 29. The ambiguous diplotypes were clarified completely by haplotype reconstruction with a probability of the reconstructed haplotype pairs of $P>0.97$. The vast majority of haplotype pairs were assigned with a high degree of confidence ($P=1.00$) with two having a probability of $P\geq0.99$ and two others $P\geq0.98$. 
Out of the 576 subjects, 130 (22.6%) subjects were homozygote and heterozygote for the wild type allele. Of whom, 24 (4.2%) subjects were homozygote for the wild type allele, 72 (12.5%) subjects were heterozygote for NAT2*4/*5 genotype, 22 (3.8%) were heterozygote for NAT2*4/*6 genotype, 7 (1.2%) were heterozygote for NAT2*4/*7 genotype and 5 (0.9%) were heterozygote for NAT2*4/*14 genotype. Whereas, 446 (77.4%) subjects were homozygote and heterozygote for the mutant alleles. Of whom, 36 (6.2%) subjects were homozygote for alleles NAT2*5 cluster, 78 (13.5%) subjects were heterozygote for NAT2*5B/*6A genotype, 177 (30.7%) subjects were heterozygote for NAT2*5/*7 clusters, 20 (3.5%) subjects were heterozygote for NAT2*5/*14 clusters, 5 (0.9%) subjects were heterozygote for NAT2*6 cluster, 34 (5.9%) subjects were heterozygote for NAT2*6A/*7B genotype, 5 (0.9%) subjects were heterozygote for NAT2*6A/*14B genotype, 14 (2.4%) subjects were homozygote for NAT2*7B/*7B genotype, 69 (11.9%) subjects were heterozygote for NAT2*7/*14 clusters and 7 (1.2%) subjects were homozygote for NAT2*14A/*14B genotype (Table 22).

Table 22: NAT2 diplotypes and genotypes in the study group. Comparison between genotypes reconstructed by PHASE and actual genotypes obtained by gene mapping

<table>
<thead>
<tr>
<th>Observed diplotype*</th>
<th>No.</th>
<th>Actual genotypes</th>
<th>PHASE reconstruction</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>00000000</td>
<td>24</td>
<td>*4/*4</td>
<td>*4/*4</td>
<td>1.00</td>
</tr>
<tr>
<td>00110000</td>
<td>4</td>
<td>*4/*5A</td>
<td>*4/*5A</td>
<td>1.00</td>
</tr>
<tr>
<td>00110100</td>
<td>52</td>
<td>*4/*5B</td>
<td>*4/*5B</td>
<td>0.98</td>
</tr>
<tr>
<td>00100100</td>
<td>6</td>
<td>*4/*5C</td>
<td>*4/*5C</td>
<td>0.98</td>
</tr>
<tr>
<td>00010000</td>
<td>11</td>
<td>*4/*5D</td>
<td>*4/*5D</td>
<td>1.00</td>
</tr>
<tr>
<td>01001000</td>
<td>19</td>
<td>*4/*6A</td>
<td>*4/*6A</td>
<td>1.00</td>
</tr>
<tr>
<td>00001000</td>
<td>5</td>
<td>*4/*6B</td>
<td>*4/*6B</td>
<td>1.00</td>
</tr>
<tr>
<td>01000010</td>
<td>8</td>
<td>*4/*7B</td>
<td>*4/*7B</td>
<td>1.00</td>
</tr>
<tr>
<td>10000000</td>
<td>6</td>
<td>*4/*14A</td>
<td>*4/*14A</td>
<td>1.00</td>
</tr>
<tr>
<td>00220000</td>
<td>2</td>
<td>*5A/*5A</td>
<td>*5A/*5A</td>
<td>1.00</td>
</tr>
<tr>
<td>00220100</td>
<td>9</td>
<td>*5A/*5B</td>
<td>*5A/*5B</td>
<td>1.00</td>
</tr>
<tr>
<td>01110001</td>
<td>18</td>
<td>5A/*7B</td>
<td>*5A/*7B</td>
<td>1.00</td>
</tr>
<tr>
<td>10110000</td>
<td>4</td>
<td>*5A/*14A</td>
<td>*5A/*14A</td>
<td>0.99</td>
</tr>
<tr>
<td>00220200</td>
<td>17</td>
<td>*5B/*5B</td>
<td>*5B/*5B</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Observed diplotypes are shown as the number of mutations identified in each individual, 0: homozygous reference, 1: heterozygous, 2: homozygous variant; the SNP order is 191, 282, 341, 481, 590, 803 and 857. Ambiguous diplotypes with at least two heterozygous loci are highlighted.

The frequencies of the various haplotypes identified among Emiratis determined from PCR-RFLP analysis of genomic DNA are shown in Table 23. The most common haplotypes by far among our population sample belonged to the *5B gene cluster and occurred with a frequency of 0.301, which is almost three times the frequency of the wild type allele (NAT2*4). Furthermore, a second slow allele belonging to the NAT2*7B, was also more common than allele NAT2*4, occurring with a frequency of 0.274.

Examining the deviation from Hardy-Weinberg Equilibrium, it was also found that the observed genotype frequencies were different from the expected one. However, the apparent numbers of alleles NAT2*7 and NAT2*14 were not in strict accordance with the Hardy-Weinberg Equilibrium. For alleles NAT2*5 and NAT2*6 the data obey the equilibrium since the result was statistically insignificant (variant allele frequency for allele NAT2*5 = 0.55, P value = 0.25; 95% CI of differences and for allele NAT2*6 = 0.31, P value = 0.98; 95% CI of differences) (table 23).

Table 23: The haplotype frequencies of NAT2
### Haplotype Table

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Frequency</th>
<th>Percentage</th>
<th>Hardy-Weinberg Equilibrium</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCTCGAG</td>
<td>*4</td>
<td>154</td>
<td>13.4%</td>
</tr>
<tr>
<td>GCCCTGGG</td>
<td>*5A</td>
<td>45</td>
<td>3.9%</td>
</tr>
<tr>
<td>GCCCTGGG</td>
<td>5B</td>
<td>347</td>
<td>30.1%</td>
</tr>
<tr>
<td>GCCCGGG</td>
<td>*5C</td>
<td>11</td>
<td>0.95%</td>
</tr>
<tr>
<td>GCCCGAG</td>
<td>*5D</td>
<td>10</td>
<td>0.86%</td>
</tr>
<tr>
<td>GTCTAGG</td>
<td>*5U</td>
<td>6</td>
<td>0.52%</td>
</tr>
<tr>
<td>GTTCAAG</td>
<td>6A</td>
<td>143</td>
<td>12.4%</td>
</tr>
<tr>
<td>GCTCAAG</td>
<td>*6B</td>
<td>4</td>
<td>0.34%</td>
</tr>
<tr>
<td>GTTCAGG</td>
<td>*6C</td>
<td>2</td>
<td>0.17%</td>
</tr>
<tr>
<td>GTTCGAA</td>
<td>*7B</td>
<td>316</td>
<td>27.4%</td>
</tr>
<tr>
<td>ACTCGAG</td>
<td>*14A</td>
<td>25</td>
<td>2.2%</td>
</tr>
<tr>
<td>ATTACGAG</td>
<td>*14B</td>
<td>89</td>
<td>7.76%</td>
</tr>
</tbody>
</table>
| Var. allele freq. = Variant allele frequency

Based on epidemiological studies, the frequency of NAT2 slow acetylation alleles in the Emirati population shows the highest frequency with 86.6%. However, in Egyptian, Indian, Omani and Jordanian populations, the frequency of the slow acetylation alleles ranged between (74% - 79%) while in the Moroccan population it was 84% (Meyer UA and Zanger UM 1997, Guaoua S et al 2014). Moreover, in American, German, Spanish, Argentinian and Saudi Arabian populations the frequency was between (72%-73%), whereas in African and Southern Brazil populations it reached 40%, in Senegalese population 60% and in Tunisian population 69%. These populations have high frequencies of mutant allele NAT2*5 and NAT2*6 and low frequencies of mutant allele NAT2*7. In Asian populations, such as Japanese, Chinese, Korean, and Thai, the frequency of the slow acetylation alleles ranged from 10 to 30% with high frequency of mutant allele NAT2*7 and low frequency of mutant allele NAT2*5 (Figure 33) (Meyer UA and Zanger UM 1997, Guaoua S et al 2014).
Figure 33: The frequency of NAT2 slow acetylation alleles among Emiratis in comparison with other populations, n = sample size, Emiratis (n = 576), Asians “Japanese (n = 79), Koreans (n = 288), Chinese (n = 120), Thai (n = 235)”, Africans (n = 97), Americans (n = 387), Tunisians (n = 100), Indians (n = 250), Egyptians (n = 200), Germans (n = 844), Omanis (n = 127), Argentineans (n = 185), Spaniards (n = 258), Saudi Arabians (n = 200), Moroccans (n = 163), Jordanians (n = 150)

Identification of Known Slow Alleles Subtype

DNA sequencing was performed for 20 samples randomly selected to confirm the accuracy of our PCR-RFLP results. Nevertheless, DNA sequencing was also performed for the remaining 6 samples that had not been clearly defined using PCR-RFLP.

Table 24 provided more detailed data of the 6 samples that had been analysed using DNA sequencing. Whereas, figure 34 illustrated DNA chromatogram that showed 510T>C variant of sample 3 DNA with well-resolved peaks and no ambiguities.
Table 24: Details of the results of the 6 samples that has been analysed using DNA sequencing

<table>
<thead>
<tr>
<th>Sample Number</th>
<th>Variants</th>
<th>Allele1</th>
<th>Allele2</th>
<th>Assigned Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 1</td>
<td>282C&gt;T 590G&gt;A</td>
<td>282C&gt;T 590G&gt;A</td>
<td>282C&gt;T 590G&gt;A</td>
<td>Homozygote NAT2<em>SU</em>SU</td>
</tr>
<tr>
<td></td>
<td>341T&gt;C 803G&gt;A</td>
<td>341T&gt;C 803G&gt;A</td>
<td>341T&gt;C 803G&gt;A</td>
<td>NAT2<em>5B</em>5B</td>
</tr>
<tr>
<td></td>
<td>481C&gt;T</td>
<td>481C&gt;T</td>
<td>481C&gt;T</td>
<td></td>
</tr>
<tr>
<td>Sample 2</td>
<td>341T&gt;C 803G&gt;A</td>
<td>341T&gt;C 803G&gt;A</td>
<td>341T&gt;C 803G&gt;A</td>
<td>Homozygote NAT2<em>5B</em>5B</td>
</tr>
<tr>
<td></td>
<td>481C&gt;T</td>
<td>481C&gt;T</td>
<td>481C&gt;T</td>
<td></td>
</tr>
<tr>
<td>Sample 3</td>
<td>282C&gt;T 590G&gt;A</td>
<td>282C&gt;T 590G&gt;A</td>
<td>282C&gt;T 590G&gt;A</td>
<td>Heterozygote NAT2<em>6C</em>6C?</td>
</tr>
<tr>
<td></td>
<td>510T&gt;C 803G&gt;A</td>
<td>510T&gt;C 803G&gt;A</td>
<td>510T&gt;C 803G&gt;A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>775A&gt;C</td>
<td>775A&gt;C</td>
<td>775A&gt;C</td>
<td></td>
</tr>
<tr>
<td>Sample 4</td>
<td>282C&gt;T 803G&gt;A</td>
<td>282C&gt;T 803G&gt;A</td>
<td>282C&gt;T 803G&gt;A</td>
<td>Heterozygote NAT2<em>SU</em>SU?</td>
</tr>
<tr>
<td></td>
<td>341T&gt;C 775A&gt;C</td>
<td>341T&gt;C 775A&gt;C</td>
<td>341T&gt;C 775A&gt;C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>481C&gt;T 733A&gt;C</td>
<td>481C&gt;T 733A&gt;C</td>
<td>481C&gt;T 733A&gt;C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>590G&gt;A</td>
<td>590G&gt;A</td>
<td>590G&gt;A</td>
<td></td>
</tr>
<tr>
<td>Sample 5</td>
<td>341T&gt;C 803G&gt;A</td>
<td>341T&gt;C 803G&gt;A</td>
<td>341T&gt;C 803G&gt;A</td>
<td>Homozygote NAT2<em>5B</em>5B</td>
</tr>
<tr>
<td></td>
<td>481C&gt;T</td>
<td>481C&gt;T</td>
<td>481C&gt;T</td>
<td></td>
</tr>
<tr>
<td>Sample 6</td>
<td>341T&gt;C 803G&gt;A</td>
<td>341T&gt;C 803G&gt;A</td>
<td>341T&gt;C 803G&gt;A</td>
<td>Homozygote NAT2<em>SU</em>SU</td>
</tr>
<tr>
<td></td>
<td>590G&gt;A 481C&gt;T</td>
<td>590G&gt;A 481C&gt;T</td>
<td>590G&gt;A 481C&gt;T</td>
<td></td>
</tr>
<tr>
<td></td>
<td>282C&gt;T</td>
<td>282C&gt;T</td>
<td>282C&gt;T</td>
<td></td>
</tr>
</tbody>
</table>

The new variants are highlighted.

Figure 34: An example of DNA chromatogram of T-510C variant of sample 3 DNA

**NAT2*5B**

Out of these 6 samples, two were homozygote for NAT2*5B, an allotypic variant of NAT2*5, that is characterized by presence of 3 mutations, 341T>C, 481C>T and 803G>A. It differs from NAT2*5 only by the presence of the 341T>C and 803G>A mutations and is indistinguishable from NAT2*4 "wild type
allele" when the criterion for identification of NAT2*4 is based only upon exclusion of the common NAT2 slow alleles. Allele carrying 341T>C and 481C>T mutations were presumed to be NAT2*5A in the absence of 803G>A mutation, while alleles carrying 282C>T, 341T>C, 481C>T and 803A>G mutations were presumed to be NAT2*5G. Accordingly, we confirmed the presence of these 3 mutations only by DNA sequencing of 2 samples.

**NAT2*5U**

Out of these 6 samples, one sample was heterozygote and two were homozygote for NAT2*5U, an allotypic variant of NAT2*5, that is characterized by presence of 5 mutations, 282C>T, 341T>C, 481C>T, 590G>A and 803>A. It differs from NAT2*5 only by the presence of the 282C>T, 341T>C, 590G>A and 803G>A mutations. Alleles carrying 341T>C, 481C>T and 803G>A mutations only were presumed to be NAT2*5B in the absence of 282C>T and 590G>A mutations, while allele carrying 282C>T, 341T>C, 481C>T and 803G>A mutations were presumed to be NAT2*5G in the absence of 590G>A mutation. Accordingly, we confirmed the presence of these 5 mutations by DNA sequencing of 3 samples.

**NAT2*6C**

Out of these 6 samples, one sample was heterozygote for NAT2*6C, an allotypic variant of NAT2*6, that is characterized by presence of 3 mutations, 282C>T, 590G>A and 803G>A. It differs from NAT2*6 only by the presence of the 282C>T and 803G>A mutations. Alleles carrying 590G>A, 282C>T and 111T>C mutations were presumed to be NAT2*6D, while allele carrying 590G>A and 803G>A mutations were presumed to be NAT2*6F in the absence of 282C>T mutation. Accordingly, we confirmed the presence of these 3 mutations only by DNA sequencing of one sample.
Unreported Variants and Unknown Slow Alleles Subtype

After the confirmation of the presence of two samples that were homozygote for \emph{NAT2*5B}, two samples were homozygote for \emph{NAT2*5U}, one sample was heterozygote for \emph{NAT2*6C/Nat2*6?} (Unreported) and one sample who was heterozygote for \emph{NAT2*5U/NAT2*5?} (Unreported). Three new variants were confirmed using DNA sequencing with repetitions. These variants have been checked in different databases like Ensembl, ExAC and GlaxC database and were found that they were not reported before. The new variants were "510T>C, 775A>C and 733A>C". They were presented as heterozygote in \emph{NAT2*5} and \emph{NAT2*6} clusters and related to nucleotides changes in the protein coding regone.

4.4.2.2 Caffeine Metabolites Ratio and NAT2 Phenotype Status

Since other studies had used (AFMU/1MX) molar ratio, in this study, we used similar measures of calculation and comparison. Caffeine phenotype categories were defined as slow, intermediate or rapid activators based on the distribution of these values with cut points at 0.5 and 2 (see aforementioned Table 10) (Grant DM et al 1984 and Woolhouse N et al 1997).

The ratio of AFMU/1MX was calculated to determine the phenotype status of NAT2 activity in each subject. Out of 576 subjects, 555 showed positive results whereas the data from the remaining 21 were missing. From those 555 subjects, 436 (78.5%) were slow acetylators for NAT2 activity "ratio < 0.5", 106 (19.1%) were intermediate "ratio between 0.5 -2" and 13 (2.4%) subjects were rapid acetylators for this enzyme "ratio > 2" (Table 25).

Table 25: The classification and the results of the acetylator status of NAT2 activity and its sub-classification based on (AFMU/1MX) ratio

<table>
<thead>
<tr>
<th>AFMU/1MX ratio</th>
<th>Acetylator status</th>
<th>No. of subjects</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 0.5</td>
<td>Slow NAT2 acetylator</td>
<td>436</td>
<td>78.5%</td>
</tr>
</tbody>
</table>
The frequency distribution histogram of urinary AFMU/1MX molar concentration ratio for 555 UAE nationals was non-normally distributed (Median = 0.2535, Skewness = 3.633 with Std error = 0.104, Kurtosis = 18.418 with Std error = 0.207) (Figure 35).

Figure 35: The distribution of subjects based on NAT2 activity (AMFU/1MX) ratio

There was a clear evidence of a bi-modal distribution of AFMU/1MX ratio with an apparent anti-mode in the region of 0.5. The mode corresponding to the slow acetylator phenotype was uniformly distributed and was assumed to be made up of individuals who possess various combinations of several different alleles, known to determine slow acetylation.

The mode corresponding to the rapid acetylator phenotype was not uniformly distributed and is assumed to include both heterozygous and homozygous rapid acetylators. The frequency distribution histogram suggests the possible existence of a second anti-mode (in the region of 2) that separates the heterozygous and homozygous rapid acetylators.

Table 26 and figure 36 shows the acetylation capacity of the 25 subgroups assigned to NAT2 allele combinations. The \( \text{NAT2}^*6A/^*6A, \text{NAT2}^*6A/^*7B, \)
NAT2*7B/*7B, NAT2*5A/*5B and NAT2*5A/*5A genotypes showed significantly the lowest median values 0.006, 0.018, 0.029, 0.095 and 0.132 respectively. Nine different genotypes associated with rapid acetylation and sixteen with slow acetylation, the most common genotypes found were NAT2*5B/*7B heterozygotes, NAT2*5B/*6A heterozygotes, NAT2*7B/*14B heterozygotes and NAT2*4/*5B heterozygotes with frequencies of 0.255, 0.135, 0.105 and 0.09, respectively.

Table 26: AFMU/1X molar ratios in subgroups assigned to NAT2 allele combinations

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>NAT2 alleles or possible allele combination</th>
<th>N</th>
<th>Median</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rapid</td>
<td>*4/*4</td>
<td>13</td>
<td>2.5430</td>
<td>2.0634</td>
<td>4.3321</td>
</tr>
<tr>
<td>Intermediate</td>
<td>*4/*5A</td>
<td>4</td>
<td>0.5788</td>
<td>0.5249</td>
<td>0.7992</td>
</tr>
<tr>
<td></td>
<td>*4/*5B</td>
<td>52</td>
<td>0.7955</td>
<td>0.5200</td>
<td>1.8045</td>
</tr>
<tr>
<td></td>
<td>*4/*5C</td>
<td>6</td>
<td>0.9055</td>
<td>0.6102</td>
<td>1.6078</td>
</tr>
<tr>
<td></td>
<td>*4/*5D</td>
<td>10</td>
<td>0.8665</td>
<td>0.6190</td>
<td>1.7049</td>
</tr>
<tr>
<td></td>
<td>*4/*6A</td>
<td>18</td>
<td>0.8435</td>
<td>0.4569</td>
<td>1.7478</td>
</tr>
<tr>
<td></td>
<td>*4/*6B</td>
<td>4</td>
<td>1.6562</td>
<td>1.6078</td>
<td>1.6790</td>
</tr>
<tr>
<td></td>
<td>*4/*7B</td>
<td>7</td>
<td>0.7326</td>
<td>0.5633</td>
<td>1.0496</td>
</tr>
<tr>
<td></td>
<td>*4/*14A</td>
<td>5</td>
<td>0.7645</td>
<td>0.5382</td>
<td>1.4808</td>
</tr>
<tr>
<td>Slow</td>
<td>*5A/*5A</td>
<td>2</td>
<td>0.1323</td>
<td>0.1210</td>
<td>0.1436</td>
</tr>
<tr>
<td></td>
<td>*5A/*5B</td>
<td>9</td>
<td>0.0956</td>
<td>0.0281</td>
<td>0.5269</td>
</tr>
<tr>
<td></td>
<td>5A/*7B</td>
<td>24</td>
<td>0.2670</td>
<td>0.0979</td>
<td>0.5445</td>
</tr>
<tr>
<td></td>
<td>*5A/*14A</td>
<td>4</td>
<td>0.4151</td>
<td>0.0815</td>
<td>0.4649</td>
</tr>
<tr>
<td></td>
<td>*5B/*5B</td>
<td>17</td>
<td>0.2455</td>
<td>0.0087</td>
<td>0.5360</td>
</tr>
<tr>
<td></td>
<td>*5B/*5C</td>
<td>3</td>
<td>0.2422</td>
<td>0.1321</td>
<td>0.2500</td>
</tr>
<tr>
<td></td>
<td>*5B/*6A</td>
<td>78</td>
<td>0.2058</td>
<td>0.0146</td>
<td>0.5488</td>
</tr>
<tr>
<td></td>
<td>*5B/*7B</td>
<td>149</td>
<td>0.2152</td>
<td>0.0120</td>
<td>0.6400</td>
</tr>
<tr>
<td></td>
<td>*5B/*14B</td>
<td>16</td>
<td>0.3586</td>
<td>0.0142</td>
<td>0.5275</td>
</tr>
<tr>
<td></td>
<td>*6A/*6A</td>
<td>4</td>
<td>0.0061</td>
<td>0.0021</td>
<td>0.1497</td>
</tr>
<tr>
<td></td>
<td>*6A/*7B</td>
<td>34</td>
<td>0.0188</td>
<td>0.0034</td>
<td>0.1996</td>
</tr>
<tr>
<td></td>
<td>*6A/*14B</td>
<td>5</td>
<td>0.3803</td>
<td>0.0356</td>
<td>0.5125</td>
</tr>
<tr>
<td></td>
<td>*7B/*7B</td>
<td>14</td>
<td>0.0298</td>
<td>0.0051</td>
<td>0.1489</td>
</tr>
<tr>
<td></td>
<td>*7B/*14A</td>
<td>9</td>
<td>0.2453</td>
<td>0.1726</td>
<td>0.4808</td>
</tr>
<tr>
<td></td>
<td>*7B/*14B</td>
<td>61</td>
<td>0.3361</td>
<td>0.0134</td>
<td>0.5787</td>
</tr>
<tr>
<td></td>
<td>*14A/*14B</td>
<td>7</td>
<td>0.2037</td>
<td>0.0147</td>
<td>0.5116</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>555</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 36: Comparison of NAT2 phenotype and genotype in 555 subjects. It illustrates the acetylation capacity of the 25 subgroups assigned to NAT2 allele combinations. Those who are homozygotes for NAT2*4 alleles show the highest AFMU/1MX ratios, those who are heterozygotes for NAT2*4 allele show intermediate AFMU/1MX ratios and those who are homozygotes or heterozygotes for the derived alleles show the slowest AFMU/1MX ratios

The resulting urinary AFMU/1X ratios and genotypic assignments of slow (S/S homozygotes), intermediate (R/S heterozygotes) and rapid acetylators (R/R homozygotes) are presented in Table 27. Using Kruskal-Wallis test, the mean rank of these groups were 217.8, 487.2 and 549 respectively, which represented strong association between NAT2 genotype and phenotype ($X^2 = 282.01$, df = 2 and P value < 0.0001; 95% CI of differences).

Table 27: AFMU/1X ratios vs. genotypes, based on the last update of NAT nomenclature scheme published by The Arylamine N-acetyltransferase Gene Nomenclature Committee according to consensus guidelines (Hein DW et al 2008). Genotypic acetylator state: S/S homozygously slow, R/S heterozygotes, R/R homozygously rapid

<table>
<thead>
<tr>
<th></th>
<th>S/S homozygotes</th>
<th>R/S heterozygotes</th>
<th>R/R homozygotes</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>436</td>
<td>106</td>
<td>13</td>
</tr>
<tr>
<td>Mean</td>
<td>0.25</td>
<td>0.89</td>
<td>2.76</td>
</tr>
<tr>
<td>SD</td>
<td>0.16</td>
<td>0.35</td>
<td>2.62</td>
</tr>
<tr>
<td>Median</td>
<td>0.22</td>
<td>0.79</td>
<td>2.54</td>
</tr>
<tr>
<td>Range</td>
<td>0.53</td>
<td>1.28</td>
<td>2.27</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.01</td>
<td>0.52</td>
<td>2.06</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.53</td>
<td>1.8</td>
<td>4.33</td>
</tr>
</tbody>
</table>
S/S homozygotes = slow acetylators, R/S heterozygotes = intermediate acetylators, R/R homozygotes = rapid acetylators

In this study, the phenotype status of NAT2 activity in the Emirati population shows the highest frequency of poor metabolizers with 78.5% slow phenotype while the percentage in other populations such as Americans, Turkish, Egyptians and Saudi Arabians were 58.1%, 57.4%, 60.5% and 72.3% respectively (Figure 37) (Hamdy et al. 2003, Djordjevic N et al. 2012).

![Figure 37: The frequency NAT2 poor metabolizers among Emiratis in comparison with other populations, n = sample size, Emiratis (n = 576), Americans (n = 255), Turkish (n = 113), Egyptians (n = 200), Saudi Arabians (n = 200)](image)

The degree of phenotype/genotype concordance was calculated by measuring the agreement between the genotype and phenotype methods based on the traditional phenotype classification (i.e. rapid/slow phenotypes), where non containing allele 4 genotypes, are considered to be a slow phenotype, and was equal to 96.2%.

**NAT2 and Degree of Parent’s Relationship**

Out of 446 subjects “homozygote or heterozygote for the mutant alleles”, 141 (31.6%) subjects were children of third degree relatives, 84 (18.8%) subjects had parents from the same tribe and 221 (49.5%) subjects were not related. On the other hand, out of 130 subjects of the “homozygote or heterozygote for the wild type
allele”, 44 (33.8%) subjects were children of third degree relatives, 21 (16.2%) subjects had parents were from the same tribe and 65 (50%) subjects were not related (Pearson Chi-Square = 0.567, degree of freedom (df) = 2 and P value = 0.753; 95% CI of differences) which was more than 0.05 representing a weak evidence of a relationship or association between genotype and degree of relationship. There was no significant correlation between degree of relationship and genotypes with correlation coefficient 0.014 value (P value = 0.729; 95% CI of differences using Spearman’s rho statistical test).

In addition, it was found that 78.1% of the subjects from those whose parents were relatives were homozygote or heterozygote for the mutant alleles whereas, 77.4% from those whose parents were not relatives were homozygote or heterozygote for the wild type allele with OR = 1.041, relative risk for those whose parents were relatives was 1.02 and relative risk for those whose parents were not relatives was 0.98 (X² = 0.040, df = 1 and P value = 0.841; 95% CI of differences).

Using Kruskal-Wallis Test, out 576 subjects, 185 subjects were classified as children with third degree relationship, 105 subjects were classified as relatives from the same tribe and 286 subjects with no relationship. There was no difference in the Mean Rank between the 3 groups which represented no effect on the NAT2 genotype (X² = 0.566, df = 2 and P value = 0.753; 95% CI of differences).

NAT2 and “Medical Illness”

To determine whether medical illness was correlated with genotypes, we found out of 57 cases, 40 (70.2%) who had medical illness were from those who were homozygote or heterozygote for the mutant alleles compared to 17 (29.8%) from those who were homozygote or heterozygote for the wild type allele (Likelihood Ratio = 7.488, degree of freedom (df) = 10 and P value = 0.679; 95% CI of differences) which was more than 0.05 representing a weak evidence of a
relationship or association between genotype and medical illness. There was no significant correlation between medical illness and genotypes with correlation coefficient of 0.007 (P value = 0.957; 95% CI of differences using Spearman’s rho statistical test).

Out of 446 subjects “homozygote or heterozygote for the mutant alleles”, 40 (8.93%) subjects had medical illness compared to 406 (91.07%) subjects were with no medical illness. Out of 130 subjects “homozygote or heterozygote for the wild type allele” 17 (13.28%) subjects were having medical illness compared to 113 (86.72%) who had no medical illness with OR = 0.64, relative risk for those with medical illness 0.672 and relative risk for those with no medical illness 1.05 (X2 = 2.115, df = 1 and P value = 0.146; 95% CI of differences).

Using Kruskal-Wallis Test, it was found that there were no differences in the mean rank between the groups which represented no effect of NAT2 genotype on PMH (X2 = 4.909, df = 10 and P value = 0.897; 95% CI of differences).

**NAT2 and Drug Sensitivity**

With regard to drug sensitivity, it was found that out of 21 cases 17 (81%) who had drug sensitivity were from those who were homozygote or heterozygote for the mutant alleles compared to 4 (19%) from those who were homozygote or heterozygote for the wild type allele (Likelihood Ratio = 1.840, degree of freedom (df) = 3 and P value = 0.606; 95% CI of differences) which was more than 0.05 representing a weak evidence of a relationship or association between NAT2 genotype and drug sensitivity. There was no significant correlation between drug sensitivity and genotypes with correlation coefficient of 0.022 (P value = 0.925; 95% CI of differences using Spearman’s rho statistical test).
Out of 446 subjects "homozygote or heterozygote for the mutant alleles", 17 (3.79%) subjects had drug sensitivity compared to 431 (96.21%) subjects with no drug sensitivity. Out of 128 subjects "homozygote or heterozygote for the wild type allele" 4 (3.13%) subjects were having drug sensitivity compared to 124 (96.88%) who had no drug sensitivity with OR = 1.223, relative risk for those with drug sensitivity 1.214 and relative risk for those with no drug sensitivity 0.993 (Likelihood ratio = 0.132, df = 1 and P value = 0.717; 95% CI of differences).

Using Kruskal-Wallis Test, it was found that no differences in the Mean Rank between the groups which represented no effect of NAT2 genotype on the Drug sensitivity ($X^2 = 1.658$, df = 3 and P value = 0.646; 95% CI of differences).

**NAT2 and Genotype distribution within the UAE Emirates**

Out of the 446 subjects who were homozygote or heterozygote for the mutant alleles 210 (46.69%) were from AD emirate, 90 subjects (20.09%) were from Dubai emirate and 146 (33.04%) were from Northern Emirates with no significant statistical difference between the cities ($X^2 = 2.247$, df = 2 and P value = 0.325; 95% CI of differences).
Chapter 5: Discussions

This study of 576 subjects of Emiratis is considered to be the first complete and comprehensive report on CYP1A2 and NAT2 phenotyping and genotyping on UAE population in which all known or common mutations or polymorphisms of these genes have been fully and thoroughly tested using HPLC and TaqMan Real Time PCR techniques for CYP1A2 gene and HPLC, PCR-RFLP techniques and DNA sequencing for NAT2 gene.

The distribution of samples were based on the distribution of population between emirates which was again based on the data provided by UAE National Bureau of Statistics of Emiratis which represented proper sample distribution all over UAE.

In the past, acetylator phenotypes were determined using different substances. Traditionally, isoniazid has often been used for acetylator phenotype determination. However, isoniazid undergoes complex metabolism and requires serial blood sampling and would not be convenient to use as a screening tool. Other substances containing sulfonamide like sulfadimidine, sulfapyridine, sulfasalazine or sulfamethazine, and based on the classical Bratton-Marshall procedures, which involve the administration of a single oral dose and collection of multiple samples or a single sample of serum or urine, have been used as well. Dapsone has also been used with a single sample of plasma being collected 2-72 hours after administration. The ratio of N-acetylprocainamide to procainamide in plasma has been used to determine acetylator phenotype. In the present study, caffeine was used because it is ubiquitous and present in popular drinks like colas, coffees and teas and can be used in school going children. It is also easier to collect urine samples than blood samples. In this study, urine samples were collected two hours after a drink of cola or coffee.
5.1 CYP1A2

The variability in liver expression of the CYP1A2 gene reached almost 40% and 60% variability in caffeine metabolism; the most often used probe drug for CYP1A2 (Gunes A and Dahl ML, 2008). Only few variants explain the phenotypic variability in CYP1A2 gene expression, contrarily, it was much more with other drug-metabolizing CYPs (Ghotbi R et al, 2007). Ghotbi R et al in 2007 observed a very low frequency of the coding sequence variants in White and Asian populations (Ghotbi R et al, 2007). Browning et al in 2010 examined CYP1A2 variation in Ethiopians, suggested that because of the overall greater incidence of variation there could be some individuals freed from any CYP1A2 activity in that population (Browning SL et al, 2010).

Generally, it has been discussed previously that the exact positioning of the cut-off point (antimode) between phenotypically slow and rapid activities seems laboratory-dependent and attributed to differences in chromatographic conditions (Bolt HM et al, 2005). In this particular population sample, the frequency distribution of the (17MX+17MU)/137MX ratio is clear trimodal.

CYP1A2*1A has been the reference (or "wild type") allele for the respective gene that associated with rapid phenotype status. The other variant alleles have been classified from this by one or more single nucleotide polymorphisms (SNPs). Based on many studies in many ethnic groups, allele *1A was the most common allele (Zhou SF et al, 2009b, Browning SL et al, 2010). Mutant alleles CYP1A2*1C "G-3860A (rs2069514)", CYP1A2*1K "T-739G (rs2069526), C-729T (rs12720461) and C-163A (rs762551)", CYP1A2*3 "2116G>A and 5347T>C (rs56276455)", CYP1A2*4 "2499A>T (rs72547516)" and CYP1A2*6 "5090C>T (rs28399424)" had low catalytic activity associated with the slow phenotype (Zhou SF et al, 2009a,
Browning SL et al 2010) and were less common than the wild type allele (Chevalier D et al 2001, Zhou H et al 2004).

CYP1A2 genotypes have been classified into three different phenotypes: "slow" (two slow alleles), "intermediate" (1 slow and 1 rapid allele), and "rapid" (2 rapid alleles). In our study we found that 1.4% of our subjects were homozygote for mutant alleles, 16.1% were heterozygote and 82.5% were homozygote for the wild type allele.

As a matter of fact, some reports suggest that many influential factors may play an important role in correlating CYP1A2 genotype to CYP1A2 phenotype. For instance, the drugs which cause induction or inhibition of CYP1A2 enzyme may affect the correlation between CYP1A2 genotype and phenotype. Similarly, the environmental factors such as cigarette smoking can affect CYP1A2 enzymatic activity as well. With regard to our study, all of the participants were non-smokers and the majority of them were aged between 16-19 years of age. Additionally, 90% of them were with no medical illness and not on any concomitant drugs. Thus, we found the high degree of concordance between phenotype and genotypes of CYP1A2 gene. Interestingly, we found that those with mutant alleles have shown slow phenotype status. Furthermore, those who were homozygote for alleles CYP1A2*1C and CYP1A2*3 were associated with the slowest enzyme activity.

Hence, when Hardy-Weinberg equation was applied, the apparent numbers of homozygous and heterozygous slow activity were not in strict accordance with the Hardy-Weinberg equilibrium which is a clear indication that individual genotypes cannot be predicted. For allele CYP1A2*1C the result was inconsistent with Hardy-Weinberg Equilibrium which may be due to an unexpectedly large number of CYP1A2*1C homozygotes. For allele CYP1A2*1K the result was consistent with Hardy-Weinberg Equilibrium. This finding may be due to the evolutionary influences,
like mate choice, mutation, selection, genetic drift, gene flow and meiotic drive as one or more of these influences are typically present in real populations. The level of consanguinity in the UAE is very high with more than 50% of marriages being consanguineous. Indeed, about 50% of the sample group reported consanguinity of their parents. This clearly influences the genotype frequencies and their deviation from Hardy-Weinberg.

Our study samples included few numbers with PMH and drug sensitivity and when tested for their relation and effect on genotype we found few significant association. Although, we found a clear associated of CYP1A2 polymorphism and epilepsy with strong statistical evidence of correlation compared to wild type carriers however, the low number of our subjects with epilepsy could be a concern and the need for another study in this regard is important.

Nevertheless, we found the results were statistically significant and differences have been found between cities in regards to genotype and alleles distribution between Emirates. We found CYP1A2 polymorphisms in AD and Dubai Emirates whereas they were absent in Northern Emirates.

5.2 NAT2

In this study, we use the ratio of AFMU/1MX to determine the acetylation status or NAT2 activity. Phenotypically, the differences in NAT2 activity (as measured by AFMU/1MX ratio) between variable populations have been documented and it was markedly high in Koreans compared to Swedes, and this may be due to a higher proportion of the NAT2*4 rapid allele in Koreans and the higher frequency of slow acetylator genotype in Swedes (Djordjevic N et al 2012). In other studies, 58.1% of Americans, 57.4% of Turkish, 60.50% of Egyptians and 72.3% of Saudi Arabians were of slow acetylator phenotype. In our study, NAT2 activity showed a high percentage of slow acetylators (78.5%).
Generally, it has been discussed previously that the exact positioning of the cut-off point (antimode) between phenotypically slow and rapid acetylators seems laboratory-dependent and attributed to differences in chromatographic conditions (Bolt HM et al 2005). In this particular population sample, the frequency distribution of the AFMU/1MX ratio is clearly trimodal.

Some studies have reported good correlation between acetylator phenotype measured by caffeine metabolite ratio and NAT2 genotype (Jetter A et al 2004, Rihs HP et al 2007); contrarily other studies have showed no correlation (Cascorbi I et al 1995, O'Neil WM et al 1997, Zhao B et al 2000, Wolkenstein P et al 2000, Bolt HM et al 2005, Djordjevic N et al 2011). In our study we found a strong statistical evidence of correlation between acetylation-phenotypes and genotypes of NAT2 gene.

The high overall degree of concordance between NAT2 genotype and phenotype thus confirms the validity of genotyping tests to predict NAT2 phenotypes within the UAE population. Genotyping offers several advantages over phenotyping methods; it is a simple and a reliable procedure, obviating the need for the use of probe drugs and is not subject to physiologically or pathologically determined variation. In this particular population sample, the frequency distribution of the AFMU/1MX ratio is clear trimodal since those who were identified as NAT2*4/*4 (wild type) homozygotes, proved to have the highest ratios. Those who were heterozygotes proved to have intermediate ratios and those who were homozygotes for the mutant alleles proved to have the slowest ratios thus providing a strong evidence of gene dosage effect.

Studies on NAT2 are obstructed by issues with unclear haplotype assignment as unphased genotypes of the NAT2 gene SNPs resulted in more than one possible haplotype pair particularly if more than one of these SNPs is
heterozygous. Because of this, PHASE v2.1.1 program was used to enable a rapid and accurate haplotype prediction. However, higher precision of the estimates and more unambiguous information may be obtained by increasing the sample size (Aqundez JA et al 2008 and Selinski S et al 2013a).

NAT2 is a polymorphic gene with around 110 NAT2 alleles being assigned official symbols by the Arylamine N-acetyltransferase Gene Nomenclature Committee, according to consensus guidelines (Hein DW et al 2008). Based on many studies in many ethnic groups, haplotype NAT2*4 carries the wild type allele at all the variable sites (polymorphisms) reported in the NAT2 literature that associated with rapid acetylator phenotype and is not the most common allele (Hein DW et al 1994, Hein et al 1995, O'Neil WM et al 2000). Mutant allele at positions 191, 341, 560 and 857 that represent mutant allele NAT2*14, NAT2*5, NAT2*6 and NAT2*7 respectively, have low catalytic activity associated with the slow acetylator phenotype and are more common than the wild type allele (NAT2*4) (Hein DW et al 1994, Hein et al 1995, O'Neil WM et al 2000). In our study, the frequency of the wild type allele (NAT2*4) among Emiratis was markedly low (0.134) which is similar to many ethnic populations including Caucasians, Egyptians, Omanis, Saudi Arabians, Indians and Moroccans. However, it is markedly considerably lower in comparison to Asians and Hispanics.

The NAT2 gene has a high frequency of functional variation, distinguishing amongst populations that are ethnically different, and has high levels of haplotype variation (Patin E et al 2006a, Mortensen HM et al 2011). The distribution of polymorphisms in our studied population showed high percentage of those with mutant alleles either homozygote or heterozygote (77.4%).

NAT2 genotypes have been classified into three different phenotypes: "slow acetylator" (two slow alleles), "intermediate acetylator" (1 slow and 1 rapid allele),
and "rapid acetylator" (2 rapid alleles) (Stanley LA and Sim E 2008). Some papers used to report rapid as any genotype containing allele NAT2*4 and slow as any non-carriers of allele NAT2*4 acetylator (Soejima M et al 2007). In addition, within the slow acetylator genotype group there is heterogeneity in phenotype due to variation in enzyme activity conferred by different alleles (Hein DW et al 1995, Cascorbi I et al 1999, Hein DW 2009, Ruiz JD et al 2012) which may affect the ability to identify significant associations (Selinski S et al 2013b).

Cartwright RA et al (1982) divided the ratio that was used to measure the phenotype status into 4 groups, (0.3 and greater) corresponded to rapid acetylators, and the other three (0.01–0.09; 0.1–0.19; and 0.2–0.29) corresponded to different levels of slow acetylator phenotype. Urinary bladder cancer risk was markedly elevated as NAT2 phenotype reduced, particularly with the slowest NAT2 phenotype (Cartwright RA et al 1982). Similar conclusion has been reported by different studies (Brockmoller J et al 1996, Okkels H et al 1997, Filadis IF et al 1999).

Interestingly, we observed that the NAT2*6A/*6A, NAT2*6A/*7B, NAT2*7B/*7B, NAT2*5A/*5B and NAT2*5A/*5A genotypes were having the lowest enzyme activity, respectively. This result was supported by Selinski et al 2013b who identified the ultra-slow genotypes. This 'ultra-slow' genotype is defined by a combination of NAT2*6A/*6A, NAT2*6A/*7B, NAT2*7B/*7B haplotype pairs (Ruiz JD et al 2012, Selinski et al 2013b). This provided a clear picture that NAT2 slow acetylator phenotype was not homogeneous, but rather multiple slow acetylator phenotypes that exist resulting from different SNPs and mechanisms.

In addition, a lot of the therapeutic agents have been found to be polymorphically acetylated in humans due to the genetic polymorphism in N-acetyltransferase activity like hydralazine, phenelzine, phenyldiamine, sulphamethazine, isoniazid, sulphasalazine, amonafide, endralazine, procainamide,
a number of sulphonamides, nitrazepam and dapsone. Although, the possibility of failed or less effective clinical response as a result of acetylation polyphorphism is very low because of the wide therapeutic window of the drugs as acetylation is a minor metabolic pathway (Yamasaki Y et al 2008, Ma JJ et al 2009) (Kuhn UD et al 2010). In another study on Japanese patients who were treated with co-trimoxazole found individuals to have higher risk of adverse events with slow acetylator status (NAT2 genotypes *6A/*6A, *6A/*7B, *7B/*7B) compared to rapid acetylators (NAT2 genotypes *4/*4, *4/*5B, *4/*5E, *4/*6A, *4/*7B) (Soejima M et al 2007).

NAT2 alleles containing the 191G>A (rs1801279), 341T>C (rs1801280), 590G>A (rs1799930), and/or 857G>A (rs1799931) missense substitutions are associated with slow acetylator phenotypes. Striking ethnic differences in the frequencies of these missense substitutions (Grant DM et al 1997) play an important role in ethnic differences in frequency of slow acetylator alleles and phenotypes (Cascorbi I and Roots I 1999). For example, the 191G>A substitution commonly represents the NAT2*14 gene cluster that is available in greater extent in African Americans and native Africans, but it is virtually absent in Caucasian populations.

In our study, the most common mutation associated with slow acetylation (341T>C) occurs with a significantly higher frequency (0.364) than other mutant variants and it was similar to what had been reported in other ethnic groups.

Moroccan, Egyptian, Spanish, Omani, Saudi Arabian, American and German populations reported high frequencies of allele NAT2*5 between (42% - 53%), whereas Argentine, African, Indian, Senegalese, Tunisian and Southern Brazil populations reported (28% - 37%). However, it was at a very low frequency in Asian populations (<7%) (Meyer U and Zanger U 1997, Guoua S et al 2014).

Indian population reported the highest frequency of allele NAT2*6 (38%) whereas, German, Omani, American, Egyptian, Argentina, Moroccan, Spanish,
Japanese and Southern Korean populations have frequencies ranging between 22% - 28.5%. However, in Senegalese, Tunisian, African and Southern Brazil populations it was at a very low frequency (10% - 18%) (Meyer U and Zanger U 1997, Guaoua S et al 2014).

In regard to allele NAT2*7 the frequency was the highest in Tunisian and Southern Korean populations 15% and 13.2%, respectively. However, it markedly declined with Argentine, African, Omani, Indian, Egyptian, Southern Brazil and Moroccan populations to (2% - 8%), although it was found to be negligible in American, German, Japanese, Spanish and Senegalese populations (<1%) (Meyer U and Zanger U 1997, Guaoua S et al 2014).

African, Senegalese, Tunisian and Moroccan populations reported the highest frequency of allele NAT2*14 between (4% - 10.3%) where it was found to be at low level with other populations (< 2%) (Meyer U and Zanger U 1997, Guaoua S et al 2014).

Moreover, in one of our previous studies conducted by Woolhouse et al (1997) we had determined the polymorphic N-acetyltransferase (NAT2) genotypes in 106 unrelated Emiratis from Eastern region of Abu Dhabi by PCR-RFLP analysis (Woolhouse et al 1997). We found more than 13 different genotypes, 4 associated with the rapid acetylator phenotype and 9 with the slow acetylator phenotype. The prevalence of slow acetylation alleles were 81.6%. In our study, we reported 27 different genotypes, 9 associated with the rapid acetylator phenotype and 18 with the slow acetylator phenotypes and the prevalence of slow acetylation alleles were 86.6%. The most common allele was NAT2*5.

Interestingly, we were able to report the presence of 3 new variants "510T>C, 775A>C and 733A>C" that have been confirmed by DNA sequencing and were associated with slow acetylation activity. These nucleotide changes represent

Hence, when Hardy-Weinberg equation was applied, the apparent numbers of homozygous and heterozygous slow acetylators were not in strict accordance with the Hardy-Weinberg equilibrium for allele NAT2*7 and NAT2*14, which is a clear indication that individual genotypes cannot be predicted, although, for allele NAT2*5 and NAT2*6 the result was consistent with Hardy-Weinberg Equilibrium. This finding may be due to the evolutionary influences, such as mate choice, mutation, selection, genetic drift, gene flow and meiotic drive, since one or more of these influences are typically present in real populations. The level of consanguinity in the UAE is very high where more than 50% of marriages are consanguineous. Indeed, about 50% of the sample group reported consanguinity of their parents. This clearly influences the genotypes frequencies and their deviation from Hardy-Weinberg.

Our samples have represented few numbers of those with PMH and drug sensitivity and when they have been tested for their relation and effect on genotype we found that the results were not supported by statistical evidence. Nevertheless, we also found the results were not statistically significant and no differences have been found between cities in regard to genotype and alleles distribution between the different Emirates.
Chapter 6: Conclusions

6.1 CYP1A2

The frequency of slow activity CYP1A2 enzyme alleles is very low among Emiratis which correlates with the presence of low frequencies of mutant alleles in CYP1A2 gene. The genotype frequency of the wild type allele is the highest in this population, followed by CYP1A2*1A/*1C and CYP1A2*1A/*1K genotypes, respectively.

Those who are homozygote for alleles CYP1A2*1C and CYP1A2*3 might be at high risk of toxicity with some drugs and some diseases as these alleles are associated with the slowest phenotype status. Consequently, genetic testing is recommended prior to prescribing medications that are largely metabolized by CYP1A2.

6.2 NAT2

There is a high percentage of slow acetylators among Emiratis which correlates with the presence of high frequencies of mutant alleles in NAT2 gene. The genotype frequency of the allele NAT2*5 was the highest in this population. Moreover, the genotype frequency of NAT2*5B/*7B, NAT2*5B/*6A, NAT2*7B/*14B and NAT2*4/*5B were the highest in this population.

Individuals who carried NAT2*6A/*6A, NAT2*6A/*7B, NAT2*7B/*7B, NAT2*5A/*5B or NAT2*5A/*5A genotypes might be at higher risk of toxicity with drugs and some diseases compared to others as these genotypes are associated with the slowest acetylation status. Consequently, genetic testing is recommended prior to prescribing medications that are largely metabolized by NAT2.
References


Altayli E, Gunes S, Yilmaz AF, Goktas S, Bek Y (2009). CYP1A2, CYP2D6, GSTM1, GSTP1, and GSTT1 gene polymorphisms in patients with bladder cancer in a Turkish population. Int Urol Nephrol. 41: 259–266.


An HR, Wu XQ, Wang ZY, Zhang JX, Liang Y (2012). NAT2 and CYP2E1 polymorphisms associated with antituberculosis drug-induced hepatotoxicity in


Buck ML (1997). The cytochrome P450 enzyme system and its effect on drug metabolism. Pediatric Pharmacotherapy. 3(5).


Butler MA, Lang NP, Young JF, (…), Kadlubar FF (1992). Determination of CYP1A2 and NAT2 phenotypes in human populations by analysis of caffeine


Chamorro JG, Castagnino JP, Musella RM, (...), de Larrañaga GF. (2013). Sex, ethnicity, and slow acetylator profile are the major causes of hepatotoxicity induced by antituberculosis drugs. *J Gastroenterol Hepatol.* 28(2): 323-328.


Fernandes MR, de Carvalho DC, dos Santos ÂK, (...), dos Santos NP (2013). Association of slow acetylation profile of NAT2 with breast and gastric cancer risk...


https://galaxc.sengenics.com/login;jsessionid=73C4635A65F960E1D0B4DCB1D3739BA8

http://www.genecards.org/

http://www.genenames.org/

http://www.imm.ki.se/CYPalleles


http://www.snpedia.com/index.php/CYP1A2


Runge D, Kohler C, Kostrubsky V, Janger D, Lehmann T, Runge D, May U, Stolz S, Strom S, Fleig W and Michalopoulos G (2000). Induction of Cytochrome P450 (CYP)1A1, CYP1A2, and CYP3A4 but Not of CYP2C9, CYP2C19, Multidrug Resistance (MDR-1) and Multidrug Resistance Associated Protein (MRP-1) by


Spinasse LB, Santos AR, Suffys PN, (…), Salles GF (2014). Different phenotypes of the NAT2 gene influences hydralazine antihypertensive response in patients with


Tiwari AK, Deshpande SN, Rao AR, Shriharsh V (2005). Genetic susceptibility to tardive dyskinesia in chronic schizophrenia subjects: I. Association of CYP1A2...


List of Publications

Papers:


Appendices

Appendix 1: Research Ethical Approval

UAEU
College of Medicine
and Health Sciences

29th April 2014

Prof. Salim Al Bastaki
Department of Pharmacology and Therapeutics
College of Medicine and Health Sciences, UAE University
Al Ain - UAE

Dear Prof. Salim,

Approval of Extension of the Study.

Re: AAM/WHREC No. 21M059 - Nat2-acetylation phenotyping/genotyping of Emiratis.

The Human Research Ethics Committee (HREC) has approved your extension of research protocol (21M059) titles as above.

The extension will be for a period of two years until May 2016.

The Committee has been organized and operates according to the Good Clinical Practice (GCP) guidelines and has the Federal Wide Assurance (FWA) for the protection of Human Subjects for International (Non-U.S.) Institutions Number 00007109.

I wish to take this opportunity to wish you success with this important study.

With kind regards,

Yours sincerely,

Dr. Fawaz C. Torab
Chair, Al Ain Medical District Human Research Ethics Committee
Appendix 2: Research Ethical Approval - extended

29th April 2014

Prof. Salim Al Bastaki
Department of Pharmacology and Therapeutics
College of Medicine and Health Sciences, UAE University
Al Ain - UAE

Dear Prof. Salim,

Approval of Extension of the Study.


The Human Research Ethics Committee (HREC) has approved your extension of research protocol (21M059) titles as above.

The extension will be for a period of two years until May 2016.

The Committee has been organized and operates according to the Good Clinical Practice (GCP) guidelines and has the Federal Wide Assurance (FWA) for the protection of Human Subjects for International (Non-U.S) Institutions Number 0000007109.

I wish to take this opportunity to wish you success with this important study.

With kind regards,

Yours sincerely,

Dr. Fawaz C. Torab
Chair, Al Ain Medical District Human Research Ethics Committee

Approved

Dr. Ahmed Alswaid

1/11/15
Appendix 3: UAE University Ethical Approval

To Whom It May Concern

This is to inform you that Prof. Salim Bastaki from the Department of Pharmacology & Therapeutics, College of Medicine and Health Sciences, UAEU, is conducting a research study titled "Polymorphic N-Acetyltransferase Phenotyping/Genotyping of Emirates".

This study has been approved by Research Committee and endorsed by Emirates Foundation for Philanthropy (21MO59).

The students will be asked to consume tea and 300 ml of caffeinated cola drinks. Urine samples collected after 2 hours. A swab of the inside of the mouth will be taken for genotyping.

In this study participation is fully voluntary & the student has the right to decline.

Should you need further details, please contact:

Prof. Salim Bastaki
Professor at CMHS, UAEU, Al Ain.
Tel No: +971 3 7137520
Mobile: +971 50 6187855

Best Regards,

Prof. Mohamed Al Balli
Deputy Vice Chancellor for Academic Affairs (Provost)
Appendix 4: Abu Dhabi Education Council Ethical Approval

Date: 10th November 2014
Ref: 10-05-2014

To: Public Schools Principals,

Subject: Letter of Permission

Dear Principals,

The Abu Dhabi Education Council would like to express its gratitude for your generous efforts & sincere cooperation in serving our dear students.

You are kindly requested to allow the researcher, Mohammad Majed Al Ahmad, to complete his research on:

Polymorphic N-Acetransferase Phenotyping/Genotyping of emirates

- Provided submitting a written consent of parents on whom the research will be conducted to the school and subsequently to ADEC's Research Office at the 6th floor.

Please indicate your approval of this permission by facilitating his meetings with the sample groups at your respected schools.

For further information: please contact Mr Helmy Seada on 03/6150140

Thank you for your cooperation.

Sincerely yours,

[Signature]

Mohamed Salim Alazzhari
Director General
Abu Dhabi Education Council

---

[Letter in Arabic]

- شرط قيام الباحث بتسليم موافقة كتابية من أولياء أمور الطلبة المختارين لتطبيق شريطة العملية على الباحث عليهم، إلى إدارتهم ومن ثم إرسالها إلى مكتبنا بال核实 بالملف الساس.
- إذا، برجى الكريم تسجيل موافقة الباحث ومساعدته على إجراء الدراسة المشار إليها.
- الاستفسار: برجى الإشارة بالبريد/ خاص سعده 02/6150140 على الهاتف 03.
- شاكرين لكم حسن تعاونكم

[Signature]
Appendix 5: Dubai Education Zone Ethical Approval

The signatory is:

United Arab Emirates Ministry of Education
Dubai Educational Zone

The approval is dated: 01/12/2023

The reference is: DED/128

Director / Ahmed Ali Al Roushi
Director of Educational System
Appendix 6: RAK Education Zone Ethical Approval

United Arab Emirates
Ministry of Education
RAK Educational Zone
وحدة الريادة التعليمية

<table>
<thead>
<tr>
<th>هاجم الشرف</th>
<th>الشرف</th>
<th>المرسوم للتعليم الثانوي</th>
<th>مؤسسة</th>
</tr>
</thead>
<tbody>
<tr>
<td>050/1966778</td>
<td>محمد خميس بن حجر</td>
<td>1</td>
<td>رأس الخيمة للتعليم الثانوي</td>
</tr>
<tr>
<td>050/4347777</td>
<td>أحمد عبد الله الشمالي</td>
<td>2</td>
<td>مكتب التعليم الثانوي</td>
</tr>
<tr>
<td>050/5100091</td>
<td>أحمد عبد الله حتة</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

المحترم

الدكتور، أحمد الزروفي

سلام عليكم ورحمة الله وبركاته...


بقبول فائق الامتنان والتقدير

مدير منطقة رأس الخيمة التعليمية

اسم: عبد الله بن حارب السويدي

www.rakedzone.ac

07.20255559 07.20255559
07.2330464 07.2330464
Appendix 7: Volunteers consent form

استمارة الموافقة

أنت مدعو للمشاركة في دراسة بحثية بعنوان "النمط الجيني والظاهري لـ Polymorphic N-acetyl transferase"

قبل أن تقرر أنتم أنه من المهم أن تفهم ماذا يحدث وما هو المطلوب منكم.

الهدف من هذه الدراسة هو التقيق في التمثيل الجيني والظاهري للأمراض (إحدى طرق الأمراض) لدى مواطني دولة الإمارات.

إذا قررتم المشاركة سوف تكون ورقة المعلومات هذه لتعفيكما وسيثبت مكان التوقيع عليها. وستكون المعلومات في الأنشاب في أي وقت دون إتباع الأسلوب.

إذا شاركتم في الدراسة مرة واحدة فقط، وسوف تكون هناك حاجة لمشاركة مشاريع أخرى على الكليات مثل الكليات وبيئية وبيئية. وسوف يتلقىون من شرب المعلومات بطلب ميد مستند للبول الخاصة بك، وسوف تؤخذ مسحة من داخل فكك للتمثيل الجيني.

إذا قررت المشاركة في الدراسة، سيتم دعوتكم إلى مقابلة قصيرة شهيرة محتوية على معلومات. أن يكون هناك أي خطر.

ولن تتلقى أي دواء، ولا أي إجراء جديد. سوف تستخدم هذه المعلومات لأغراض البحث العلمي فقط، المعلومات التي ستكون على يد هذه الدراسة قد تستند في في الاستخبارات للعلاج بالعفرين بصروره فضلًا اعتماداً على النمط الجيني للأمراض، وستبقى جميع المعلومات التي يتم جمعها عنك في سرية تامة.

لا يدفع أي مبلغ لإجراء هذا البحث.

للإبلاغ عن المعلومات، يمكنك الاتصال ب:

الدكتور سليم سلطان
استاذ في علم العفرين
كلية الطب، جامعة العربي المتحدة
17666
المكملة: 03
7672030
faks: 03
sba@auaeu.ac.ae

الاسم الكامل للمشاركة بالبحث:

التوقيع:

التاريخ:
Appendix 8: Data collection form

Acetylation Phenotyping and Genotyping Study

Name: ________________________________ ID no: ________________

Nationality: UAE

Are father and mother related? Yes __ No __

If the answer to above is yes what is the relation?

1. Father’s father and mother’s father are brothers.
2. Father’s father and mother’s mother are brother and sister.
3. Father’s father and mother’s mother are sisters.
4. Father’s father and mother’s father are sister and brother.
5. Father’s father and mother’s grandfather are brothers.
6. Father’s father and mother’s grandmother are brother and sister.
7. Father’s father and mother’s grandmother are sisters.
8. Father’s father and mother’s grandfather are sister and brother.
9. Other close relation ______________________
10. Same tribe

Emirate

Abu Dhabi
Dubai
Sharjah
RAK
Ajman
UAQ
Fujairah

Age

Years
Months
Sex Male Female
Smoker

Yes No
Appendix 8: Data collection form - continue

Existence of any medical condition

| Yes | No |

If answer to above is yes, what is (are) the medical condition(s)?

Any mild, moderate or severe reaction to any drug?

| Yes | No |

If the answer to the above question is yes, what is the name of the drug?

If the answer to the above question is yes, what was the reaction that occurred?
Appendix 8: Data collection form – continue

دراسة أشكال التمييز الجيني الاستباغ

الاسم : ____________________________
الرقم الدراسي/الهوية: ____________________________

الجنسية: ____________________________

إماراتي: ____________________________

هل الأب والأم ذات الأصلة؟

نعم: ____________________________
لا: ____________________________

إذا كان الجواب أعلاه ينتمي ما هي العلاقة؟

1. والد الأب وأم الإخوة.
2. والد الأب وأم الأم إخوة.
3. أم الأب وأم الأم إخوة.
4. أم الأب وأم الأم إخوة.
5. والد الأب والجد للأم إخوة.
6. والد الأب والجد للأم إخوة.
7. أم الأب والجد للأم إخوة.
8. أم الأب والجد للأم إخوة.
9. والد الأب والجد للأم إخوة.
10. والد الأب والجد للأم إخوة.
11. غير ذلك.
12. نفس القبيلة.

الإمارة: ____________________________

أم القوين: ____________________________

الفجيرة: ____________________________

أبو ظبي: ____________________________

الشارقة: ____________________________

عجمان: ____________________________

رأس الخيمة: ____________________________

الإمارة: ____________________________

القبيلة: ____________________________

العمر: ____________________________

الشهر: ____________________________

السنة: ____________________________

ذكر: ____________________________

أنثى: ____________________________

مذكور: نعم: ____________________________
لا: ____________________________

أذكرها إذا وجد

هل يوجد أي حالة مرتبطة

نعم: ____________________________
لا: ____________________________

أذكر الاسم الدواء ووصف الحالة

هل يوجد أي حالة مرتبطة أدواء

نعم: ____________________________
لا: ____________________________

أذكر الاسم الدواء