

# Targeted and untargeted liquid chromatography-mass spectrometry (LC-MS) metabolic profiling in population cohort studies

Thesis submitted by Deepti Sood

For the degree of Doctor of Philosophy (PhD) of Imperial College London  
on 29<sup>th</sup> June 2016

Supervisors: Dr Hector Keun, Dr Elizabeth Want, & Dr Muireann Coen

Computational Systems Medicine  
Department of Surgery and Cancer  
Faculty of Medicine  
Imperial College London

## **Declaration of Originality**

I declare herewith, that this thesis and the research to which it refers is solely my work. I have fully referenced and acknowledged any ideas discussed by other authors and have used inverted commas for all text quoted from other publication and books. Furthermore, I confirm that any third-party figures used in this thesis were reconstructed and fully acknowledged in accordance with good academic practices.

29<sup>th</sup> June 2016

Deepti Sood

## Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Work done in Chapter 2 is published in British Journal of Cancer

Sood, D. *et al.* CYP3A7\*1C allele is associated with reduced levels of 2-hydroxylation pathway oestrogen metabolites. *British Journal of Cancer* 116, 382–388 (2017).

## Abstract

Incorporating liquid chromatography-mass spectrometry (LC-MS) based metabolite profiling into molecular epidemiological studies holds the potential to identify biomarkers that link genetics, environmental exposures (the exposome) and phenotypes to multifactorial diseases. This thesis applied and examined three LC-MS platforms for three different kinds of studies in the context of exposome research.

A targeted, fully quantitative LC-MS/MS method was applied to quantify 14 urinary endogenous oestrogen metabolites; the results demonstrated identification of metabolic phenotype for the genetic polymorphism in CYP3A gene associated with breast cancer in pre-menopausal women. A combined targeted LC-MS/MS and direct infusion-MS/MS method were used to explore the metabolic response to prenatal exposures to persistent organic pollutants (POPs) in 1<sup>st</sup> trimester maternal and cord serum. The results showed an association of POPs with concentration variation in citrulline and diacyl phosphatidylcholines in both maternal and cord serum and kynurenine/tryptophan ratio in mothers indicating altered indoleamine 2, 3-dioxygenase enzyme activity. Finally, an untargeted LC-MS metabolite profiling pilot study was conducted to extend the limited knowledge of the urinary metabolites level temporal variability in healthy children. Despite high precision, the data demonstrated large inter- and intra-individual variation in metabolite levels.

Through these studies, the broader challenges for the successful application of LC-MS metabolome analysis in identifying molecular biomarkers from population studies are explored. These specifically include data comparability, lack of longitudinal variability data and pre-analytical biases. In summary, this thesis illustrates the value of LC-MS to exposome research and identifies the strengths and weakness of different experimental strategies used to characterise metabolic phenotypes in population studies.

## **Acknowledgments**

It is my great pleasure to take this opportunity to express my sincere gratitude to so many people without whom this PhD research and writing this thesis wouldn't have been possible. First and foremost, I would like to thank my supervisor Dr Hector Keun for giving me this opportunity and his continuous support throughout my PhD. I am ever so grateful to you Hector; this PhD wouldn't have happened if you didn't see a fit for me in your group. I would also like to acknowledge wholeheartedly my other two supervisors Dr Elizabeth Ward and Dr Murieann Coen for their insightful comments which widened my research from various perspectives. My sincere thanks go to the funding projects of this PhD, HELIX and Envirogenomarkers and their members outside Imperial for their support and co-operation.

My sincere thanks also go to Dr Alexandros Siskos, Dr Pooja Jain, Dr Macrelo Segeura and Mr Mark Bennett for their invaluable guidance, support and motivation. You all provided me with the right guidance and equipped me with the correct knowledge and tools I needed to conduct my research successfully. Also, I would like to thank Dr Olivia Fletcher for the excellent collaboration and opportunity I was given to conduct joint research with Institute of Cancer Research.

I thank my friends Nicole, Kate and Abi for always being there for me. Finally, I would like to thank my family: my parents, my sister and Jerome for supporting me throughout writing this thesis and my life in general.

## CONTENTS

|                                                                                                                                             |    |
|---------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Declaration of Originality</b> .....                                                                                                     | 2  |
| <b>Copyright Declaration</b> .....                                                                                                          | 3  |
| <b>Abstract</b> .....                                                                                                                       | 4  |
| <b>Acknowledgments</b> .....                                                                                                                | 5  |
| <b>List of Figures</b> .....                                                                                                                | 10 |
| <b>List of Tables</b> .....                                                                                                                 | 13 |
| <b>Abbreviations</b> .....                                                                                                                  | 15 |
| <b>Chapter 1. General Introduction</b> .....                                                                                                | 17 |
| 1.1 Metabolic Profiling .....                                                                                                               | 17 |
| 1.1.1 Size and diversity of the metabolome.....                                                                                             | 19 |
| 1.1.2 Physiological Range and dynamics of the metabolome.....                                                                               | 19 |
| 1.1.3 Genetic and environmental exposure and health .....                                                                                   | 20 |
| 1.2 Technologies in Metabolic Profiling.....                                                                                                | 23 |
| 1.2.1 Liquid chromatography (LC) .....                                                                                                      | 25 |
| 1.2.1.1 Theory of chromatography .....                                                                                                      | 27 |
| 1.2.2 Direct Infusion (DI) or Flow infusion analysis (FIA) .....                                                                            | 30 |
| 1.2.3 Mass spectrometry .....                                                                                                               | 31 |
| 1.2.4 The Triple Quadrupole (TQ or QQQ) mass analyser.....                                                                                  | 33 |
| 1.2.5 The Time of Flight (TOF) mass analyser.....                                                                                           | 34 |
| 1.3 Experimental strategies in metabolic profiling.....                                                                                     | 36 |
| 1.3.1 Targeted metabolite profiling .....                                                                                                   | 36 |
| 1.3.1.1 Quantitative analysis and construction of a calibration curve .....                                                                 | 37 |
| 1.3.2 Untargeted metabolic profiling .....                                                                                                  | 40 |
| 1.3.2.1 Role of QC samples in untargeted metabolic profiling .....                                                                          | 41 |
| 1.4 Conceptual applications of targeted and untargeted LC-MS metabolic profiling .....                                                      | 44 |
| 1.4.1 Functional genomics .....                                                                                                             | 46 |
| 1.4.2 Biomarker discovery.....                                                                                                              | 47 |
| 1.4.3 Exposome studies.....                                                                                                                 | 48 |
| 1.5 Statistical analysis used in this thesis.....                                                                                           | 49 |
| 1.6 Current challenges in biomarker discovery with metabolic profiling: characterising the clinically recognised metabolic phenotypes ..... | 51 |
| 1.7 Aim and outline of the thesis.....                                                                                                      | 53 |
| 1.8 References.....                                                                                                                         | 55 |

|                                                                                                                                                                                                                         |           |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| <b>Chapter 2: Application of an LC-MS/MS method for quantitative analysis of fourteen endogenous oestrogens in urine from premenopausal females: application to carriers of the CYP3A7*1C allele (AC genotype).....</b> | <b>61</b> |
| 2.1 Introduction .....                                                                                                                                                                                                  | 61        |
| 2.1.1 Oestrogen metabolism .....                                                                                                                                                                                        | 63        |
| 2.1.1.1 The 2-hydroxylation pathway .....                                                                                                                                                                               | 66        |
| 2.1.1.2 The 4-hydroxylation pathway .....                                                                                                                                                                               | 67        |
| 2.1.1.3 The 16-hydroxylation pathway .....                                                                                                                                                                              | 67        |
| 2.1.2 The role of genetic variation in oestrogen metabolism and risk of breast cancer .....                                                                                                                             | 68        |
| 2.1.3 A Liquid chromatography-tandem mass spectrometry (LC-MS/MS) method to quantify endogenous oestrogen metabolites (EMs) .....                                                                                       | 69        |
| 2.1.4 Hypothesis and Aim.....                                                                                                                                                                                           | 71        |
| 2.1.5 Objectives .....                                                                                                                                                                                                  | 71        |
| 2.2 Materials and Methods .....                                                                                                                                                                                         | 73        |
| 2.2.1 Materials .....                                                                                                                                                                                                   | 73        |
| 2.2.2 Enzymatic hydrolysis buffer and derivatisation reagent.....                                                                                                                                                       | 74        |
| 2.2.3 Preparation of calibration curve standards and quality control samples .....                                                                                                                                      | 74        |
| 2.2.4 Sample preparation: Hydrolysis, Extraction and Derivatisation.....                                                                                                                                                | 75        |
| 2.2.5 High-Performance Liquid Chromatography- Tandem Mass Spectrometry (HPLC-MS/MS) analysis.....                                                                                                                       | 77        |
| 2.2.6 Quantification of oestrogen metabolites .....                                                                                                                                                                     | 79        |
| 2.2.7 Intra/Inter assay accuracy and precision of the measured oestrogen metabolites ..                                                                                                                                 | 80        |
| 2.2.8 Study population .....                                                                                                                                                                                            | 80        |
| 2.2.9 Statistical analysis.....                                                                                                                                                                                         | 82        |
| 2.3 Results .....                                                                                                                                                                                                       | 85        |
| 2.3.1 HPLC-MS/MS chromatographic profiles of Oestrogen metabolites in quality control and follicular phase premenopausal urine sample.....                                                                              | 85        |
| 2.3.2 Calibration curve and limit of quantitation (LOQ) .....                                                                                                                                                           | 86        |
| 2.3.3 Method validation .....                                                                                                                                                                                           | 87        |
| 2.3.4 Intra-class correlations (ICCs) of measured oestrogen metabolites.....                                                                                                                                            | 89        |
| 2.3.5 Urinary levels of oestrogen metabolites .....                                                                                                                                                                     | 90        |
| 2.3.6 Associations between urinary oestrogen metabolite levels and carriers of the CYP3A7*1C allele (AC genotype) compared to non- carriers (AA genotype) .....                                                         | 92        |
| 2.4 Discussion .....                                                                                                                                                                                                    | 96        |
| 2.5 Strengths and limitations.....                                                                                                                                                                                      | 100       |
| 2.6 Conclusion.....                                                                                                                                                                                                     | 101       |

|                                                                                                                                                                                                                                                                  |                                                                                                                                   |     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------|-----|
| 2.7                                                                                                                                                                                                                                                              | References.....                                                                                                                   | 102 |
| <b>Chapter 3: Targeted mass spectrometry (LC-MS/MS and FIA-MS/MS) based metabolic profiling exploring the biochemical effects of persistent organic pollutants exposure in maternal and umbilical cord serum within the RHEA mother-child cohort study .....</b> |                                                                                                                                   |     |
| 3.1                                                                                                                                                                                                                                                              | Introduction .....                                                                                                                | 107 |
| 3.1.1                                                                                                                                                                                                                                                            | Persistent organic pollutants (POPs).....                                                                                         | 107 |
| 3.1.1.1                                                                                                                                                                                                                                                          | PCBs (Polychlorinated biphenyls).....                                                                                             | 108 |
| 3.1.1.2                                                                                                                                                                                                                                                          | HCB (Hexachlorobenzene).....                                                                                                      | 109 |
| 3.1.1.3                                                                                                                                                                                                                                                          | DDE (Dichlorodiphenyldichloroethylene).....                                                                                       | 110 |
| 3.1.2                                                                                                                                                                                                                                                            | Prenatal exposure to persistent organic pollutants and pregnancy outcomes .....                                                   | 111 |
| 3.1.3                                                                                                                                                                                                                                                            | Metabolic profiling to identify biomarkers of exposure to POPs .....                                                              | 112 |
| 3.1.4                                                                                                                                                                                                                                                            | Biocrates AbsoluteIDQ P180 assay .....                                                                                            | 113 |
| 3.1.5                                                                                                                                                                                                                                                            | Hypothesis and Aim.....                                                                                                           | 114 |
| 3.1.6                                                                                                                                                                                                                                                            | Objectives .....                                                                                                                  | 114 |
| 3.2                                                                                                                                                                                                                                                              | Materials and Methods .....                                                                                                       | 115 |
| 3.2.1                                                                                                                                                                                                                                                            | Materials .....                                                                                                                   | 115 |
| 3.2.2                                                                                                                                                                                                                                                            | Biocrates AbsoluteIDQ p180 assay .....                                                                                            | 115 |
| 3.2.2.1                                                                                                                                                                                                                                                          | Internal standards .....                                                                                                          | 116 |
| 3.2.2.2                                                                                                                                                                                                                                                          | Calibration standards and Quality control samples .....                                                                           | 116 |
| 3.2.2.3                                                                                                                                                                                                                                                          | Zero samples (Phosphate saline buffer) to calculate limit of detection (LOD), derivatisation reagent and extraction solvent ..... | 117 |
| 3.2.2.4                                                                                                                                                                                                                                                          | Sample preparation workflow.....                                                                                                  | 117 |
| 3.2.2.5                                                                                                                                                                                                                                                          | Ultra performance liquid chromatography tandem mass spectrometry (UPLC-MS/MS) method.....                                         | 119 |
| 3.2.2.6                                                                                                                                                                                                                                                          | Flow injection analysis tandem mass spectrometry (FIA-MS/MS) method .....                                                         | 119 |
| 3.2.2.7                                                                                                                                                                                                                                                          | Data processing .....                                                                                                             | 120 |
| 3.2.3                                                                                                                                                                                                                                                            | Study population: the RHEA mother-child cohort in Crete, Greece .....                                                             | 121 |
| 3.2.3.1                                                                                                                                                                                                                                                          | Organochlorine POPs exposure assessment in 1st-trimester maternal serum ...                                                       | 121 |
| 3.2.3.2                                                                                                                                                                                                                                                          | RHEA cohort sample sets analysed.....                                                                                             | 122 |
| 3.2.4                                                                                                                                                                                                                                                            | Quality control assessment criteria for the measured metabolites .....                                                            | 124 |
| 3.2.5                                                                                                                                                                                                                                                            | Statistical analysis.....                                                                                                         | 125 |
| 3.3                                                                                                                                                                                                                                                              | Results .....                                                                                                                     | 126 |
| 3.3.1                                                                                                                                                                                                                                                            | Inter-batch precision of the reliably measured metabolites.....                                                                   | 126 |
| 3.3.2                                                                                                                                                                                                                                                            | Metabolite profile comparison between maternal serum and cord serum .....                                                         | 127 |
| 3.3.3                                                                                                                                                                                                                                                            | Associations between 1st-trimester organochlorine POPs exposure and metabolite profile of 1st-trimester maternal serum .....      | 131 |

|                                                                                                                                                                                                                                   |                                                                                                                |            |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------|------------|
| 3.3.4                                                                                                                                                                                                                             | Associations between 1st-trimester organochlorine POPs exposure and the metabolite profile of cord serum ..... | 134        |
| 3.4                                                                                                                                                                                                                               | Discussion .....                                                                                               | 137        |
| 3.5                                                                                                                                                                                                                               | Strengths and limitations.....                                                                                 | 144        |
| 3.6                                                                                                                                                                                                                               | Conclusion.....                                                                                                | 146        |
| 3.7                                                                                                                                                                                                                               | References.....                                                                                                | 147        |
| <b>Chapter 4. Untargeted LC-MS metabolic profiling to determine sources of analytical and short-term temporal variation in urinary metabolite measurements: application to the Human Early-Life Exposome (HELIX) project.....</b> |                                                                                                                | <b>152</b> |
| 4.1                                                                                                                                                                                                                               | Introduction .....                                                                                             | 152        |
| 4.1.1                                                                                                                                                                                                                             | The Human Early-Life Exposome (HELIX) project.....                                                             | 152        |
| 4.1.2                                                                                                                                                                                                                             | Untargeted metabolic profiling for characterising the human exposome.....                                      | 154        |
| 4.1.3                                                                                                                                                                                                                             | Sources of metabolite level variability .....                                                                  | 155        |
| 4.1.4                                                                                                                                                                                                                             | Aim .....                                                                                                      | 156        |
| 4.1.5                                                                                                                                                                                                                             | Objectives .....                                                                                               | 156        |
| 4.2                                                                                                                                                                                                                               | Materials and Methods .....                                                                                    | 158        |
| 4.2.1                                                                                                                                                                                                                             | Materials .....                                                                                                | 158        |
| 4.2.2                                                                                                                                                                                                                             | Study design .....                                                                                             | 158        |
| 4.2.3                                                                                                                                                                                                                             | LC-MS sample preparation and untargeted metabolic profiling .....                                              | 160        |
| 4.2.4                                                                                                                                                                                                                             | LC-MS metabolite identification.....                                                                           | 162        |
| 4.2.5                                                                                                                                                                                                                             | Data normalisation for analytical and biological variability .....                                             | 164        |
| 4.2.6                                                                                                                                                                                                                             | Temporal biological variability .....                                                                          | 166        |
| 4.3                                                                                                                                                                                                                               | Results .....                                                                                                  | 168        |
| 4.3.1                                                                                                                                                                                                                             | Metabolites identified.....                                                                                    | 168        |
| 4.3.2                                                                                                                                                                                                                             | LC-MS analytical variation.....                                                                                | 169        |
| 4.3.3                                                                                                                                                                                                                             | LC-MS metabolome temporal biological variation.....                                                            | 173        |
| 4.4                                                                                                                                                                                                                               | Discussion .....                                                                                               | 178        |
| 4.5                                                                                                                                                                                                                               | Strengths and limitations.....                                                                                 | 183        |
| 4.6                                                                                                                                                                                                                               | Conclusion .....                                                                                               | 184        |
| 4.7                                                                                                                                                                                                                               | References.....                                                                                                | 185        |
| <b>Chapter 5. Summary and Discussion.....</b>                                                                                                                                                                                     |                                                                                                                | <b>188</b> |
| 5.1                                                                                                                                                                                                                               | References .....                                                                                               | 193        |
| <b>Appendix.....</b>                                                                                                                                                                                                              |                                                                                                                | <b>194</b> |

## List of Figures

|                                                                                                                                                                                                                                                                                                                                                                                                                                     |           |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| <b>FIGURE 1.1 THE COMPLEX BIDIRECTIONAL 'OMICS' STRUCTURE OF SYSTEM BIOLOGY SHOWING INTERACTIONS OF GENOME, TRANSCRIPTOME, PROTEOME AND METABOLOME WITH INTRINSIC ENVIRONMENTAL FACTORS TO PRODUCE A PHENOTYPE.....</b>                                                                                                                                                                                                             | <b>17</b> |
| <b>FIGURE 1.2 (A) MAJOR SOURCES OF EXPOSOME INFLUENCES ON HUMAN METABOLOME AND HUMAN HEALTH. (B) SUB-CATEGORIES OF EXPOSOME EXPLORED IN THIS THESIS. ....</b>                                                                                                                                                                                                                                                                       | <b>23</b> |
| <b>FIGURE 1.3 (A) SCHEMATIC DIAGRAM OF THE PRINCIPLE OF LIQUID CHROMATOGRAPHY-MASS SPECTROMETRY (LC-MS). A HIGH-LEVEL VIEW OF WORKING OF LIQUID CHROMATOGRAPHY-MASS SPECTROMETRY (LC-MS) WHERE ANALYTES IN A SAMPLE ARE FIRST SEPARATED IN THE TIME DOMAIN IN LC AND THEN ACCORDING TO THEIR MASS IN MS. (B) 3-DIMENSIONAL LC-MS DATA OUTPUT DESCRIBING ANALYTE RETENTION TIME, MASS TO CHARGE (M/Z) RATIOS AND INTENSITY. ....</b> | <b>26</b> |
| <b>FIGURE 1.4 THE GRAPHICAL REPRESENTATION OF THE VAN DEEMTER PLOT .....</b>                                                                                                                                                                                                                                                                                                                                                        | <b>29</b> |
| <b>FIGURE 1.5 A SCHEMATIC DIAGRAM OF A TYPICAL TRIPLE QUADRUPOLE (TQ OR QQQ) MASS ANALYSER.....</b>                                                                                                                                                                                                                                                                                                                                 | <b>33</b> |
| <b>FIGURE 1.6 A SCHEMATIC DIAGRAM OF A TYPICAL TIME OF FLIGHT (ToF) MASS ANALYSER .....</b>                                                                                                                                                                                                                                                                                                                                         | <b>34</b> |
| <b>FIGURE 1.7 CALCULATION OF ANALYTE CONCENTRATION IN THE TEST SAMPLE USING EXTERNAL REFERENCE STANDARDS AND INTERNAL CALIBRATION STANDARDISATION CURVE. <math>R_A</math> =PEAK RESPONSE OF ANALYTE, <math>R_{IS}</math> =PEAK RESPONSE OF INTERNAL STANDARDS, <math>C_A</math> =CONCENTRATION OF THE ANALYTE AND <math>C_{IS}</math> =CONCENTRATION OF INTERNAL STANDARDS.....</b>                                                 | <b>39</b> |
| <b>FIGURE 1.8 SCHEMATIC WORKFLOWS OF TARGETED (LC-MS/MS) AND UNTARGETED (LC-MS) METABOLIC PROFILING .....</b>                                                                                                                                                                                                                                                                                                                       | <b>43</b> |
| <b>FIGURE 1.9 SUMMARY OF THREE LC-MS PLATFORMS INCORPORATING METABOLOME ANALYSIS INTO THREE DIFFERENT STUDIES CARRIED OUT IN THIS PhD PROJECT. THIS DIAGRAM ALSO HIGHLIGHTS STRENGTHS AND LIMITATIONS OBSERVED FOR EACH METHOD IN THIS PhD. ....</b>                                                                                                                                                                                | <b>45</b> |
| <b>FIGURE 2.1. PATHWAY OF STEROID HORMONE SYNTHESIS AND ENDOGENOUS OESTROGEN METABOLISM IN HUMANS. 64</b>                                                                                                                                                                                                                                                                                                                           |           |
| <b>FIGURE 2.2 OXIDATIVE METABOLISM SITES IN OESTRADIOL (E2) BY PHASE I CYTOCHROME P450 ENZYMES.....</b>                                                                                                                                                                                                                                                                                                                             | <b>65</b> |
| <b>FIGURE 2.3 OVERALL METHOD SYNOPSIS FOR THE QUANTITATIVE MEASUREMENT OF 14 ENDOGENOUS OESTROGEN METABOLITES IN URINE BY HPLC-ESI-MS/MS. (B) MOLECULAR STRUCTURE EXAMPLE OF OESTRONE-GLUCURONIDE SHOWING HYDROLYSIS AND DERIVATISATION REACTIONS TO ANALYSE OESTRONE-DANSYL.....</b>                                                                                                                                               | <b>76</b> |
| <b>FIGURE 2.4 STUDY DESIGN SHOWING SELECTION AND NUMBER OF SAMPLES ANALYSED IN THIS PhD. N= 729 URINE SAMPLES FROM HEALTHY PREMENOPAUSAL WOMEN IN WHICH OESTRONE GLUCURONIDE (E1G) AND PREGNANEDIOL GLUCURONIDE (PG) WERE MEASURED AND 642 SNPs WERE ANALYSED.....</b>                                                                                                                                                              | <b>82</b> |
| <b>FIGURE 2.5 HPLC-MS/MS MRM CHROMATOGRAPHIC PROFILE OF DANSYLATED EMs. (A) QC SAMPLE AT 2NG/ML AND (B) FOLLICULAR PHASE PREMENOPAUSAL URINE SAMPLE. THE PEAKS CORRESPONDING TO EACH 14 EMs LISTED IN MATERIALS AND METHODS ARE INDICATED WITHIN THE CHROMATOGRAPHIC PROFILES.....</b>                                                                                                                                              | <b>85</b> |
| <b>FIGURE 2.6 INTER-BATCH PRECISION CALCULATED AS A PERCENTAGE COEFFICIENT OF VARIANCE (CV %) IN QC SAMPLES (N=18; CONCENTRATION 2 NG/ML) OF EM MEASUREMENT AFTER HYDROLYSIS, EXTRACTION, DERIVATISATION AND HPLC-MS/MS STEPS.....</b>                                                                                                                                                                                              | <b>88</b> |

|                                                                                                                                                                                                                                                                         |     |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| FIGURE 2.7 HEATMAP OF SPEARMAN CORRELATION COEFFICIENTS FOR INDIVIDUAL EMS AND 2-, 4- AND 16-HYDROXYLATION PATHWAYS, IN WOMEN (N=30) WHO WERE CARRIERS OF THE CYP3A7*1C ALLELE (AC GENOTYPE) AND IN WOMEN (N=30) WHO WERE NON-CARRIERS (AA GENOTYPE) OF RS45446698..... | 94  |
| FIGURE 3.1 MOLECULAR STRUCTURE OF PCB (POLYCHLORINATED BIPHENYLS) .....                                                                                                                                                                                                 | 108 |
| FIGURE 3.2 MOLECULAR STRUCTURE OF HCB (HEXACHLOROBENZENE).....                                                                                                                                                                                                          | 109 |
| FIGURE 3.3 MOLECULAR STRUCTURE OF DDT (DICHLORODIPHENYLTRICHLOROETHANE) & DDE (DICHLORODIPHENYLDICHLOROETHYLENE).....                                                                                                                                                   | 110 |
| FIGURE 3.4 BIOCRATES ABSOLUTEIDQ P180 SAMPLE PREPARATION WORKFLOW.....                                                                                                                                                                                                  | 118 |
| FIGURE 3.5 STUDY DESIGN OF THE 1ST-TRIMESTER MATERNAL SERUM AND UMBILICAL CORD SERUM SAMPLES ANALYSED IN THIS PhD FROM RHEA MOTHER-CHILD COHORT USING BIOCRATES ABSOLUTEIDQ P180 .....                                                                                  | 123 |
| FIGURE 3.6 CLUSTER ANALYSES OF QC SERUM SETS ACCORDING TO THEIR METABOLITE PROFILES (RHEA QCs AND COMMERCIAL SERUM QCs). .....                                                                                                                                          | 127 |
| FIGURE 3.7(A) PRINCIPAL COMPONENTS ANALYSIS (PCA) SCORES PLOT DISCRIMINATING THE METABOLIC PROFILES IN N=67 1ST-TRIMESTER MATERNAL SERUM.....                                                                                                                           | 127 |
| FIGURE 3.7(B) PRINCIPAL COMPONENTS ANALYSIS (PCA) LOADINGS PLOT CORRESPONDING TO THE FIRST TWO PCs OF THE PCA MODEL. ....                                                                                                                                               | 128 |
| FIGURE 3.4 BAR GRAPH COMPARING MEAN CONCENTRATIONS OF THE SUM OF AMINO ACIDS (AA) IN 1ST TRIMESTER MATERNAL AND CORD SERUM AMINO ACIDS. ....                                                                                                                            | 129 |
| FIGURE 3.5 BAR GRAPH COMPARING MEAN CONCENTRATIONS OF SELECTED METABOLITE RATIOS IN THE 1ST TRIMESTER MATERNAL AND CORD SERUM.....                                                                                                                                      | 130 |
| FIGURE 3.6 $\beta$ COEFFICIENTS OF 1ST-TRIMESTER MATERNAL SERUM METABOLITES ASSOCIATED WITH PRENATAL POPs EXPOSURES ( $\Sigma$ PCBs, HCB AND DDE). ....                                                                                                                 | 134 |
| FIGURE 3.7 $\beta$ COEFFICIENTS FOR THE METABOLITES SIGNIFICANT ( $p \leq 0.05$ ) IN ADJUSTED REGRESSION MODELS FOR CORD SERUM IN RELATION TO THE MATERNAL SERUM CONCENTRATION OF $\Sigma$ PCBs, HCB, DDE AND OVERALL SCORE.....                                        | 136 |
| FIGURE 3.8 A SCHEMATIC DIAGRAM OF PHOSPHATIDYLCHOLINE STRUCTURE.....                                                                                                                                                                                                    | 139 |
| FIGURE 4.1 HUMAN EARLY-LIFE EXPOSOME (HELIX) PROJECT CONCEPTUAL FRAMEWORK.....                                                                                                                                                                                          | 153 |
| FIGURE 4.2 CHILDREN URINE (N=20) MORNING VOID AND BEFORE BED NIGHT VOID SAMPLE COLLECTION AND DAILY POOL (50:50 MORNING AND NIGHT URINE) STRATEGY ACROSS SIX DAYS .....                                                                                                 | 159 |
| FIGURE 4.3 SCHEMATIC DIAGRAM SHOWING VARIOUS STEPS FOLLOWED TO IDENTIFY METABOLITES IN UNTARGETED LC-MS METABOLIC PROFILING .....                                                                                                                                       | 163 |
| FIGURE 4.4 SCHEMATIC DIAGRAM SHOWING DATA PRE-TREATMENT STEPS FOLLOWED TO NORMALISE ANALYTICAL AND BIOLOGICAL VARIATIONS.....                                                                                                                                           | 165 |
| FIGURE 4.5 LIST OF METABOLITES ASSIGNED IN POSITIVE AND NEGATIVE IONISATION MODE URINE DATA OF N=20 HEALTHY CHILDREN AND QC SAMPLES.....                                                                                                                                | 168 |

|                                                                                                                                                                                                                                                                                              |     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| FIGURE 4.6 INTER-BATCH PRECISION (CV %) OF THE PEAK AREAS OF THE 17 METABOLITES ASSIGNED IN POSITIVE IONISATION MODE DATA FROM 51 POOLED QCS ACROSS SIX PLATES. ....                                                                                                                         | 169 |
| FIGURE 4.7 INTER-BATCH PRECISION (CV%) OF THE NORMALISED PEAK AREA DATA (QC-RLSC, CREATININE-ADJUSTED AND LOG TRANSFORMED) OF THE 17 METABOLITES ASSIGNED IN POSITIVE IONISATION MODE DATA FROM 51 POOLED QCS ACROSS SIX PLATES.....                                                         | 169 |
| FIGURE 4.8 PCA SCORES PLOT IS SHOWING THE DISTRIBUTION OF 51 POOLED QCS (RED CIRCLES) AMONG 360 CHILDREN URINE SAMPLES (GREY CIRCLES) FROM SIX BATCHES. RAW DATA, I.E. PEAK AREAS OF 17 METABOLITES WERE USED IN THIS ANALYSIS.....                                                          | 170 |
| FIGURE 4.9 PCA SCORES PLOT SHOWING CLUSTERING OF 51 POOLED QCS FROM SIX SEPARATE BATCHES AFTER NORMALISATION. COLOUR CODE REPRESENTS BATCHES 1 TO 6.....                                                                                                                                     | 171 |
| FIGURE 4.10 INTER-BATCH PRECISION (CV%) OF THE RAW PEAK AREAS OF THE 17 METABOLITES IDENTIFIED IN POSITIVE IONISATION MODE DATA FROM 45 POOLED QCS ACROSS FIVE PLATES (BATCH 1, 2,3, 4 AND 6), EXCLUDING QCS FROM BATCH 5. ....                                                              | 172 |
| FIGURE 4.11 INTER-BATCH PRECISION (CV%) OF THE NORMALISED PEAK AREAS OF THE 17 METABOLITES IDENTIFIED IN POSITIVE IONISATION MODE 45 POOLED QCS ACROSS FIVE PLATES (BATCH 1, 2,3, 4 AND 6), EXCLUDING QCS FROM BATCH 5.....                                                                  | 172 |
| FIGURE 4.12 RETENTION TIME TREND OBSERVED FOR TYROSINE IN ALL 51 QC SAMPLES ACROSS SIX BATCHES. ....                                                                                                                                                                                         | 173 |
| FIGURE 4.13 LINE GRAPHS ARE SHOWING INTRA-INDIVIDUAL VARIATION IN LEVELS OF TYROSINE AND TRYPTOPHAN ACROSS SIX DAYS IN N=20 CHILDREN URINE SAMPLES.....                                                                                                                                      | 177 |
| APPENDIX FIGURE 1A BAR GRAPH COMPARING AVERAGE CONCENTRATIONS OF MATERNAL AND CORD SERUM AMINO ACIDS. X-AXIS REPRESENTS INDIVIDUAL AMINO ACID METABOLITES AND Y-AXIS REPRESENTS THEIR MEAN CONCENTRATION IN ( $\mu$ M). ....                                                                 | 198 |
| APPENDIX FIGURE 1B BAR GRAPH COMPARING AVERAGE CONCENTRATIONS OF MATERNAL AND CORD SERUM BIOGENIC AMINES. X-AXIS REPRESENTS INDIVIDUAL BIOGENIC AMINE METABOLITES AND Y-AXIS REPRESENTS THEIR MEAN CONCENTRATION IN ( $\mu$ M). ....                                                         | 198 |
| APPENDIX FIGURE 1C. BAR GRAPH COMPARING AVERAGE CONCENTRATIONS OF MATERNAL AND CORD SERUM LYSO-GLYCEROPHOSPHOCHOLINES (LYSO PC). X-AXIS REPRESENTS INDIVIDUAL LYSO-GLYCEROPHOSPHOCHOLINES METABOLITES AND Y-AXIS REPRESENTS THEIR MEAN CONCENTRATION IN ( $\mu$ M). ....                     | 199 |
| APPENDIX FIGURE 1D & 1E. BAR GRAPH COMPARING AVERAGE CONCENTRATIONS OF MATERNAL AND CORD SERUM DIACYL-GLYCEROPHOSPHOCHOLINES (PC AA) AND PLASMOGENS-GLYCERPHOSHPOCHOLINE (PC AE). X-AXIS REPRESENTS INDIVIDUAL METABOLITES AND Y-AXIS REPRESENTS THEIR MEAN CONCENTRATION IN ( $\mu$ M)..... | 200 |
| APPENDIX FIGURE 1F. BAR GRAPH COMPARING AVERAGE CONCENTRATIONS OF MATERNAL AND CORD SERUM SPHINGOMYELINS (SM). X-AXIS REPRESENTS INDIVIDUAL SPHINGOMYELINS AND HYDROXYL SPHINGOMYELINS (SM-OH) METABOLITES AND Y-AXIS REPRESENTS THEIR MEAN CONCENTRATION IN ( $\mu$ M). ....                | 201 |
| APPENDIX FIGURE 1G. BAR GRAPH COMPARING AVERAGE CONCENTRATIONS OF MATERNAL AND CORD SERUM ACYLCARNITINE. X-AXIS REPRESENTS INDIVIDUAL CARNITINE METABOLITES AND Y-AXIS REPRESENTS THEIR MEAN CONCENTRATION IN ( $\mu$ M). ....                                                               | 201 |

## List of Tables

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |            |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| <b>TABLE 2.1 CONCENTRATION OF DEUTERATED OESTROGEN METABOLITES (D-EM) INTERNAL STANDARD SOLUTION IN ALL THE SAMPLES.....</b>                                                                                                                                                                                                                                                                                                                                                                                             | <b>75</b>  |
| <b>TABLE 2.2 CONCENTRATIONS OF OESTROGEN METABOLITE (EM) STANDARD SOLUTION MIXTURES IN EIGHT CALIBRATION SAMPLES.....</b>                                                                                                                                                                                                                                                                                                                                                                                                | <b>75</b>  |
| <b>TABLE 2.3 GRADIENT ELUTION WITH MOBILE PHASE A (WATER + 0.2% FORMIC ACID) AND MOBILE PHASE A (METHANOL + 0.2% FORMIC ACID), TOTAL RUN TIME 77 MINUTES PER SAMPLE.....</b>                                                                                                                                                                                                                                                                                                                                             | <b>78</b>  |
| <b>TABLE 2.4 MULTIPLE REACTION MONITORING (MRM) TRANSITIONS MONITORED FOR THE PROTONATED [M+H]<sup>+</sup> EM-DANSYL AND D-EM-DANSYL METABOLITE AND THEIR RETENTION TIMES.....</b>                                                                                                                                                                                                                                                                                                                                       | <b>79</b>  |
| <b>TABLE 2.5 INTER-BATCH ACCURACY (RECOVERY OF THE KNOWN AMOUNT OF EM IN PERCENTAGE) AND PRECISION OF EM MEASUREMENT AFTER HYDROLYSIS, EXTRACTION, DERIVATISATION AND HPLC-MS/MS STEPS. THE VALUES SHOWN IN THIS TABLE ARE FROM ONE BATCH AND ARE REPRESENTATIVE OF OTHER FIVE BATCHES.....</b>                                                                                                                                                                                                                          | <b>88</b>  |
| <b>TABLE 2.6 ICCs (95% CONFIDENCE INTERVALS) FOR INDIVIDUAL AND GROUPED EMs, EXPRESSED AS RELATIVE TO CREATININE AND LOG-TRANSFORMED LEVELS IN N=30 WOMEN WITH COMMON (RS45446698-A) ALLELE OR AA GENOTYPE. ....</b>                                                                                                                                                                                                                                                                                                     | <b>89</b>  |
| <b>TABLE 2.7 GEOMETRIC MEAN AND 5<sup>TH</sup>-95<sup>TH</sup> PERCENTILE RANGES OF INDIVIDUAL AND GROUPED URINARY EM EXPRESSED AS RELATIVE (CREATININE-ADJUSTED AND FOLLICULAR DAYS AVERAGED) LEVELS IN N=30 WOMEN WITH COMMON (RS45446698-A) ALLELE OR AA GENOTYPE. SELECTED EM RATIOS ARE ALSO REPORTED. PERCENTAGE OF THE TOTAL EM EXPRESSED AS ARITHMETIC MEAN (CREATININE ADJUSTED, FOLLICULAR DAYS AVERAGED AND LOG<sub>e</sub> TRANSFORMED) AND THEIR 5<sup>TH</sup>-95<sup>TH</sup> PERCENTILE RANGES. ....</b> | <b>91</b>  |
| <b>TABLE 2.8 PERCENTAGE DIFFERENCE IN EM LEVELS IN 30 PREMENOPAUSAL WOMEN WHO WERE CARRIERS OF THE CYP3A7*1C ALLELE (AC GENOTYPE) COMPARED TO WOMEN WHO WERE NON-CARRIERS (AA GENOTYPE) OF RS45446698. THE RESULTS SHOW THE OUTPUT OF LINEAR REGRESSION ANALYSIS WITH AND WITHOUT ADJUSTING FOR COVARIATES** AGE, BMI AND PARITY.....</b>                                                                                                                                                                                | <b>93</b>  |
| <b>TABLE 3.1 LIST OF FIVE CLASSES OF METABOLITES MEASURED USING BIOCRATES ABSOLUTEIDQ p180 ASSAY, IN PARENTHESIS NUMBER OF METABOLITES MEASURED IN EACH CLASS, MODE OF MEASUREMENT (LC OR FIA), QUANTIFICATION (ABSOLUTE OR SEMI) AND APPENDIX LIST OF INDIVIDUAL METABOLITES MEASURED. ....</b>                                                                                                                                                                                                                         | <b>116</b> |
| <b>TABLE 3.2 GRADIENT ELUTION WITH MOBILE PHASE A (WATER + 0.2% FORMIC ACID) AND MOBILE PHASE B (ACETONITRILE + 0.2% FORMIC ACID), TOTAL RUN TIME 4.15 MINUTES PER SAMPLE.....</b>                                                                                                                                                                                                                                                                                                                                       | <b>119</b> |
| <b>TABLE 3.3 FIA-MS/MS ISOCRATIC ELUTION METHOD .....</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                | <b>120</b> |
| <b>TABLE 3.4 INTER-BATCH PRECISION OF RHEA QCs AVERAGE CV% ACROSS FIVE BIOCRATES P180 PLATES (EXCLUDING METABOLITES BELOW LOD IN MORE THAN 30% OF SAMPLES AND METABOLITES WITH INTER-DAY CV% &gt;20%.....</b>                                                                                                                                                                                                                                                                                                            | <b>126</b> |
| <b>TABLE 3.5 UNADJUSTED AND ADJUSTED MULTIPLE LINEAR REGRESSION MODELS FOR 1ST-TRIMESTER MATERNAL SERUM POPs (ΣPCBs, HCB AND DDE) AND A METABOLITE OF INTEREST IN 1ST-TRIMESTER MATERNAL SERUM IN N=67 WOMEN WITH HEALTHY PREGNANCY FROM THE RHEA PREGNANCY COHORT. ....</b>                                                                                                                                                                                                                                             | <b>133</b> |
| <b>TABLE 3.6 UNADJUSTED AND ADJUSTED MULTIPLE LINEAR REGRESSION MODELS FOR CORD SERUM METABOLITES IN RELATION TO MATERNAL SERUM CONCENTRATIONS OF ΣPCBs, HCB AND DDE.....</b>                                                                                                                                                                                                                                                                                                                                            | <b>135</b> |

|                                                                                                                                                                                                                                                                                                                    |            |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| <b>TABLE 4.1 GRADIENT ELUTION WITH MOBILE PHASE A (WATER + 0.1% FORMIC ACID) AND MOBILE PHASE B (ACETONITRILE + 0.1% FORMIC ACID), TOTAL RUN TIME 12 MINUTES PER SAMPLE (INCLUDING 2 MINUTES RE-EQUILIBRATION TIME).....</b>                                                                                       | <b>161</b> |
| <b>TABLE 4.2 ICCs (95% CONFIDENCE INTERVAL) FOR 17 METABOLITES IN THE MORNING, NIGHT AND DAILY POOL SAMPLES OF 9 CHILDREN. THE WITHIN AND BETWEEN VALUES REPRESENT WITHIN THE INDIVIDUAL AND BETWEEN INDIVIDUAL VARIANCES. ....</b>                                                                                | <b>175</b> |
| <b>APPENDIX TABLE 1. LIST OF METABOLITES MEASURED USING BIOCRATES P180 LC-MS/MS ASSAY .....</b>                                                                                                                                                                                                                    | <b>194</b> |
| <b>APPENDIX TABLE 2. LIST OF METABOLITES MEASURED USING BIOCRATES P180 FIA-MS/MS ASSAY .....</b>                                                                                                                                                                                                                   | <b>194</b> |
| <b>APPENDIX TABLE 3. LIST OF 28 RATIOS OR SUMS OF SELECTED METABOLITES WITH SIGNIFICANCE OR PROXY TO PROVIDE DETAILED INSIGHT INTO A WIDE RANGE OF DISORDERS OF THE METABOLIC DISEASES.....</b>                                                                                                                    | <b>196</b> |
| <b>APPENDIX TABLE 4 SPEARMAN CORRELATION RHO VALUES AND P VALUES REPORTED, METABOLITES SIGNIFICANTLY ASSOCIATED (<math>P \leq 0.05</math>) WITH DIFFERENT POPs (<math>\Sigma</math>PCBs, HCB AND DDE) IN 1ST TRIMESTER MATERNAL SERUM IN N=67 WOMEN WITH HEALTHY PREGNANCY FROM THE RHEA PREGNANCY COHORT.....</b> | <b>202</b> |
| <b>APPENDIX TABLE 5. SPEARMAN CORRELATION RHO VALUES AND P VALUES REPORTED FOR METABOLITES SIGNIFICANTLY (<math>P \leq 0.05</math>) ASSOCIATED WITH DIFFERENT POPs (<math>\Sigma</math>PCBs, HCB AND DDE) IN CORD SERUM IN N=86 NEONATES WITH HEALTHY BIRTH OUTCOME FROM THE RHEA PREGNANCY COHORT.....</b>        | <b>203</b> |
| <b>APPENDIX TABLE 6. LIST OF 120 CHEMICAL STANDARDS DETECTED .....</b>                                                                                                                                                                                                                                             | <b>204</b> |

## Abbreviations

|                       |                                                  |
|-----------------------|--------------------------------------------------|
| <b>AA</b>             | Arachidonic acid                                 |
| <b>AhR</b>            | Aryl hydrocarbon receptor                        |
| <b>C<sub>18</sub></b> | Octadecyl carbon chains                          |
| <b>CID</b>            | Collision-induced dissociation                   |
| <b>CQC</b>            | Commercial quality control                       |
| <b>CYP</b>            | Cytochrome                                       |
| <b>DDE</b>            | Dichlorophenyldichloroethylene                   |
| <b>DHA</b>            | Docosahexaenoic acid                             |
| <b>DI</b>             | Direct infusion                                  |
| <b>EI</b>             | Electron impact ionisation                       |
| <b>EMs</b>            | Oestrogen metabolites                            |
| <b>ER</b>             | Oestrogen receptors                              |
| <b>ESI</b>            | Electrospray ionisation                          |
| <b>FIA- MS/MS</b>     | Flow injection analysis tandem mass spectrometry |
| <b>GC-MS</b>          | Gas chromatography-mass spectrometry             |
| <b>GWAS</b>           | Genome-wide association study                    |
| <b>HCB</b>            | Hexachlorobenzene                                |
| <b>HELIX</b>          | Human Early Life Exposome                        |
| <b>HPLC</b>           | High performance liquid chromatography           |
| <b>HILIC</b>          | Hydrophilic interaction liquid chromatography    |
| <b>ICC</b>            | Intraclass correlation                           |
| <b>IUGR</b>           | Intrauterine growth restriction                  |
| <b>LC-MS</b>          | Liquid chromatography-mass spectrometry          |
| <b>LC-MS/MS</b>       | Liquid chromatography-tandem mass spectrometry   |
| <b>LOD</b>            | Limit of detection                               |
| <b>LOQ</b>            | Limit of quantification                          |
| <b>MRM</b>            | Multiple reaction monitoring                     |
| <b>MS</b>             | Mass spectrometry                                |
| <b>MS/MS</b>          | Tandem mass spectrometry                         |
| <b>m/z</b>            | mass to charge ratio                             |
| <b>NMR</b>            | Nuclear magnetic resonance spectroscopy          |
| <b>PC</b>             | Principal Component                              |
| <b>PC aa</b>          | Phosphatidylcholines diacyl                      |
| <b>PC ae</b>          | Phosphatidylcholine acyl alkyl                   |
| <b>PCA</b>            | Principal Component Analysis                     |
| <b>PCB</b>            | Polychlorinated biphenyls                        |
| <b>POPs</b>           | Persistent organic pollutants                    |
| <b>QC</b>             | Quality control                                  |
| <b>QQQ or TQ</b>      | Triple quadrupole                                |
| <b>Q-ToF</b>          | Quadrupole-Time of Flight                        |
| <b>RPC</b>            | Reversed phase chromatography                    |
| <b>RQC</b>            | RHEA quality control                             |
| <b>SGA</b>            | Small for gestational age                        |
| <b>SMs</b>            | Sphingomyelins                                   |

|             |                                         |
|-------------|-----------------------------------------|
| <b>SNP</b>  | Single nucleotide polymorphism          |
| <b>S/N</b>  | signal to noise ratio                   |
| <b>ToF</b>  | Time of flight                          |
| <b>UPLC</b> | Ultra-performance liquid chromatography |

# Chapter 1. General Introduction

## 1.1 Metabolic Profiling

Since the completion of the human genome project, there has been a paradigm shift towards a more holistic approach (systems biology) in modern biochemistry to understand the complete system (from cell or tissue to an organism) and how it gives rise to function. The interaction of diet and environmental factors with variations at the level of the genome, proteome, transcriptome, metabolome are complex and dynamic producing a phenotype (a functional endpoint) that can be measured by metabolic profiling [1, 2]. This is summarised in Figure 1.1 and shows how metabolome fits into the structure of system biology.

The metabolome represents the terminal downstream product of the genome and exposure. It consists of the quantitative complement of low-molecular-weight organic and inorganic chemical molecules (metabolites) within a cell, tissue, organelle, biofluid or an entire organism [1]. Metabolites are functional entities, reactants, intermediates or products of enzyme-mediated biochemical reactions and have a clear role in growth, maintenance or in the normal functioning of an organism in a specific physiological state [2].

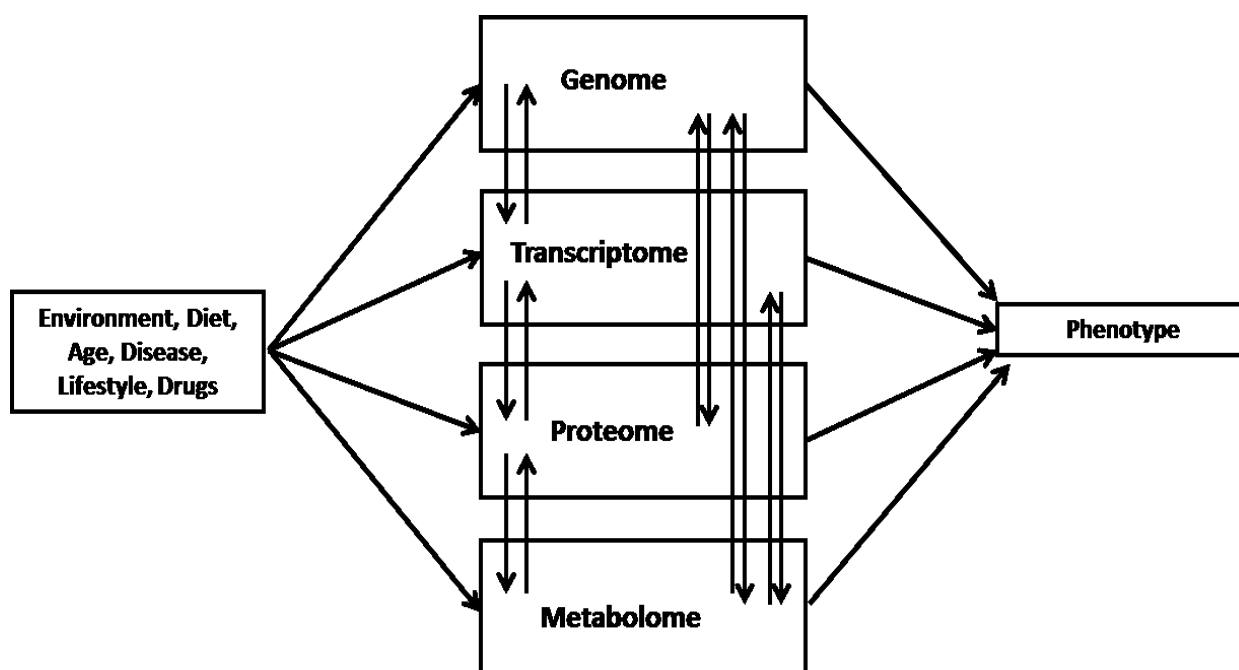


**Figure 1.1 The complex bidirectional 'omics' structure of system biology showing interactions of genome, transcriptome, proteome and metabolome with intrinsic environmental factors to produce a phenotype. Figure reconstructed from Ref 2.**

The concept that the status of an individual's health (including the concepts of homeostasis and allostasis) and disease can reflect/result in changes in the concentration of metabolites in biological fluids related to a genetic or environmental perturbation is central to the field of

metabonomics [3]. Metabonomics is a 'top-down' systems biology approach to study the metabolome and has been defined by Nicholson *et al.* [4] as: '*The quantitative measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification*'. This approach and definition is closely related to that of 'metabolomics', and the terms are often used interchangeably [3]. Both metabonomics and metabolomics are widely applied in disease diagnostics [3, 5, 6] and are generally referred to as a 'metabolic profiling' approach, a term that will be used in this thesis.

The metabolome is considered as a close reflection of a phenotype, and metabolic profiling of biological fluids such as urine and serum holds the potential to define the metabolic phenotype. The term metabolic phenotype has been described as metabolite products reflecting the interaction between several factors such as individual's genome, diet, environmental exposures and gut microbiome [7, 8]. Analysis of the changes in the concentrations/levels of metabolites consequently can lead to disease diagnosis and prognosis at the individual level or be used as a risk factor at the population level [7, 8]. In human studies, metabolic profiling has been successfully applied for the identification of potential diagnostic biomarkers (change in metabolite levels as a consequence of the disease) for cardiovascular disease [2, 10, 11]. It has also been used to investigate the effects of pharmaceutical interventions to predict drug efficacy [2, 12, 13]. In the field of inborn errors of metabolism, an extensive repertoire of clinically applicable metabolite biomarkers are currently being used as diagnostic and prognostic biomarkers [14]. In clinical chemistry, a small number of metabolites such as a glucose test for diabetes, cholesterol for cardiovascular health and a creatinine test for normal kidney function are used routinely as biomarkers to assess an individual's health.

However, at present, there are still significant challenges in metabolic profiling for the characterisation of clinically accepted and useful metabolic phenotypes. The metabolome has several unique characteristics making metabolic profiling demanding and challenging. In contrast to the genome, the metabolome is highly dynamic and has a high chemical diversity. Also, the wide physiological range (concentration ranges at which metabolites occur in *in vivo*), the size of the human metabolome, temporal resolution (kinetics e.g. rate at which metabolites degrade over time, their half-life) and compartmentation of metabolome are all major issues in biological systems [15, 16]. Another major challenge is to understand the influence of exogenous input, i.e. environmental (non-genetic) factors in relation metabolite concentrations and its subsequent effect on health. All the above-mentioned factors contributing towards metabolome complexity are discussed in detail below.

### **1.1.1 Size and diversity of the metabolome**

Metabolic profiling of metabolites within the mass range of 50 to >1500 Daltons encompasses immense chemical diversity. According to Human Metabolome Database (HMDB) [17], there are currently approximately 5000 different classes and subclasses of metabolites (~42000 metabolites) reported in human urine and blood samples. These include the most commonly studied endogenous metabolite classes such as amino acids, amines, steroid hormones, nucleotides, antioxidants, organic acids, vitamins and lipids. The most common exogenous metabolite classes include pharmaceutical drugs (~1600 compounds reported in DrugBank 4.0 database) [18], persistent organic pollutants, exogenous compounds originating from diet (nutrients), microbiota, and other exogenous xenobiotics [17]. Within each class metabolites have different physiochemical properties including polarity (polar/non-polar), solubility or reactivity. The most common and frequently applied methods in studying these metabolites is according to their chemical core structure (e.g. fatty acids), by presence in the same metabolic pathway (e.g. glycolysis) or both (e.g. oestrogens). The biological function of each metabolite can change drastically with a simple exchange of its functional group.

### **1.1.2 Physiological Range and dynamics of the metabolome**

The normal occurring physiological concentrations of metabolites in urine or blood can easily range over ten to eleven orders of magnitude e.g. steroid hormones (pg/mL) to sugars, lipids, glucose, urea (mg/mL) [19]. The physiological ranges of specific metabolites can further increase or decrease depending on the metabolite class, individual's age, gender, and diet. Some of the steroids hormones are known to vary across 10 to 100-fold in concentration among individuals. Detecting and identifying metabolites with 100 to 1000-fold differences in concentration makes defining the metabolome very demanding.

As metabolic processes vary over certain time periods, this significantly influences metabolite concentrations and their measurements over time. For example, the half-life for the thyroid hormone T3 is up 5 to 7 days, and T4 is 1-day whereas the half-life of catecholamines in blood is a few minutes [20, 21]. Consequently, any measurements taken over different time courses can vary and introduce reproducibility issues within studies.

Also, the total size of a metabolome is hard to estimate using any one analytical platform. According to 'Human urine' and 'Human serum' metabolome databases there are approximately 3000 and 4000 confirmed metabolites respectively measured or detected using combination of following analytical platforms: Liquid chromatography-mass spectrometry (LC-MS), Nuclear magnetic resonance spectroscopy (NMR), Gas chromatography-mass spectrometry (GC-MS),

Direct flow injection tandem mass spectrometry (DFI-MS/MS) and High performance liquid chromatography (HPLC) [22, 23]. These databases are not completely inclusive and do not include sufficient information to comprehensively retrieve all the metabolites that define biological samples [24].

### 1.1.3 Genetic and environmental exposure and health

It is now widely accepted that individual's health status is determined by both the genetic and the environmental exposures. Various genome-wide association studies (GWAS) have shown how candidate gene approach is useful in determining genetic variants of chronic diseases. However, these genetic variants only explain or account for a small proportion of the multifactorial chronic diseases and provide a limited understanding of disease causality [25]. It is now shown that nearly 90% of the multifactorial chronic diseases are influenced and explained by 'non-genetic' environmental factors (including diet, lifestyle, occupation, pharmaceutical drugs, behavioural, environmental exposures etc.) [26, 27, 28, 29]. Historically the major challenges in establishing a link between environmental exposures and disease risk in population studies have been that the environment is poorly characterised regarding accurate definition and measurement of spatial, temporal and intra-/inter- individual variations of the exposures [30]. Exposure misclassification may positively or negatively bias the prediction of relative risk of the disease depending on geographical, environmental, behavioural, socioeconomic, genetic factors etc.; furthermore, traditional research in this field has mainly focused only on few pollutants chosen *a priori*.

To address these limitations, the concept of 'exposome' was first developed by Wild [31] in 2005 which '*at its most complete... encompasses life-course environmental exposures (including lifestyle factors) from the point of conception forwards*'. It has also been suggested to be the environmental (non-genetic) equivalent of the human genome. The concept has later been expanded, in the context of a holistic approach to epidemiological studies by Wild [32] in 2012, and to include all detectable chemicals in a given human biofluid by Rappaport and Smith 2010 [33]. In 2013 Miller and Jones [34] further refined the definition of exposome as '*The cumulative measures of environmental influences and associated biological responses throughout the lifespan, including exposures from environment, diet, behaviour and endogenous process*' to differentiate between exposures and their associated biological responses.

The main purpose of exposome research is to identify environmental risk factors which may increase overall burden of diseases from small group of individuals to larger population. Thus, exposome research (in epidemiological studies) would subsequently also help to develop appropriate public health policies [32, 35]. The exposome approach is different from traditional exposure estimation in two ways. Firstly, it shifts the focus from an insufficient selection of

chemical pollutants towards the totality of wide range of exposures, e.g. diet, geography, behaviour, lifestyle, occupational/environmental exposures etc. across different longitudinal timeframes. Second is it incorporates the 'omics' platforms to understand the effect of exposures on individuals' biological profiles.

The application of exposome broadly falls under three overlapping non-genetic exposures categories of which two are exogenous, and one is endogenously contributing to the disease risk. The first one is specific external exposures which are historically the main focus of epidemiological, environmental risk factors studies including environmental pollutants, radiation, chemical contaminants, pharmaceutical drugs, smoking, alcohol, and occupational exposures. The second is general external exposures such as education, socioeconomic status, climate, and psychological status etc. Thirdly are the internal conditions or endogenous exposure of the body which may have a negative effect on the cellular environment of the individual such as imbalanced hormone metabolism, inflammation, oxidative stress, lipid peroxidation, ageing and gut microflora [32]. It has been hypothesised that the three categories of exposome described above share a qualitative and quantitative association with the biological profiles disclosed by 'omics' techniques including metabonomics/metabolomics [32, 33, 36].

It is worth noting that human metabolome encompasses chemical metabolites originating from diverse biological responses and origin and may not be simply designated as 'endogenous', 'exogenous' and 'xenobiotic' chemical entities. This concept is addressed by Nicholson and Wilson [37] to describe the idea of 'continuum of metabolism'. It explains that origin of many metabolites or metabolic pathways may be linked to both genetics and wider environmental inputs. For example, hippuric acid excreted in urine is derived from liver conjugation of glycine to benzoic acid, so from this perspective is 'endogenous'. However, the benzoate can be produced by gut microbial processing of plant phenolics (and 'exogenous' source) or from xenobiotics. Therefore a more integrated view of metabolism is required.

Metabonomics/metabolomics is a vital part of exposome research as it provides a means to link the full diversity and dynamics of the metabolome with the genetic and multifaceted environment of an individual [38]. Analysis of human metabolome provides information on the metabolic responses to specific exposures which could subsequently help in predicting disease risk. On the other hand, the exposome provides a useful framework for metabonomics studies as we currently lack a detailed understanding of different sources of variability for metabolic phenotypes [38]. Metabolic profiling within large epidemiological studies will allow the construction of robust models that to define the usual/normal level of metabolic phenotype variability in relation to environmental exposures.

Incorporating metabolic profiling in environmental exposure studies (or exposome studies) has particular utility in the molecular epidemiology research. The concept of molecular epidemiology was formalised in 1993 by Schulte and Perera in "Molecular Epidemiology: Principles and Practice" [39]. This field establishes a framework to integrate molecular biology into epidemiological research to encompass utilisation of biomarkers from both genetic and environmental risk [40]. Molecular epidemiology holds the potential to improve the understanding of mechanisms of disease by identifying molecular biomarkers and genes that are the causation of developing diseases in populations [40]. In the last decade, molecular epidemiology approach has flourished taking advantage of technological advances made in 'omics' platforms and their usage in epidemiological studies. Most published work in the field of molecular epidemiology studies has focused on cancer. Several molecular epidemiology cancer studies have reported *KRAS* mutations for colon, and colorectal cancers in population studies [40, 41, 42], and other studies have reported on how specific lifestyle exposures, e.g. smoking, environmental carcinogens and diet effect genetic variants in *KRAS* tumours [40, 43, 44].

The conceptual framework 'Meet in the middle' approach has been proposed by Vineis [45] to determine the associations between exposure, response and disease outcome in molecular epidemiology studies. According to this concept, the link between exposure and disease can be strengthened by identifying specific exposure biomarkers in samples from prospective studies (before the disease onset) which also correlate (predict) disease outcome. This can also improve understanding of the exposure-disease pathway and consequently help to design better prevention methods. This approach was used to provide proof of concept evidence for the existence of intermediate biomarkers linking protective dietary fibre exposure and colon cancer in a prospectively collected European Prospective Investigation into Cancer and Nutrition (EPIC) cohort an epidemiological-metabolic profiling study within EU FP7 EnviroGenomarkers project [46].

However, the application of metabolic profiling in molecular epidemiology and exposome studies is fraught with its own set of challenges. Large-scale studies studying relatively subtle effects in the normal population are potentially more susceptible to the detrimental effects of pre-analytical factors, (lack of) standardisation of assays between laboratories, limited coverage of the metabolome, poor quantitative accuracy, lack of longitudinal data that describes 'normal/usual' levels of metabolic phenotype variability, rate of throughput and cost-effectiveness. Considering these challenges, one of the aims of this PhD thesis is to examine different strategies for the characterisation of metabolic phenotypes and identification of metabolic biomarkers in exposome and molecular epidemiological studies.

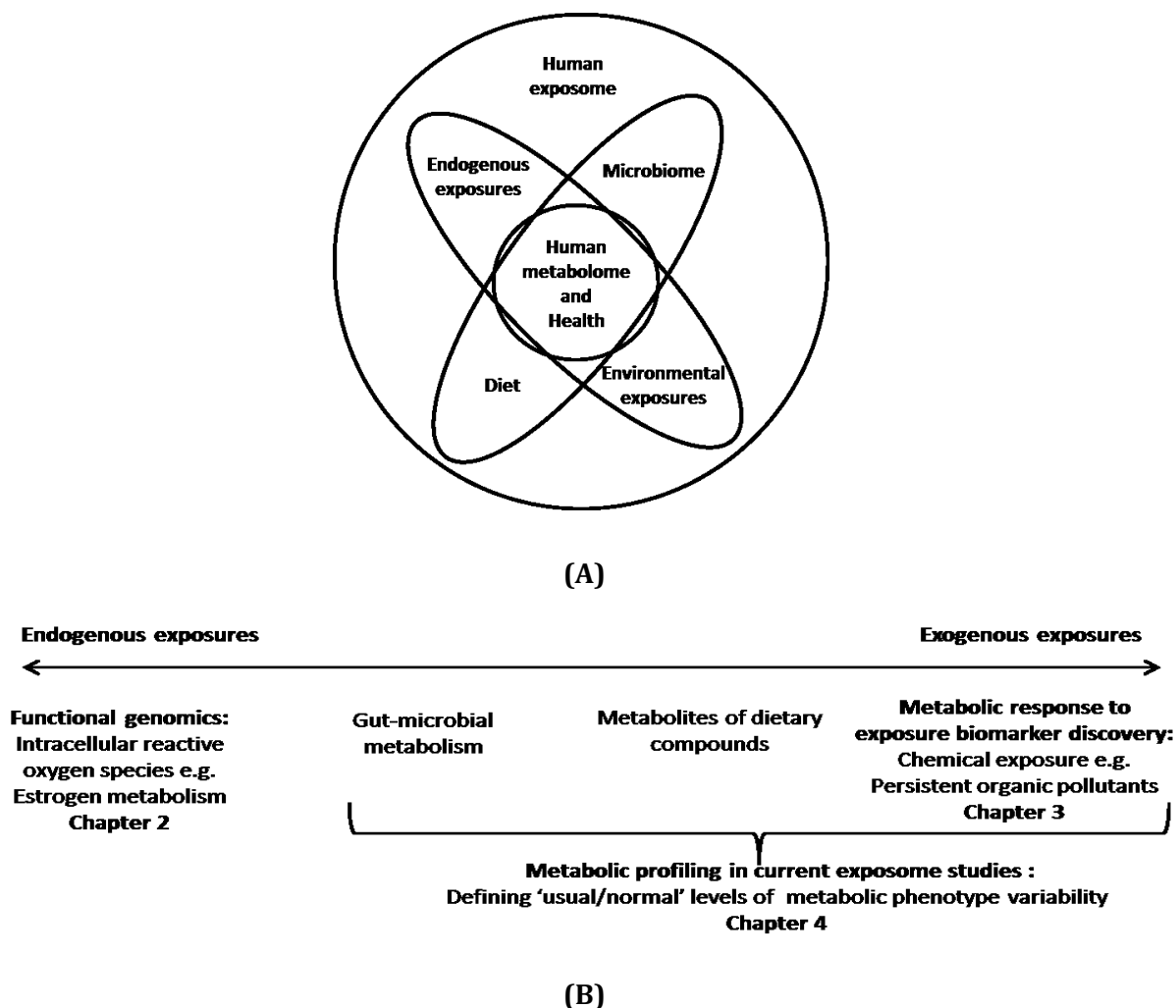


**Figure 1.2 (A) Major sources of exposome influences on human metabolome and human health. (B) Sub-categories of exposome explored in this thesis.**

Figure 1.2(A) summarises major sources of exposome influences on human health and the human metabolome. Figure 1.2(B) summarises three sub-categories of the exposome which are explored in this thesis using liquid chromatography-mass spectrometry (LC-MS) based metabolic profiling. These include endogenous exposure, via a functional genomics study demonstrating how the intracellular concentration of oestrogen metabolites can reveal metabolic phenotypes for genetic variant identified as a risk for breast cancer. The second is exogenous chemical exposure, through a biomarker discovery study to identify metabolic response to persistent organic pollutants in mother-child cohort. The third is a holistic exposome study to analyse temporal variability in the metabolic phenotype of healthy children.

## 1.2 Technologies in Metabolic Profiling

All the characteristics discussed above makes it impossible to define a metabolome by using any one analytical platform. A variety of technologies are used, each offering different advantages and disadvantages depending on the biological question to be answered, experimental design and

metabolites of interest to be measured. However, the two principal technologies used for metabolic profiling are Mass Spectrometry (MS) and Nuclear Magnetic Resonance spectroscopy (NMR) [47, 48].

MS and NMR are complementary technologies as they measure a different subset of the metabolome, uncovering different metabolite classes. Both MS and NMR metabolic profiling has been used and appraised in the context of large population-based studies [49, 50]. They are also well suited to meet the requirements for some of the challenges in large population studies such as the need for high throughput analysis, the cost associated with these, and sample volume availability. NMR possesses the advantages of high reproducibility, robustness, a non-destructive nature with minimal sample preparation and can provide identification and quantification of 50-150 of the most abundant metabolites in biofluids [51, 52, 53]. In contrast to NMR, the main advantages of MS profiling are its sensitivity, possessing the ability to detect metabolites in the femtomolar concentration range in a biological sample [47, 48]. However, MS is also prone to greater variability such as ion suppression thus constraining the reproducibility of studies. Also, MS often cannot distinguish between isomers of a compound with same  $m/z$  ratio, e.g. geometric and optical isomers, and additional chromatographic separation and/or MS/MS fragmentation is required.

To separate individual metabolites from a complex biofluid, MS is often combined with chromatographic techniques such as Liquid Chromatography (LC) and Gas Chromatography (GC). When MS is paired with LC or GC, a 3-dimensional dataset is generated with axes of intensity (ion count), retention time and mass to charge ( $m/z$ ) ratio as shown in Figure 1.3(A & B). In the case of NMR, the resulting datasets are typically 2-dimensional where metabolite signals are reported on a frequency axis of parts per million (ppm) versus intensity [51]. The widespread availability of MS platforms, which are lower in cost to establish than NMR, has made it a popular technique for metabolic profiling.

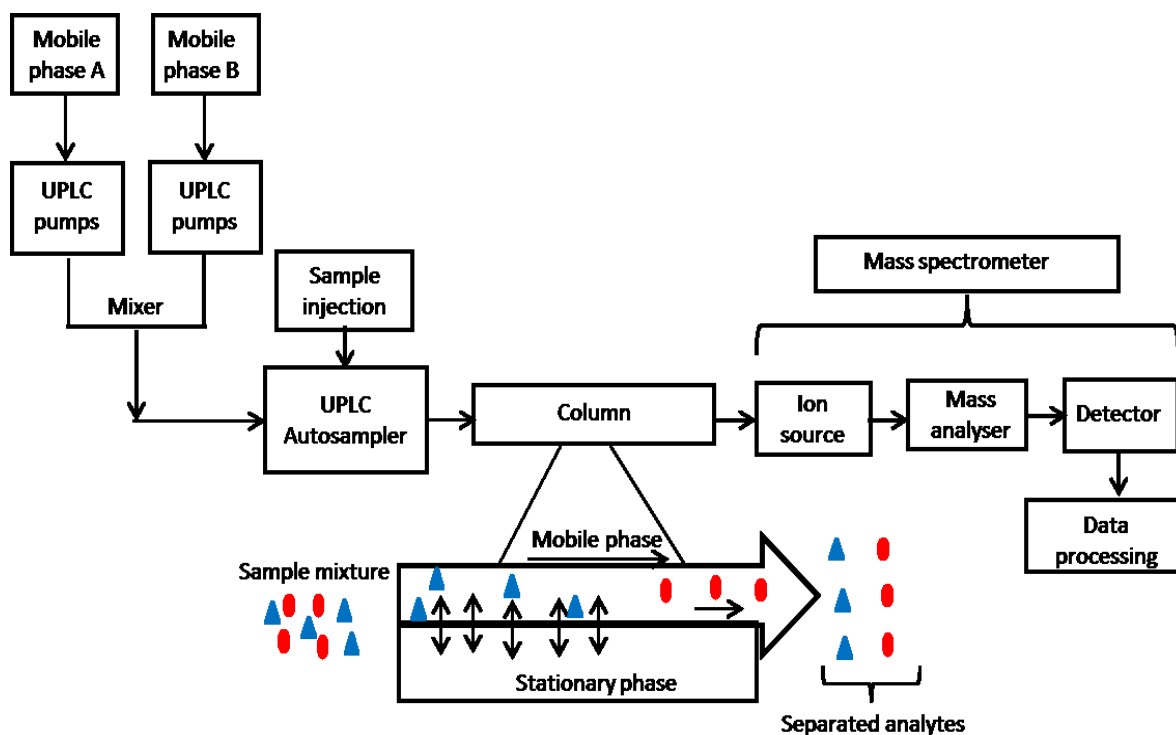
In practice, the choice between Liquid chromatography coupled with mass spectrometry (LC-MS) and Gas Chromatography coupled with mass spectrometry (GC-MS) is influenced by the chemical nature of the metabolites of interest, sample matrix and cost. GC-MS is a preferred technique for less labile and more volatile metabolites (e.g. lipids, flavonoids, carotenoids, fatty acids). LC-MS can analyse much wider pool of metabolites which are not volatile (e.g. organic acids, organic amines, nucleotides, polyamines, amino acids, alcohols, alkaloids). Electron impact ionisation the most common ionisation for GC-MS and is considered very reproducible, suffering less from ion suppression. Electrospray ionisation (ESI) the most common ionisation technique for LC-MS, suffers from ion suppression which may underestimate or fail to detect co-eluting analytes.

Despite this, the use of GC-MS for the discovery of unknown metabolites has largely been superseded by LC-MS due in part to complex sample preparation which usually involves derivatisation.

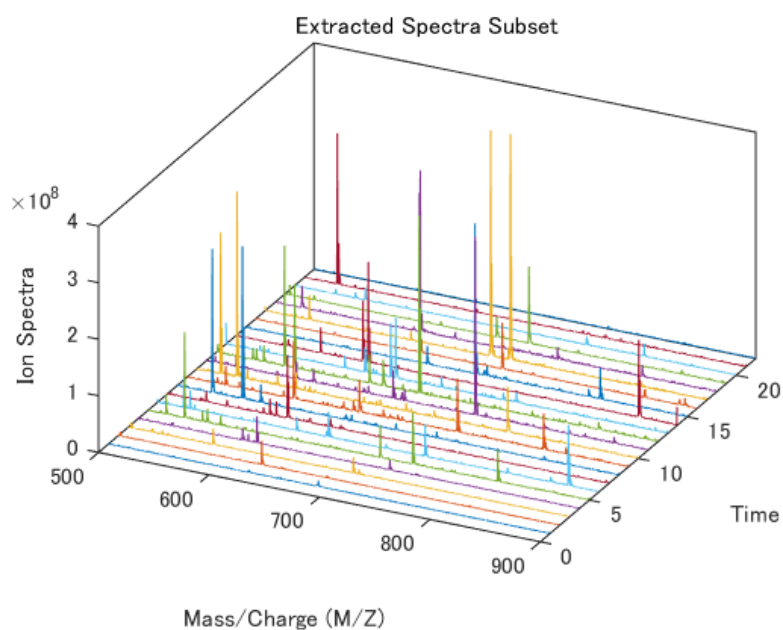
### **1.2.1 Liquid chromatography (LC)**

Liquid chromatography is an important step before mass spectrometry. It helps to reduce the possibility of ion suppression effects by restricting the number of analytes which are ionised simultaneously [54, 55]. Ion suppression in MS is known to negatively affect accuracy, precision and detection response of low abundance metabolites due to competition for ionisation efficiency between the metabolites of interest and other endogenous components present in the matrix (e.g. salts, proteins, lipids) [54]. Furthermore, using LC helps to separate the molecular species which exist in complex biofluid matrixes that are isomers. Therefore, combining LC with MS (LC-MS) helps to make metabolome data more accurate and specific.

The principle of column-based LC is summarised in Figure 1.3 (A), and Figure 1.3 (B) shows 3-dimensional data acquired using LC-MS. The chromatographic separation is achieved by the differential affinity between the stationary phase (static) which is packed in a cylindrical column made using fused silica or polymer coated inside the column and the mobile phase (moving). After injecting the sample into the LC column, the analytes from the biological sample gets mixed with the mobile phase and interacts to a varying degree with the stationary phase also known as column media or resin.



(A)



(B)

Figure 1.3 (A) Schematic diagram of the principle of liquid chromatography-Mass Spectrometry (LC-MS). A high-level view of working of liquid chromatography-mass spectrometry (LC-MS) where analytes in a sample are first separated in the time domain in LC and then according to their mass in MS. (B) 3-dimensional LC-MS data output describing analyte retention time, mass to charge ( $m/z$ ) ratios and intensity. (B) taken from MATLAB plot gallery.

The analytes in the mobile phase interact or bind strongly to the stationary phase if they have similar physiochemical characteristics (e.g. polarity, charge, molecular size, hydrophobicity, specific binding affinity or combination of all) and thereby their passage through the column is delayed or prevented (retention). Components with similar physiochemical characteristics as the mobile phase will move faster through the column and the separation is achieved. Therefore, elution of an individual molecular entity as a peak from the column is a result of chemical selectivity between molecules travelling through the column, interacting with the stationary phase and the mobile phase. By using an isocratic elution (constant mobile phase composition) or a gradient elution (where the composition of the mobile phases is changed to alter the strength of the interaction of analyte of interest between the mobile phase and stationary phase [58]), analytes are separated then analysed by the MS.

### 1.2.1.1 Theory of chromatography

Plate theory (older; proposed by Martin & Synge in 1941) and Rate theory (currently used, developed by van Deemter in 1956) are the main two theories that apply to chromatography.

Plate theory proposes that the chromatographic column contains a large number of adjacent separate imaginary layers called theoretical plates (N). The analyte establishes separate equilibrium between stationary and mobile phase at each theoretical plate. The motion of the analyte down the column happens by movement of the equilibrated mobile phase (assuming equilibration is very fast) from one plate to another. According to plate theory greater separation of peaks occurs when [59, 60]:

- Number (N) of theoretical plates is large
- Plate height (H) or height equivalent to theoretical plates (HETP) becomes smaller

HETP can be calculated as follows:

$$HETP = \frac{L}{N}$$

L= length of the column

Half height method represented by the following formula is used to calculate the N for the real column after examining an eluted peak. Depending upon the solutes, retention time and peak width at half height different plates numbers are calculated for the columns.

$$N = 5.55 * t_R^2 / w_{1/2}^2$$

$t_R^2$  = retention time;  $w_{1/2}$  = peak width at half height

Rate theory, on the other hand, is based on the time taken by the analyte to equilibrate between the stationary and the mobile phase. Rate theory follows the van Deemter equation for plate height, which is considered most appropriate to explain the process at work inside the column [59, 60]. The van Deemter equation describes how the main factors (termed in the van Deemter equation as A, B, & C) contributes toward a phenomenon called band broadening (loss of peak efficiency observed as wider analytical band) are related with peak height and the average velocity of the mobile phase [59, 60, 61].

The Van Deemter is mathematically expressed as:

$$H_{TEP} = A + B/v + Cv$$

Where,  $v$  = linear velocity of the mobile phase flowing through the column

A = Eddy diffusion is the first contributing factor towards band broadening. It reflects the quality of the column packing (column with small or big particle size, shape, pore structure and column diameter). As analytes in the solute will travel unequal distances at random depending on the size of the stationary particles, analytes take a different path of different lengths leading to band broadening. The effect of Eddy diffusion can be minimised by using well-packed columns with smaller and uniformly packed stationary phase particles.

B = Longitudinal diffusion is the diffusion of a concentrated analyte molecule in a solute from the centre to its edges causing band broadening. The adverse effect of longitudinal diffusion is more pronounced at the low-velocity mobile phase. Hence, using mobile phase at higher flow rates with narrower column will reduce band broadening caused by longitudinal diffusion. Other factors to affect this diffusion are mobile phase viscosity and temperature.

C = Resistance against mass transfer is the equilibration of the analyte between the stationary phase and the mobile phase. Since it takes some time to restore continuous equilibrium, there is always a shift in the concentration of the analyte between stationary and mobile phase. In this diffusion increase in the velocity of the mobile phase flowing through the column contributes towards band broadening. Using lower flow rate of mobile phase (antagonising longitudinal diffusion) and using smaller diameter stationary phase particles minimises the band broadening caused by mass transfer.

The graphical representation of van Deemter equation (H- $v$  plot) showing the relationship between plate height as a function of mobile phase velocity is illustrated in Figure 1.4. The graph shows that while Eddy diffusion (A) and Longitudinal diffusion (B) has no to negligible effect on

the velocity of the mobile phase, under normal operating conditions. The resistance against mass transfer has a considerable effect on the linearity of the mobile phase velocity.

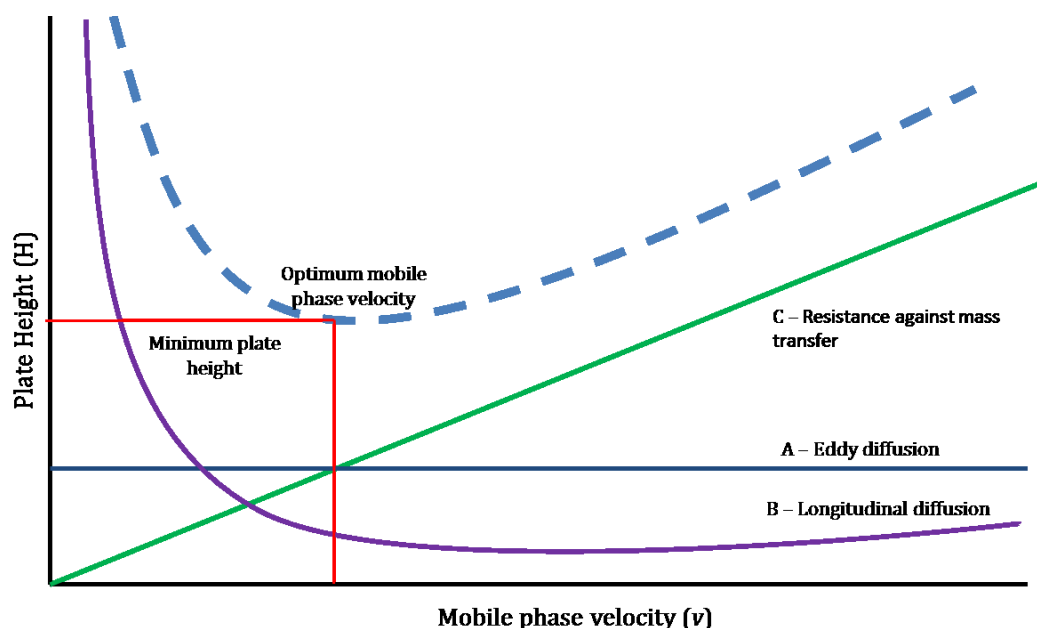


**Figure 1.4 The graphical representation of the Van Deemter plot**

The H-v plot is of considerable use in determining an ideal mobile phase flow rate at which optimal column efficiency can be attained. When the empirically determined H-v plot (efficiency vs velocity or mobile phase flow) is considered the general rules of thumb are:

- The column efficiency increases with the smaller plate height (H), however, even with the lowest plate height if the flow rate is below optimum it contributes to very long runtime and the column efficiency is adversely affected by the longitudinal diffusion (B term in van Deemter equation).
- At very high flow rates, the column efficiency drops hence peak separation is compromised.

The decrease in the particle size (diameter less than or equal to  $2\mu\text{m}$ ) and using more uniformly packed columns allow an increase in the flow rate and reduce the runtime without compromising overall efficiencies or less band broadening across a much higher flow rates. This decrease is observed as the analyte spends less time between the column particles leading to shorter travelled diffusion paths.

The development Ultra Performance Liquid Chromatography (UPLC), which uses columns of particle size less than  $2\mu\text{m}$  in diameter, has had a large impact on metabolic profiling studies. Smaller particles within the column packaging increase the maximum resolution possible and the range of flow rates over which efficient separation can be achieved, compared to conventional

High-Performance Liquid Chromatography (HPLC) [2]. However, as the particle size decreases more force/back-pressure is required by the pump to move the mobile phase at a given velocity. Traditional HPLC with a larger particle size column (e.g. 4  $\mu\text{m}$ ) typically operates with continuous mobile phase flow below 6000 psi. UPLC pump systems in combination with a smaller particle size ( $\leq 2\mu\text{m}$ ) are capable of operating constant mobile phase flow at pressure limits of 12000-24000 psi and need to do so to achieve better resolution in the same time as HPLC or the same resolution in reduced time.

Technological advances in LC and the impact on metabolic profiling studies are discussed in detail by Wilson *et al.* [56] and Forcisi *et al.* [57].

Chromatographic separation of individual molecular entities can be achieved by two different types of gradient elution: isocratic elution and gradient elution. Isocratic elution uses homogenous mobile phase throughout the separation runtime. When using isocratic elution the formulation of the mobile phase has chemical selectivity towards just a few selected molecular entities, therefore it is conventionally used more for targeted analysis. On the other hand, gradient elution uses a mixture of different solvents (mobile phases) in various proportion to achieve separation. It is considered to have more versatile usage in metabolic profiling as it gives the flexibility to separate chemically diverse molecular entities without the need to change the stationary phase. The two most common type of chromatography applied for metabolic profiling are reversed phase chromatography (RPC) and hydrophilic interaction liquid chromatography (HILIC). The most common applied stationary phase for RPC are alkyl hydrocarbon (typically  $\text{C}_{18}$  octadecyl) bonded silica columns. The reversed phase chromatographic separation is achieved by employing a polar or aqueous (high water content) mobile phase and gradient elution is achieved as the content of organic mobile phase increases (methanol or acetonitrile). The stationary phase used in reversed phase is hydrophobic which is ideal for the separation of non-polar metabolites including fatty acids and lipids [2, 58]. Depending upon the different chemistries of the column, UPLC-MS can provide the detection of thousands of metabolite features in a given biological sample.

### **1.2.2 Direct Infusion (DI) or Flow infusion analysis (FIA)**

Direct infusion or flow infusion analysis mass spectrometry (DI-MS or FIA-MS) is a high-throughput technique used in metabolic profiling studies as an alternative to LC-MS [2]. With no chromatography data processing is relatively more simple and robust. In FIA, a 'plug' of sample is introduced through radial and convection-diffusion into the LC solvent carrier stream using an isocratic elution. The plug disperses into and mixes with the solvent carrier under *laminar flow* conditions to form a concentration gradient. The important operating parameters that control the

magnitude of the dispersion are sample injection volume, flow rate and tubing bore size. Highly reproducible injected sample volume is vital for the optimum functioning of FIA thus full loop injections are used to yield better recoveries and better precision.

Data are captured as a full-width spectrum, generally in both positive and negative ionisation mode (depending on the instrument) and an average chromatogram is calculated as the central portion of the single infusion peak containing all the metabolites. Data pre-processing is limited to aligning single summoned or averaged mass spectra for each sample [2, 62].

FIA-MS-MS has already been shown to be fit for purpose and useful in metabolic profiling of polar and lipophilic metabolites especially in lipidomics [63, 64]. Both FIA-MS (full scan spectra) and FIA-MS/MS can be achieved in single quadrupole TOF, triple quadrupole, ion trap, and Orbitrap instruments. Definitive metabolite identification is often limited due to the absence of chromatography where stereoisomers of metabolites with same accurate mass but different chemical structures will be detected as a single  $m/z$  (e.g. hexoses, glucose and fructose). This problem can be addressed by FIA-MS/MS which uses multiple reaction monitoring (MRM; explained in the next section) scans in quadrupole instruments. FIA-MS and FIA-MS/MS are more susceptible to ion suppression [62].

### **1.2.3 Mass spectrometry**

One of the most compatible options to analyse chromatographic eluting molecular entities of interest into specific isolated chemicals is via mass spectrometry (MS). Mass spectrometry (MS) measures the molecular mass to charge ( $m/z$ ) ratio to differentiate, detect and quantify compounds. A mass spectrometry consists of three main parts: an ion source, a mass analyser and a detector [65]. While the function of the ion source is to convert the molecule of interest into gas-phase ions (positively or negatively charged), the mass analyser separates these ions according to their mass to charge ( $m/z$ ) ratio. The separated ions are then captured and measured by the detector.

Several options for ion sources are available and are categorised into hard and soft ionisation techniques [66]. 'Hard' ionisation techniques such as electron impact (EI) ionisation are generally used in conjunction with gas chromatography (GC) where a volatile eluent is introduced into the source under vacuum. This technique allows a high degree of fragmentation by bombarding a beam of electrons into the molecule causing ions to fragment. It is considered as a good technique for structural elucidation. However, it is not suitable for use in conjunction with Liquid chromatography due to the 'hard' nature of the EI which introduces a high degree of fragmentation of the intact analyte, and it can be difficult to see a molecular peak with high

accuracy. In contrast, the most commonly used 'soft' ionisation technique is electrospray ionisation source (ESI) which allows the coupling of an LC to MS. It is capable of extensively ionise a large number of molecular entities present in biofluids. ESI operates under atmospheric pressure and causes minimal ion fragmentation. The liquid flow from LC is nebulised into a droplet spray, and charge exchange in solution forms a large number of ions. The main ions are formed by the non-covalent addition or loss of hydrogen  $(M+H)^+$  or  $(M-H)^-$ . The adduct ions such as sodium  $(Na)^+$ , potassium  $(K)^+$  or ammonium  $(NH_4)^+$  in positive mode and chloride  $(Cl)^-$  in negative mode are generally formed by interaction with other molecular species/salt content of the matrix. Positive and negative ions are formed via selecting the electric polarity of the source and also depending on the physiochemical properties of the analyte [2, 66, 67].

Generated ions are then accelerated by the atmospheric pressure within the ion source via kinetic energy to the vacuum region (mass analyser and detector) of the MS. These ions are separated according to their mass to charge ratio ( $m/z$ ) in the mass analyser. Ions are deflected in the magnetic or electric field of the mass analyser according to their masses (lighter ions get deflected more than heavier ions, and so the greater the ion is charged, the more it gets deflected). According to electrodynamics, when subject to magnetic or electric fields, ions with same  $m/z$  move in the same path in a vacuum. Finally, the ion mass detector captures the separated ions at a predefined scan range and signals are converted to a digital output.

The three main mass analysers generally used in metabolic profiling studies are categorised as: a triple quadrupole, QQQ (e.g. AB SCIEX Triple Quad™ 6500 used for targeted LC-MS/MS metabolic profiling of oestrogen metabolism in Chapter 2 and Waters Xevo® tandem quadrupole stepWave™ used for targeted LC-MS/MS and FIA-MS/MS in Chapter 3); a quadrupole-time of flight, Q-TOF (e.g. Waters Xevo® G2-S Q-TOF used for untargeted metabolic profiling Chapter 4); and orbitrap systems (e.g. Thermo Exactive™ Plus Orbitrap). Each mass analyser varies in performance and uses different methods to separate ions according to  $m/z$ , by exploiting the properties of ion mass and charge [65]. For a detailed review of principles of triple quadrupole (QQQ) and time of flight (TOF) mass analysers, refer to Watson J.T *et al.* [65] and Ardrey R.E [68]. For comparison of sensitivity, accuracy, mass range, resolution, fragmentation of the different mass analysers mentioned, refer to Prasad *et al.* [69].

QQQ mass analysers are used in the targeted analysis to analyse all ions over small window range (typically 0.5 Da, 'nominal' mass) achieving high specificity and sensitivity through selective reaction monitoring (described below). TOF analysers are routinely used for untargeted metabolic profiling as these possess the ability to simultaneously capture a broad range of  $m/z$

values at high mass resolution; fast data acquisition speed also makes them highly compatible with high-resolution chromatography.

#### 1.2.4 The Triple Quadrupole (TQ or QQQ) mass analyser

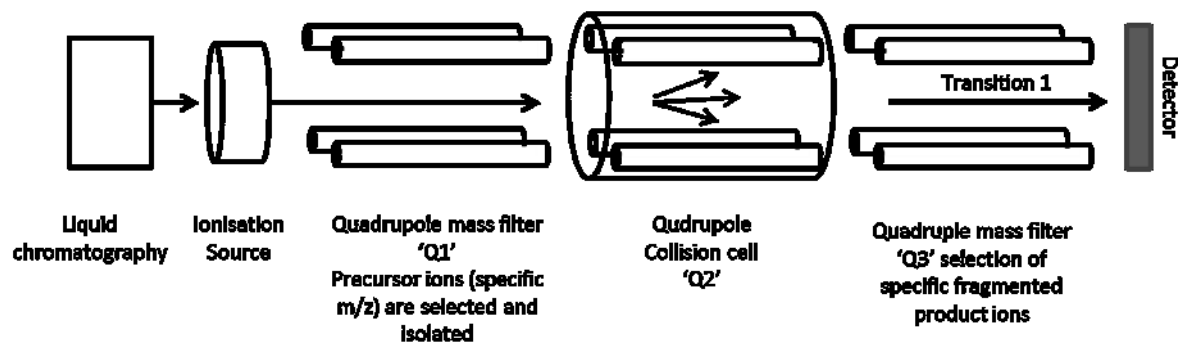


**Figure 1.5** A schematic diagram of a typical triple quadrupole (TQ or QQQ) mass analyser

The basic principle of a single quadrupole mass analyser is that it varies radio frequency voltages between two opposite pairs of parallel cylindrical rods to maintain trajectories of ions and force them to travel between the four rods. The ions of specific  $m/z$  ratio or range of  $m/z$  ratios reach the detector depending on the ratio of the applied voltages [65]. When three quadrupoles are connected in a linear series as shown in Figure 1.5, it is called a triple quadrupole (TQ or QQQ) mass analyser.

Triple quadrupole mass spectrometers have high sensitivity, making them ideal for targeted metabolite profiling. It allows targeted metabolite identification and accurate quantification by acquiring highly resolved and accurate MS/MS spectra. Tandem mass spectrometry or MS/MS or MS<sup>2</sup> refers to the combination of two high-resolution mass analysers (Q1 and Q3) in series with a collision chamber (Q2) in between which is used for fragmentation of ions.

Triple quadrupole MS is one of the most common MS systems for the routine quantification of metabolites at low abundance (femtogram range). This is achieved by employing multiple reaction monitoring (MRM) scan analysis which is unique to triple quadrupole instruments. MRM involves monitoring specific pairs of precursor and product ions called transitions [68, 69]. The metabolite precursor ions are generated in the ESI source and then accelerated to the first quadrupole (Q1) where the precursor ions are isolated according to a pre-defined  $m/z$ . The second quadrupole (Q2) acts as a fragmentation chamber or a collision cell, where the selected precursor ions are fragmented to product ions typically by colliding the ions with a neutral gas (e.g. argon); this process is called collision-induced dissociation (CID). Finally, the product ions are accelerated to the third quadrupole (Q3) or second MS which acts as a mass filter, separating

specific  $m/z$  of the product ion, which is then directed towards the ion detector to get measured. The resulting MS/MS spectrum consists of ion intensity versus peak from specific  $m/z$  product ion from the selected precursor ions without any other interference from chemical moieties in the sample matrix. When interfaced with chromatography technology, the identity of a specific metabolite is further confirmed by the retention time.

Triple quadrupole MS/MS experiments can be classified according to four scan modes: product ion scan, precursor ion scan, constant neutral loss scan and selected reaction monitoring scan. The aim of the product ion scan (PIS) used to determine structurally significant fragment ions, where MS/MS experiment is performed to find all the product ions which result from the breaking of the selected precursor ion. In the case where the mixture of components belonging to a specific compound class needs to be identified, the precursor-ion scan is used where MS/MS experiments are performed to find possible precursor ions that fragment to produce specific product ions. Constant neutral loss scan (CNLS) is where both precursor and product ions differ by the constant neutral loss. This is used when the specific neutral loss is characteristic of a class of compounds [68, 69]. Selected reaction monitoring (SRM) scan follows the same principle as MRM scanning described above.

### 1.2.5 The Time of Flight (TOF) mass analyser

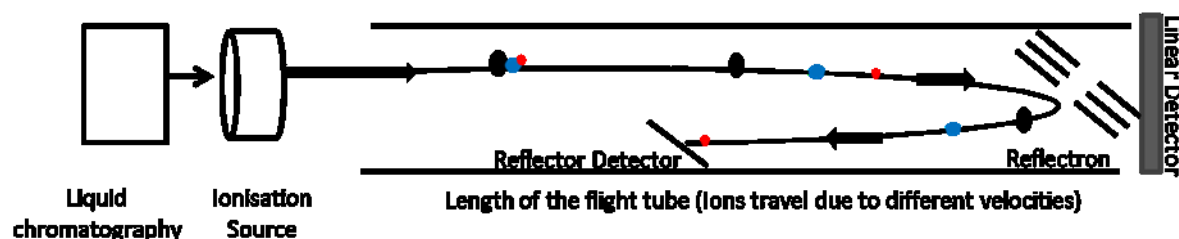


Figure 1.6 A schematic diagram of a typical time of flight (ToF) mass analyser

The basic principle of a time of flight (ToF) mass analyser is that it separates ions according to their  $m/z$  value depending upon the time taken to reach the known distance to the detector. Figure 1.6 summarises the working principle of typical ToF mass analysers. A known strength of the electromagnetic field is applied to accelerate all the ions present in the source. The ions with the same charge will acquire the same amount of kinetic energy; hence the velocity of each ion is proportional to square root of its mass. Thus, the  $m/z$  value of an ion is calculated according to a time measurement it takes to travel the flight tube of the instrument (more charged or heavier ions travels at a lower speed and vice versa) [68].

The advantage of TOF mass spectrometers, in addition to its simplicity to operate, is its capability of fast scanning. Therefore, it is increasingly being used with UPLC to acquire high

chromatographic mass resolution untargeted metabolic profiling data. In contrast to quadrupole mass analysers in which a particular  $m/z$  ratio or range of  $m/z$  ratios enter the mass analyser, in TOF, it is essential that at a known time, all ions of all  $m/z$  ratios present in the source get transferred to the mass analyser simultaneously and instantaneously. This allows the accurate measurement of the time of flight of all ions. By allowing sufficient time for all the ions of interest to reach the detector, a complete mass spectrum is acquired. After the delay of few milliseconds, another single pulse of ions is accelerated to the analyser to acquire full scan mass spectra.

A single MS or full scan mass spectra provide the full mass range or maximum amount of information for all the ions that are ionised in the source, separated in the mass analyser and which reach the detector. Depending on the instrument the resolution of ions with a similar  $m/z$  ratio is improved by using two techniques; either by increasing the length of the flight tube i.e. distance an ion takes to travel between source and detector or by using an ion mirror known as a 'reflectron' which reflects the ion beam to a second detector [68].

A Q-TOF configuration combines the properties of both quadrupole and time-of-flight analysers. In Q-ToF instruments, the quadrupole is the first mass analyser which allows the pre-selection of  $m/z$  values, followed by the collision cell for collision-induced dissociation (CID) for the fragmentation. Time of flight mass analysers is then used to capture the full high-resolution spectrum of pre-selected fragmented ions. The use of quadrupole and collision cell is optional (depending upon experimental design), their use can be disabled to allow the ions to pass to the ToF mass analyser directly. Application of Q-ToF instruments can be used for structure elucidation, identification and product ion scan.

### **Common terms used in mass spectrometry**

Some of the common terms used in MS are defined here:

Extracted ion chromatogram (chromatogram of a single ion), base peak chromatogram (displays most intense peak in each spectrum) and total ion chromatogram (mass spectral summed intensities from the same scan).

Mass spectrum as detected by mass spectrometry is an output of distribution of ions arranged in order of their mass to charge ratios. It consists of  $m/z$  ratio on x-axis and intensity of y-axis.

Mass resolution is defined as the value of  $m/\Delta m$ , where  $m$  is  $m/z$  of the ion to be measured, and  $\Delta m$  is the difference in Daltons between ion of interest and ion from which it has to be separated. The high mass resolution provides better separation of chromatographic peaks and the ability to separately detect metabolites with similar masses from that of accurate mass.

Accuracy refers to the ratio between the mass error calculated as the difference in between measured and theoretical mass to that of measured mass. Deviation inaccuracy is specified in parts per million (ppm).

Sensitivity is defined as the response of the analytical system measured per amount of sample injected into the system. Selectivity is the ability of MS to determine  $m/z$  or peak of interest without any interference from other components in the matrix.

Signal to noise ratio (S/N) is defined as the ratio of the analyte signal to an intensity signal of the noise measured in the blank. Scan rate is time interval at which mass detector detects a given mass ranges; higher scan rate enables detection of more features per chromatographic signal.

Limit of detection (LOD) is defined as the lowest concentration of an analyte that can be detected reliably and at the same time can be discriminated from the concentration/signal obtained in a blank sample. Generally, it is accepted as three times the signal detected in the blank sample.

Limit of quantification (LOQ) is the lowest concentration of an analyte that can be measured reliably with predefined precision, accuracy, selectivity and sensitivity. In other words, the value at which the difference between two different amount of analyte can be reasonably differentiated. LOQ can be considered same value as LOD, but it cannot be lower than LOD.

A detailed summary of terminologies used in chromatography and mass spectrometry can be found in Ardrey R.E [68] and Murray *et al.* [70].

### **1.3 Experimental strategies in metabolic profiling**

Depending on the biological question to be answered, two generalised experimental strategies which employ LC-MS are applied in metabolic profiling. One is 'targeted metabolite profiling', with preselection of metabolites and often hypothesis-driven, and the other is a more data-driven/hypothesis generating 'untargeted or global metabolic profiling' with limited or no *a priori* biological knowledge [2, 24]. Each of the approaches is complementary to one another and has their inherent advantages and disadvantages which are discussed below. The differences in the LC-MS workflow of both the strategies regarding sample preparation, data capturing, and data processing are summarised in Figure 1.8.

#### **1.3.1 Targeted metabolite profiling**

Targeted metabolic profiling refers to the quantification of absolute concentrations of a pre-defined group of chemically and biochemically characterised metabolites that can be related to one or more metabolic pathways of interest [2, 22, 71]. Targeted metabolic profiling can provide

a focused hypothesis testing strategy leading to an investigation of a particular pathway e.g. there are metabolic differences in oestrogen metabolism in women with differential risk of breast cancer [72]. It also provides the advantage of controlling or analysing multiple variables, such as the dynamic range of metabolite concentrations and controlling the influence of ion suppression/matrix effects of analysed molecules.

The most common analytical platform applied for targeted metabolic profiling is liquid chromatography-tandem mass spectrometry (LC-MS/MS) which is the most sensitive method for quantifying metabolites as low as femtogram/mL in concentration. Also, it provides high specificity, accuracy and precision. This approach is not new and has been used since the late 1960s as part of bio-analytical chemistry to quantify a range of specific metabolites such as amino acids and glucose in urine and serum samples [22, 71]. The use of isotopically labelled internal standards in LC-MS/MS experiments is effective in compensating for inefficiencies/loss of analyte during sample preparation steps especially extraction. It also compensates for ion suppression matrix effects and increases quantification accuracy [73].

#### **1.3.1.1 Quantitative analysis and construction of a calibration curve**

A calibration curve is used in a quantitative analysis to determine the instrument response obtained from an analyte in a test (or unknown) sample and comparing that with the concentration of a known amount of analyte in reference standards. To obtain reliable quantification from a targeted LC-MS method a correct selection of calibration standards, their preparation, calibration strategy, use of internal standards, regression of calibrator response and calculation of the concentration of analytes in all unknown samples using the regression model are vital [74, 75].

In most targeted LC-MS assays there are two types of standards used: External reference standards and internal standards. An external standard calibration is constructed using a series of calibration solutions different from the actual test samples and instrument response is measured within the same run. In general, to minimise the effect of ion suppression or enhancement, external calibration solutions are preferably prepared in a matrix reflecting the same composition as the test sample. However, it is worth noting that the physiological composition of the matrix-matched calibrant will not be exactly same as that of test sample. It is often seen as the best compromise achievable to reduce the possibility of matrix suppression. For example, steroid stripped urine or serum which is passed through a carbon filter will be a cleaner matrix than actual patient urine or serum. In some cases, use of a blank matrix (e.g. calibrants solutions prepared in water) can be advantageous where the more stable matrix is required to measure the response of abundant matrix compounds [72, 75, 76]. Plus, a choice of matrix-

matched or using a blank matrix to prepare calibration standards also depends on the nature of the instrument and its sensitivity to measure the analyte. Other aspects which are important to constructing a calibration curve are:

- External calibration standards should be authentic reference standards of the analytes to be measured in the test sample with known purity.
- The number of calibration standards: For method development, at least seven different concentrations should be prepared as duplicates. Subsequent calibrant numbers depend on the established range of the analytes to be measured and resources available.
- The calibration range must cover the physiological concentration of the analyte found in the test sample. Ideally, in most of the test samples, concentrations should fall between the middle concentration ranges of the calibration curve.

To determine the concentration of the analyte in the test sample from the external calibration curve, a regression line is plotted with X-axis= concentrations of the analyte in the test samples ( $C_A$ ) and Y-axis is the response of the analyte ( $R_A$ , usually peak area or peak height). The same volume of the test sample is injected, and its peak area response is used to determine the concentration. Figure 1.7 illustrates how the concentration of the analyte in the test sample is measured using both external reference standards and internal standardisation curve.

The internal standardisation is used as internal calibration of the measurements. In internal standardisation, an additional set of compounds called internal standards of same known concentration are added (referred as 'spiking') to all the test samples including external calibration standards early in the sample preparation process. Internal standards are used to compensate for potential sources of error such as any inconsistency or loss during sample preparation especially when sample preparation involves multiple steps such as extraction, derivatisation etc.; instrument drift; change in sample injection volume or variability in adducts profiles between test samples and calibrants or ionisation variability [74, 75, 76].

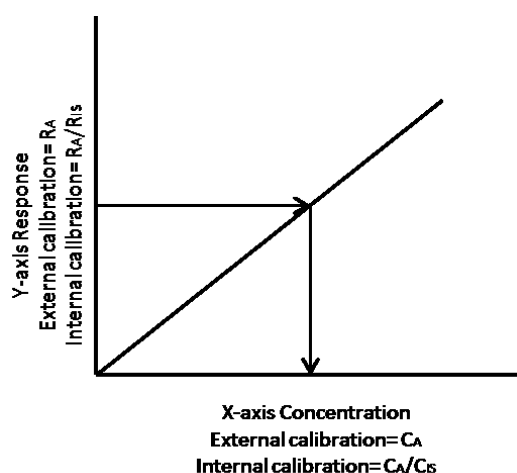


**Figure 1.7 Calculation of analyte concentration in the test sample using external reference standards and internal calibration standardisation curve.  $R_A$  =peak response of analyte,  $R_{IS}$  =peak response of internal standards,  $C_A$  =concentration of the analyte and  $C_{IS}$  =concentration of internal standards**

The two major considerations when selecting internal standards: internal standards should never be naturally present or found in the test sample and internal standards should be very similar to target analyte in the test sample in terms of its physicochemical properties and structure. In addition, there should not be any mass spectral overlap between the natural target analyte and internal standards. Therefore, for quantification, isotopically labelled standards should have a monoisotopic mass of at least 3 Da higher than that of the natural analyte in the test sample to avoid signal overlap [74, 75, 76]. Furthermore, labelled standards must be stable during sample preparation well resolved, compatible with MS detector, as the analyte of interest in the test sample under same experimental conditions. For these reasons, it is preferred to use isotopically labelled internal standards.

The rationale behind adding the same known amount of internal standards at the beginning of the experiment to all the test and calibration standard samples is that is the internal standards should experience same changes as the analyte of interest. By constructing a calibration curve based on the ratio of the responses of the analyte to the internal standards and ratio of concentrations of analyte and internal standards, much of the experimental variations can be removed. Thus, using internal standards improves data quality, precision and accuracy of the measurements. In internal standard calibration, two chromatograms are measured one for internal standard and other for external reference standards. As shown in Figure 1.7 the concentration of the analyte in the test sample is measured by paring the ratio of peak response of two peaks one of the internal standards and other of reference standard to the ratio of concentrations [76].

Targeted metabolic profiling has undoubtedly played a significant role in the development of the metabonomics/metabolomics field especially with the advances made in MS technologies. Targeted LC-MS/MS metabolite screening led to the discovery of the naturally occurring physiological concentrations of metabolites in one-carbon metabolism [77], quantification of carboxylic acids, amino acids, sugar phosphates and nucleotides. It was used to determine and quantify the coverage of metabolites in the central metabolism of glycolysis, pentose phosphate pathway, and tricarboxylic acid cycle [78]. Targeted profiling of citric acid metabolites was recently identified as metabolic signatures of myocardial ischaemia [79] and increased concentrations of a combination of three amino acids were predicted to be a risk for future diabetes [80].

Overall targeted metabolic profiling is a 'bottom-up' systems biology approach where quantification of specific metabolites and their interactions within pathways are studied to give information on whole-organism functional integrity with significantly higher accuracy and precision [2]. However, this approach also has few limitations, for example, not all multifactorial diseases have specific hypotheses stating which metabolites or metabolic pathways are related to pathophysiological changes and the potential of identification of unknown novel metabolites related to the disease outcome is eliminated. A greater level of sample preparation is required to separate the metabolites of interest from other matrix components and may require lengthy hydrolysis and derivatisation steps. In addition, preparation of calibration curves with authentic chemical standards of different concentrations is extensive but necessary for the absolute quantification of metabolites. A typical workflow of LC-MS targeted metabolic profiling is summarised in Figure 1.8.

### **1.3.2 Untargeted metabolic profiling**

In contrast to targeted metabolite profiling, untargeted metabolic profiling is global in scope. The main objective of this approach is to detect and measure relative changes in response or intensities (not absolute concentration) of as many metabolites as possible across multiple metabolite classes in a given biological sample, with the aim to obtain maximum coverage of the metabolome [2, 22, 81]. Subsequent data processing and analysis provide novel insight into the changes in the metabolome which generates a hypothesis to the biological question for e.g. if there are any metabolic differences between humans diagnosed with cancer in comparison to healthy controls. Further targeted metabolic profiling studies are then performed to test the hypothesis generated by the metabolites or metabolic pathways of interest and to validate any findings in independent sample sets.

The minimum amount of sample preparation is required in order to avoid the loss of any specific metabolite class. LC-MS-based untargeted metabolic profiling tends to measure any molecule (known or novel) that ionises within a specific mass range including metabolites with structures yet to be elucidated, hence the use of internal standards are not required. Furthermore, constructions of calibration curves are also not required. Metabolite quantification is determined by measuring relative changes in the integrated mass ion intensities (signal an ion generates at the detector) i.e. their peak areas. In these studies, the obtained results are considered 'fit for the purpose' in terms of accuracy and precision but much lower than for targeted metabolic profiling acquired data.

Given its sensitivity, many significant advances have been made in the last decade using LC-MS based untargeted metabolic profiling such as identification of novel substrates of different

enzymes: identification of both known and novel vitamin E urinary metabolites in mice which are altered upon activation of Pregnane X receptor (PXR), xenosensor which regulates expression of enzymes that are involved in the metabolism of xenobiotics [82]. It has also been useful in characterising metabolic pathways which are altered in mouse models of human multifactorial diseases such as cardiac disease [83], obesity and diabetes [84, 85]. Another application where untargeted metabolic profiling is proving useful, is the discovery of novel drug targets referred as 'therapeutic metabolomics' [22, 86]; for example, the discovery of the mutation in the gene encoding cytosolic isocitrate dehydrogenase 1 (IDH1) enzyme which leads to increased levels of the metabolite 2-hydroxyglutarate in cancer cells. Its inhibition can halt the conversion of the tumour of nervous tissue to a more lethal highly invasive glioma in the brain [22, 87].

The current major limitations of the LC-MS-based untargeted metabolic profiling strategy include chemical identification of thousands of peaks that are generated. Each peak represents a metabolite feature which corresponds to the ion with both unique mass to charge ratio and retention time. When using ion intensities as the sole means to quantify any molecule, complications such as ion suppression and retention time drift arise, leading to reproducibility issues [81]. This makes data processing, data analysis and identification of metabolites exceedingly complex and time-consuming. However, major progress has been made in the last decade to correct for analytical variations such as retention time drift and to account for deregulated peaks using software such as XCMS and Mzmine [88, 89].

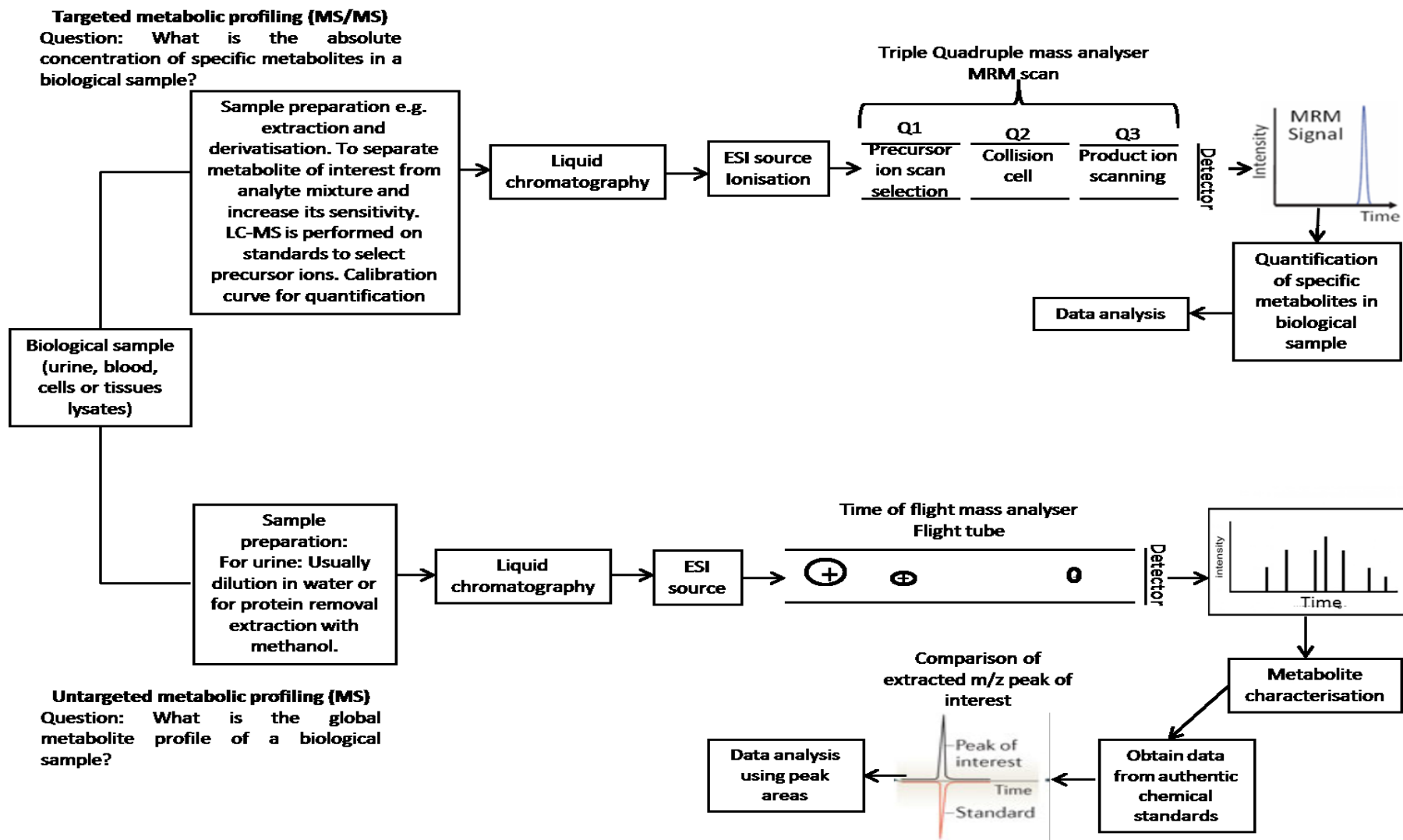
#### **1.3.2.1 Role of QC samples in untargeted metabolic profiling**

In large-scale, untargeted LC-MS studies where hundreds to thousands of samples are analysed over several batches, great care in experimental design is necessary to avoid significant challenges of inconsistency in the signal. These inconsistencies can be observed due to issues ranging from long-term reproducibility of the instrument, necessary maintenance breaks; degree and timing of signal drift not being consistent across all metabolites, mass accuracy error, and depending on biofluid under analysis. For reproducible and reliable metabolic profiling analyses, it is vital that stringent approaches are used to get stable retention time, mass accuracy and signal intensity. For these reasons, it is essential in untargeted profiling to incorporate adequate quality control (QC) samples for column conditioning and data quality assessment [90, 91].

Ideally, a QC should be a representative of all the samples analysed in the study so it covers the wide range of metabolites that are present in test samples. In addition, typically QC should be of same sample matrix as the test biological sample. Hence, QC samples are generally pooled samples (pooled aliquots of all the test samples) of all the biological samples getting analysed [90,

91]. In studies where enough samples are not available to prepare a pool of QCs, a commercial QC sample of the same biological matrix can be used.

In an untargeted LC-MS metabolic profiling, it is advisable that 5-10 QC samples are run at the beginning of the batch to condition the chromatographic column [90, 91]. The new column will not give the ideal representation of the samples at the beginning of the run due to small changes in the chromatographic conditions. In untargeted metabolic profiling where the use of internal standards is not practical. QC samples are periodically (ideally after every 10 test samples) run throughout the analytical batches to provide robust quality assurance for each metabolite. The QC response during data processing stages can be used to evaluate the quality of the data by evaluation reproducibility of characterised peaks and removing peaks with poor reproducibility, to correct for signal drift by algorithms using QCs signal response and merging the data from several batches together before statistical analysis [92, 93].



## 1.4 Conceptual applications of targeted and untargeted LC-MS metabolic profiling

Targeted and untargeted LC-MS based metabolic profiling has been increasingly applied to wide areas of research. In this PhD, the conceptual applications of LC-MS metabolic profiling are highlighted in three fields, related to human health. Figure 1.9 summarises the conceptual model implemented in this PhD project by developing and incorporating three LC-MS platforms for three different studies each in the context of functional genomics, biomarker discovery and exposome research. The figure also shows the strengths and limitations of each LC-MS platform mainly in relation to the time required for sample preparation; a panel of metabolites measured; time required for data analysis; data comparability and reliability of the quantification of detected metabolites and cost-effectiveness.

Acknowledging the widespread availability of the LC-MS analytical platform coupled with greater sensitivity compared with NMR and ability to analyse more diverse compounds than GC-MS, LC-MS was chosen as a preferred analytical technique in this PhD. Depending on the biological questions to be answered the experimental strategy ranged between targeted and untargeted analysis. On the targeted end, an LC-MS/MS method was applied to quantify 14 urinary endogenous oestrogen metabolites. While on the untargeted end of the spectrum urinary LC-MS global metabolic profiling method developed Want *et al.* [90] by was used. Biocrates Absolute*IDQ180* a commercial kit which allows detection and identification of 183 metabolites fell in the middle of targeted and untargeted analysis spectrum. The applications of metabolic profiling in the field of functional genomics, biomarker discovery and exposome studies are also briefly discussed in this section.

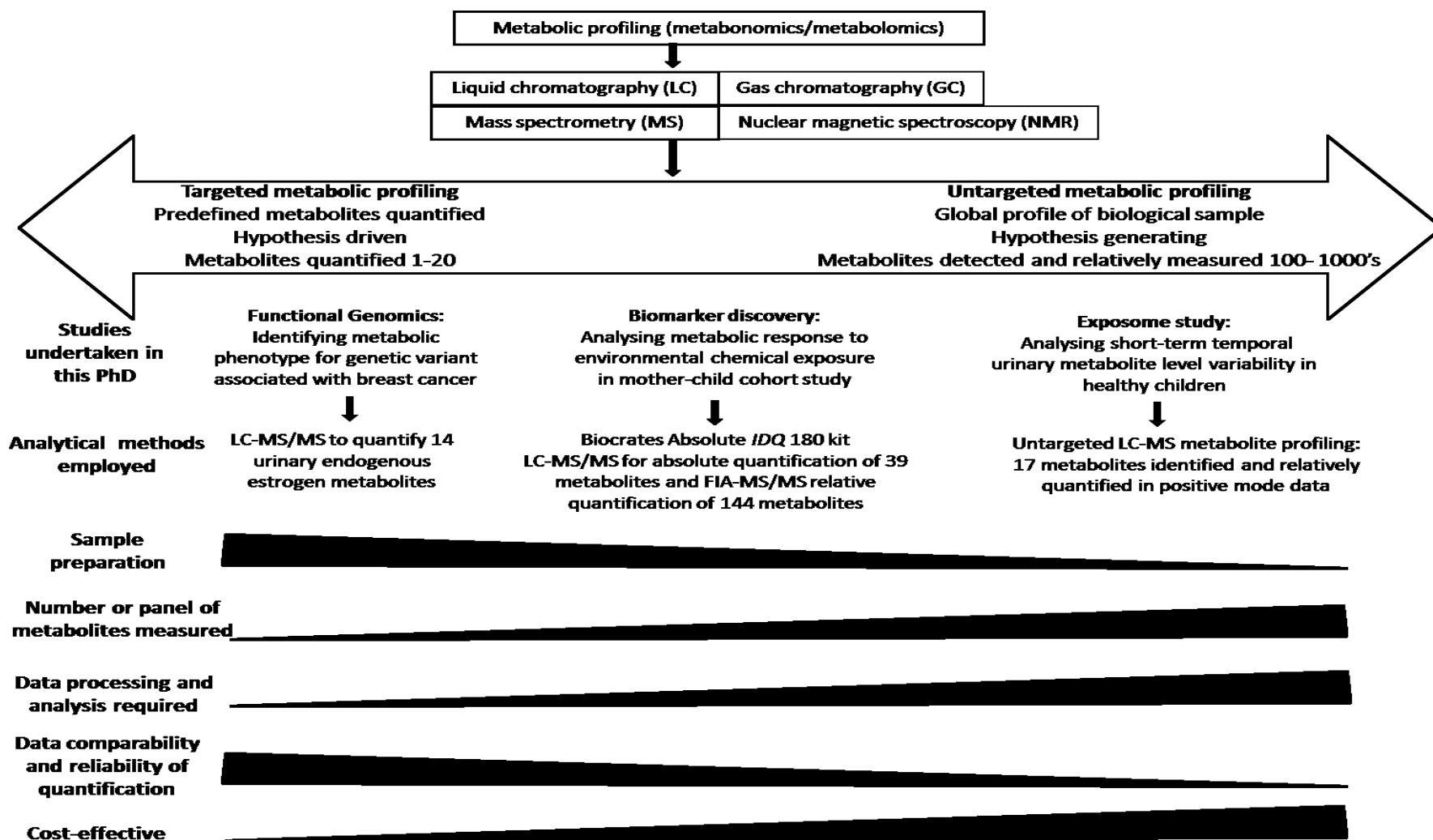


Figure 1.9 Summary of three LC-MS platforms incorporating metabolome analysis into three different studies carried out in this PhD project. This diagram also highlights strengths and limitations observed for each method in this PhD.

### 1.4.1 Functional genomics

Functional genomics is a field of genomics described by Gibson and Muse [94] as '*biochemical or physiological properties and functions of all genes and genetic products in organism's genome*'. The definition has been expanded by Pevsner [95] to include properties and functions of mRNAs, proteins and non-genetic elements encoded by genes. Advancements in techniques to analyse single nucleotide polymorphism (SNP) and mutations in various genes and their products within biological samples have driven the field forward. While several Genome-wide association studies (GWAS) have identified various disease causative/risk loci but they lack understanding and information on their functional biological processes [64]. Metabolic profiling is emerging as a powerful tool for functional genomics studies to establish the functional linkages between genes and their phenotypic expressions [96].

The realisation that obtaining genome sequence is not enough to elucidate the relationship between gene expression and its phenotypic outcome, has triggered an increased interest in using metabolic profiling approaches. One approach that provides an advantage of a holistic understanding of biological behaviours of many multifactorial diseases is to measure the concentration of metabolites of specific pathways which gets influenced by genetic variations. In order to achieve this genetic variation i.e., mutation or SNP is introduced into a system to observe the unknown effect of gene function on phenotype i.e. metabolome. By measuring the alternations in concentrations of metabolic phenotypes functions may be assigned to genetic variation of the gene [97]. Such approach has been applied in cancer research to study the different stages, subtypes of breast cancer tumour [97], to determine the effect of drug intervention [5, 98], and to examine the clinical response to therapy [99].

A review by Suhre & Gieger [64] has reported a list of various GWAS which used hypothesis-free metabolic profiling approach (mainly using Biocrates Absolute kit) to find the associations between genetic variants and their influence on metabolites. These metabolic traits were termed as genetically influenced metabolotypes (GIMs). In another study Suhre *et al.* [100] listed 37 genetic loci associated with human metabolite level individuality in human serum in German KORA and British Twins UK cohort studies. It has been hypothesised that GIMs identified in GWAS-metabolic profiling studies is essential to facilitate gene-based personalised clinical therapy especially in clinical studies e.g. pharmacogenomics to monitor/tailor the response to drug treatment and in nutrigenomics studies [64, 100].

However, there are some challenges in applying metabolic profiling in functional genomics studies such as how many metabolites need to be measured to find an alignment of a genetic variant with related altered metabolic pathways [99]. How to analyse and best use untargeted

metabolomics data and deal with the correlation between metabolites, and most importantly how to treat data acquired from different experimental techniques [64]. Given the complexity and diversity of the metabolome instead of measuring all the metabolites, a useful alternative approach could be to use LC-MS to measure absolute quantification of unique metabolic pathways which are *priori* hypothesised to be related given genetic variant. In this way, the alignment of genetic variant and altered metabolic phenotypes will allow the identification of other related pathways for further investigation.

### 1.4.2 Biomarker discovery

National Institutes of Health, in the US under the initiative of Biomarkers definition workgroup 2001 [101] has defined biomarkers as: '*A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention*'. Conceptual applications of biomarkers include an indicator of disease prognosis, monitoring/predicting staging of the disease, diagnostic tool to distinguish patients from the healthy population, predicting disease risk factors and monitor clinical response to drugs. Since metabolome represents the best reflection of an individual's phenotype at a specific time. Assessments of metabolic profiles in bio-fluids provide a powerful approach to discover biomarkers that could improve disease diagnosis and prognosis, which potentially addresses an important clinical need [102].

Incorporating LC-MS metabolic profiling into large population cohort studies such as European Prospective Investigation into Cancer and Nutrition (EPIC) [103], National Health and Nutrition Examination Survey (NHANES) [104] and Framingham Heart Study [105, 106] has led to the identification of a number of putative biomarkers [107]. LC-MS metabolic profiling has been extensively used to quantify the response of a wide range of exposures on biochemical perturbation consequentially influencing human health. Among them, most studied are dietary exposures [10, 109, 110] and environmental chemical exposures, e.g. dioxin or dioxin-like compounds [107, 111, 112].

However, LC-MS metabolic profiling in mother-child cohort studies related to environmental chemical exposure has lagged behind despite many findings systematically reporting adverse pregnancy outcomes. Metabonomics in pre-/post-natal health is not a new concept, however metabolic profiling using a high-throughput LC-MS platform to distinguish the metabolome of healthy pregnancy from complicated pregnancy to improve neonatal health is still in its infancy [113]. The first review of LC-MS and NMR-based metabolic profiling in pregnancy research was published in 2009 [114]. In the last decade, LC-MS metabolic profiling has been used on biological samples (e.g. plasma, serum, urine, placenta and cord blood) obtained from pregnancy women

and neonates for metabolomics interrogations; to find biomarkers of preeclampsia, preterm delivery, malformations, small for gestational age and hypoxic-ischemic encephalopathy [113, 115]. Overall, the results emphasised the clinical potential of targeted and untargeted LC-MS metabolic profiling in the complex context of pregnancy where mother and foetus are interacting for mutual benefit. The impact of environmental exposures in developmental periods of early life may play an important role in exerting susceptibility of disease in later life of an individual. LC-MS metabolic profiling methods hold the potential to quantifying the metabolic response to some of the persistent chemical exposures can be used in understanding how they regulate human metabolism and in the assessment of risk factors.

### 1.4.3 Exposome studies

Untargeted metabolic profiling approaches using analytical techniques such as LC-MS and NMR have demonstrated their applicability in the better understanding of how environmental exposures influence health outcomes, which is a critical part of the exposome [38, 116]. However, LC-MS metabolic profiling in current EU-funded initiative exposome project i.e. Human Early life exposome study (HELIX discussed in Chapter 4) is still at early stages both in the context of analytical methods employed and its conceptual framework. One of the early initiatives and challenges to advance the field of metabolic profiling in exposome research is to define normal metabolite level variations over longitudinal time and within the same population. To accurately detect metabolites that establish the link between exposures, 'omics' profiles and health status it is vital to distinguish what is a significant deviation from normal metabolites level variation and background noise across the population.

Metabolites levels are known to vary over time within individuals due to the impact of a wide variety of factors such as genetics, age, BMI, diet, and lifestyle/environmental exposures. A 'cross-omics' study by Gruden *et al.* [117] using GC-MS and LC-MS/MS metabolic profiling to investigate inter-intra-individual variability in three repeat plasma samples of 16 individuals over one month showed high intra-individual variability in 20% of metabolites (234 total measured) with gender having strongest effect on metabolites levels followed by BMI and age. The metabolites that were shown be highly variable within same individuals were hormones, lipids, some amino acids and amines. Therefore, caution needs to be taken when these metabolites are of interest in population studies. For the appropriate study design of metabonomics-exposome studies, it is critical to characterise different sources of the normal level of metabolites variations over a short and long period of time and figuring out ways to minimise their effect.

## 1.5 Statistical analysis used in this thesis

Data pre-processing methods used to calculate metabolites concentrations in Chapter 2 and 3, and peak area in Chapter 4 of this PhD are discussed separately in their respective chapters. The general introduction to statistical data analysis methods used in this PhD is explained in this section. The specific choice of statistical analysis methods depends on the particular study design and question to be answered.

Spearman rank correlation coefficient was used in Chapter 2 to calculate the association between urinary oestrogen metabolites and in Chapter 3 to calculate the associations between metabolites concentrations and persistent organic pollutant exposure. It is one of the simplest non-parametric (does not require characterisation of distribution of the variables being passed) statistical tests used to measure the rank based degree of association between two continuous variables  $x$  and  $y$  (Bivariate) [118, 119]. The two variables:  $x$  and  $y$  are considered to have a monotonic relationship which means they are positively correlated; if the value of  $y$  increases with an increase in  $x$  and negatively correlated if the value of  $y$  decreases with an increase in  $x$  and vice versa. Spearman correlation coefficients denoted by  $\rho_s$  is reported to describe the strength of the association, which varies between -1 to +1. The closer the value of  $\rho_s$  is to 0, the weaker the associations between the bivariate variables; and the closer the  $\rho_s$  is to -1 or 1, the stronger the association between bivariate variables is considered [119, 120].

Principal component analysis (PCA) is one of the most common multivariate statistical analysis techniques used in metabolic profiling studies [121, 122]. It is often used as the first step in order to visualise the multivariate data and to investigate the overall trends, including any clustering or outliers within the data [122]. PCA is an unsupervised method which means it does not assume any prior knowledge which is not present in the original data matrix; hence it provides an unbiased view of the data. PCA generates Principal components (PCs) variables to reduce the multi-dimensional data into the low-dimensional model. PCs are used to explain the variance within a dataset; each PC explains the underlying latent variable of the original variables in a linear combination.

Each PC consists of two matrices: one is set of scores (T) or score plot and the other one is set of loadings (T) or loadings plot. The scores plot describes each data point of the spectrum per observation. It projects the overall summary of each spectrum/observations or samples and how they are related to one another, thereby allowing the visualisation of any clustering within the observations or samples and identification of any outliers. The loading plot represents a linear combination of variables/properties making up the PCs (a means to interpret the pattern seen in the score plot). The PCs are generated in sequential order to explain the order of variance within

the dataset, the highest possible amount of variance explained is modelled by first PC followed by the second PC and so on. The PCA scores and the loadings plot were used in Chapter 3 of this thesis to discriminate between the metabolic profiles of the 1<sup>st</sup>-trimester maternal serum and umbilical cord serum. The PCA scores plot was also used in Chapter 4 of this thesis to explore the reproducibility of quality control samples across six batches.

Linear regression models are often used in metabolic profiling studies to observe the relationship between the dependent variable (also called the response, or regressor variable, denoted as  $y$ ) e.g. metabolite concentrations or peak areas and the independent variable (also called the explanatory or regressor variable, denoted as  $x$ ) e.g. exposure level of persistent organic pollutants, genotype or disease [123]. When there is more than one independent variable to be analysed for e.g. cofounders or covariates (usually denoted by  $x_1, x_2, x_3$  and so on), it is called a multiple regression model.

The following formula expresses the simple regression model:

$$y = \beta_0 + \beta_1 x + \varepsilon \text{ [Ref 124]}$$

Where  $y$  represents the dependent variable e.g. metabolite level often  $\log_{10}$  –transformed,  $\beta_0$  is the  $y$  intercept,  $\beta_1$  is the slope of the regression line,  $x$  is the independent variable e.g. exposure level of persistent organic pollutant and  $\varepsilon$  represents residual error [124].

The general formula for multiple regression models is:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \varepsilon \text{ [Ref 124]}$$

Where  $y$  represents the dependent variable e.g. metabolite level often  $\log_{10}$  –transformed,  $\beta_0, \beta_1, \beta_2, \beta_3$  are regression coefficients,  $x_1, x_2, x_3$  are the independent variable e.g. exposure level of persistent organic pollutant, diet, age etc. and  $\varepsilon$  represents residual error [124].

The main goals of regression models are to establish the strength of relationship between the dependent variables and independent variables, to investigate whether  $y$  could be predicted by set values of  $x_1, x_2$  and so on and to determine accurate response by screening which independent variables  $x_1, x_2$  etc. are more important than the others to explain the response in the dependent variable.

When regression models are conducted, three main outputs are presented. R-square ( $R^2$ ) also known as the coefficient of determination is a measure (percent) of total variance explained by the dependent variable in a linear model or in other words it provides the strength of the relationship between the model and the dependent variable. It could be presented as  $R^2 =$

explained variation/total variation and ranges between 0 to 100%, higher the  $R^2$  values better the data fits the model. The test or significance of the model i.e. whether if the overall relationship is statistically significant is represented by F-test and its P-value. If the F-test's P-value is less than 0.05 it is generally considered as a good fit for the model.  $\beta_0$  or regression beta coefficient can be negative or positive and if statistically significant represents the measure of mean change in the dependent variable for one unit of change in the independent variable [123, 124]. If the  $\beta_0$  is positive it can be interpreted as for every 1-unit increase in the independent variable, the dependent variable increases by the value of beta coefficient and if it is negative that for every 1-unit increase in the independent variable, the dependent variable decreases by the value of beta coefficient.

## **1.6 Current challenges in biomarker discovery with metabolic profiling: characterising the clinically recognised metabolic phenotypes**

With the incorporation of 'omics' platforms into exposome and molecular epidemiology studies, it has become clear that relation between genetics, lifestyle/environmental exposures, and disease are optimally studied from systems perspective [3]. However, despite more than 150,000 papers published on the topic of biomarker discovery using 'omics' technologies, genomics and proteomics have produced less than 100 clinically recognised biomarkers [125].

Metabolic profiling holds the potential to provide molecular biomarkers that link genetic and environmental exposures and phenotype to multifactorial disease endpoints. At present, there are still significant challenges in metabolic profiling when answering the biological question and identifying clinically accepted/useful metabolic biomarkers [102, 125, 126]. The number of metabolites used as biomarkers to detect disease prognosis and diagnosis in clinical settings ranges between few traditional metabolites such as glucose, thyroxin (T4), urea, creatinine, cholesterol, cortisol etc. and repertoire of metabolites for monogenetic disease inborn error of metabolism. There is a critical need for all form of biomarkers for multifactorial diseases such as neurodegenerative disorders, breast and colon cancers, type 2-diabetes, metabolic syndrome, and heart diseases.

Since 2005 nearly 3,000 papers are published on the topic of metabonomics/metabolomics biomarker discovery and 200 published papers on metabolic profiling in exposome and epidemiology [Scopus search]. However, the majority of the published studies didn't lead to any useful clinical biomarker discovery. Most of the studies get hampered by the fact that a precise, comparable and standardised method to characterise highly dynamic metabolic phenotype in large populations is missing which is vital for the discovery of molecular biomarkers. This lack of

rigorously standardised metabolic profiling methodologies in epidemiology studies can be the pitfalls of exposome and molecular epidemiology studies. Briefly, some the current key barriers include but are not limited to:

- Experimental strategies targeted vs. untargeted metabolic profiling
- Standardisation of assays and data comparability
- Lack of longitudinal data on metabolic phenotype variability
- Pre-analytical factors

For LC-MS metabolic profiling to realise its full potential in novel biomarkers discovery in population studies which can later be successfully translated into clinical settings it should overcome these challenges. With proper metabolic phenotyping methodology combined with genotype data, the field of metabonomics will play a vital role in long-term systems diagnostics leading to personalised healthcare.

## 1.7 Aim and outline of the thesis

The overall aim of this PhD thesis was to address and examine the potential and limitations of targeted and untargeted LC-MS metabolic profiling in relation to the discovery of clinically useful/acceptable biomarkers in exposome and molecular epidemiological studies. The goal was to assess and explore three different LC-MS metabolic profiling platforms to overcome some of the key barriers discussed in the above commentary in novel biomarkers discovery in population studies for the general scientific community and their application in exposome studies.

To achieve this, aim this PhD incorporated three LC-MS metabolic profiling platforms for three different kinds of studies each addressing novel research questions pertaining to categories of exposome research. The first study incorporated application of targeted LC-MS metabolic profiling in functional genomics research to define metabolic phenotypes of genetic variant identified as risk factors for breast cancer. The second study was a biomarker discovery study which used semi-targeted LC-MS to explore the impact of persistent organic pollutants on metabolic profiles in mother-child cohort study. The third was holistic exposome study which used untargeted LC-MS metabolic profiling method to analyse the source of metabolic phenotype variability in healthy children.

### Outline of the thesis

This PhD thesis consists of five chapters in total. This includes Chapter 1 which gives background to the main contents covered in this thesis and framework of this PhD project. Chapters Chapter 2, 3, and 4 each contains an introduction, methodology and results of three different studies undertaken in this PhD. Each chapter finishes with result interpretation and their biological significance, strength and limitations and conclusion of each study. Chapter 5 will summarises key research findings, discusses the overall conclusion of the PhD and what it has contributed to the field of metabonomics and exposome research.

### Chapter 2

The specific objectives of the **Functional genomics** study were:

- I. To develop a robust and accurate targeted LC-MS method for the quantification of fourteen endogenous urinary metabolites of oestrogen metabolism.
- II. To measure the absolute concentrations of individual oestrogen metabolites and explore the patterns of grouped oestrogen metabolites, their pathways and selected ratios in follicular phase urine of healthy pre-menopausal women.

- III. To investigate the role of a genetic polymorphism, SNP (rs45446698) which mapped to the Cytochrome P450 3A (CYP3A) haplotype in the regulation in levels of individual, grouped oestrogen metabolites and their pathways in women who were a carrier of *CYP3A7\*1C* allele than those to who were non-carriers.

### **Chapter 3**

The specific objectives of the **Biomarker discovery study** were:

- I. Use a targeted LC- and FIA-MS/MS-based metabolic profiling assay (Biocrates AbsoluteIDQ p180) which allows comprehensive coverage, identification and quantification of 145 metabolites across four metabolite classes to explore and compare the metabolome of healthy first-trimester maternal serum and foetal umbilical cord serum.
- II. To explore whether the prenatal low-dose exposure of persistent organic pollutants is associated with metabolite profiles in 1<sup>st</sup>-trimester maternal serum and cord serum of healthy pregnancies.
- III. To explore the links of discriminant metabolites of exposure to molecular pathways that could potentially have an adverse influence on pregnancy outcome and infantile or juvenile development.

### **Chapter 4**

The specific objectives of the **Exposome study** were:

- I. The primary goal of this study was to provide key information required to effectively design and optimise metabolic profiling analysis specifically in the context of HELIX exposome project. For this purpose, a pilot study was designed and conducted to evaluate the sources of variability in untargeted LC-MS urinary metabolites. These sources were experimental variability and two components of temporal biological variability (intra and inter-individual variations).
- II. To build a chemical reference library to automate the identification of mass spectra peaks. This was achieved by linking the retention time and masses of chemical standards with mass spectra peaks.

To investigate the reliability of the pooled diurnal time points from the same individual compared to analysing individual time points (morning and night) separately to get a reliable and reproducible representation of an individual's metabolite profile.

## 1.8 References

1. Oliver, S. G. *et al.* Systematic functional analysis of the yeast genome. *Trends Biotechnol.* 16, 373–378 (1998).
2. Dunn, W. B. *et al.* Systems-level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chem. Soc. Rev.* 40, 387–426 (2011).
3. Nicholson, J.K. *et al.* "System biology: metabonomics". *Nature.* 23, 1054–1156 (2008).
4. Nicholson, J. K. *et al.* 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica*, 29, 1181 – 9 (1999).
5. Nicholson, J. K. *et al.* Metabonomics: a platform for studying drug toxicity and gene function. *Nat. Rev. Drug Discov.* 1, 153–161 (2002).
6. Idle, J. R. & Gonzalez, F. J. Metabolomics. *Cell Metab.* 6, 348–351 (2007).
7. Holmes, E. *et al.* Metabolic Phenotyping in Health and Disease. *Cell* 134, 714–717 (2008).
8. Holmes, E. *et al.* Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature* 453, 396–400 (2008).
9. Nicholson, J. K. *et al.* Metabolic phenotyping in clinical and surgical environments. *Nature* 491, 384–92 (2012).
10. Wang, Z. *et al.* Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* 472, 57–63 (2011).
11. Koeth, R. A. *et al.* Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nat. Med.* 19, 576–85 (2013).
12. Kaddurah-Daouk, R. *et al.* Pharmacometabolomic mapping of early biochemical changes induced by sertraline and placebo. *Transl. Psychiatry* 3, e223 (2013).
13. Mamas, M. *et al.* The role of metabolites and metabolomics in clinically applicable biomarkers of disease. *Arch. Toxicol.* 85, 5–17 (2011).
14. Mak, C. M. *et al.* Inborn errors of metabolism and expanded newborn screening: review and update. *Crit. Rev. Clin. Lab. Sci.* 50, 142–62 (2013).
15. Ramautar, R. *et al.* Human metabolomics: strategies to understand biology. *Curr. Opin. Chem. Biol.* 1–6 (2013).
16. Issaq, H. J. *et al.* Analytical and statistical approaches to metabolomics research. *J. Sep. Sci.* 32, 2183–2199 (2009).
17. The Human metabolome database version 3.6. <http://www.hmdb.ca/>
18. Law, V. *et al.* DrugBank 4.0: Shedding new light on drug metabolism. *Nucleic Acids Res.* 42, (2014).
19. Becker, S. *et al.* LC-MS-based metabolomics in the clinical laboratory. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 883–884, 68–75 (2012).
20. Neepa, Y.C. *et al.* Role of thyroid hormones in human and laboratory animal reproductive health. *Birth defects research h.* 68, 479–91 (2003).
21. Grouzmann, E. *et al.* Disappearance rate of catecholamines, total metanephrines, and neuropeptide Y from the plasma of patients after resection of pheochromocytoma. *Clinical chemistry*, 47, 1075–82 (2001).
22. Bouatra, S. *et al.* The human urine metabolome. *PLoS One* 8, e73076 (2013).
23. Psychogios, N. *et al.* The human serum metabolome. *PLoS One* 6, e16957 (2011).
24. Patti G.J. *et al.* Innovation: Metabolomics: the apogee of the omics trilogy. *Nature Reviews Molecular Cell Biology* 13, 263–269 (2012).

25. Visscher, P. M. et al. Five years of GWAS discovery. *Am. J. Hum. Genet.* 90, 7–24 (2012).
26. Willett, W. C. et al. Balancing life-style and genomics research for disease prevention. *Science* 296, 695–698 (2002).
27. Lichtenstein, P. et al. Environmental and Heritable Factors in the Causation of Cancer - Analyses of Cohorts of Twins from Sweden, Denmark, and Finland. *N. Engl. J. Med.* 343, 78–85 (2000).
28. Lozano, R. et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: A systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 380, 2095–2128 (2012).
29. Ezzati, M. et al. Behavioral and Dietary Risk Factors for Noncommunicable Diseases. *N. Engl. J. Med.* 369, 954–964 (2013).
30. Elliott, P. & Wartenberg, D. Spatial epidemiology: Current approaches and future challenges. *Environ. Health Perspect.* 112, 998–1006 (2004).
31. Wild C.P. Complementing the genome with an exposome the outstanding challenges of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol Biomarkers Prev.* 14, 1847 (2005).
32. Wild, C. P. The exposome: From concept to utility. *Int. J. Epidemiol.* 41, 24–32 (2012).
33. Rappaport, S.M. and Smith, M.T. Epidemiology. Environment and disease risks. *Science* 330, 460–461 (2010).
34. Miller, G. W. & Jones, D. P. The nature of nurture: refining the definition of the exposome. *Toxicol. Sci.* 137, 1–2 (2014).
35. Kontis, V. et al. Contribution of six risk factors to achieving the 25×25 non-communicable disease mortality reduction target: a modelling study. *Lancet* 384, 427–37 (2014).
36. Vineis, P. & Perera, F. Molecular epidemiology and biomarkers in etiologic cancer research: the new in light of the old. *Cancer Epidemiol. Biomarkers Prev.* 16, 1954–65 (2007).
37. Nicholson, J. K. & Wilson, I. D. Opinion: understanding 'global' systems biology: metabonomics and the continuum of metabolism. *Nat. Rev. Drug Discov.* 2, 668–676 (2003).
38. Athersuch, T.J. & Keun, H.C. Metabolic profiling in human exposome studies. *Mutagenesis* 30 (6): 755–762 (2015).
39. Schulte, P.A. & Perera, F.P. Molecular epidemiology: principles and practice. San Diego, CA, USA: Academic Press, Inc.; 1993.
40. Slattery, LM.; Editorial: The science and art of molecular epidemiology. *J Epidemiol Community Health.* 56, 728–729 (2002).
41. Samowitz, WS. et al. Relationship of K-ras mutations in colon cancers to tumour location, stage and survival: a population-based study. *Cancer Epidemiol Biomarkers Prev.* 9, 1193–8 (2000).
42. Rashid, A. et al. Genetic epidemiology of mutated K-ras proto-oncogene, altered suppressor genes, and microsatellite instability in colorectal adenomas. *Gut* . 44, 826–33 (1999).
43. Slattery, ML, et al. Associations between Cigarette smoking, lifestyle factors, and microsatellite instability in colon tumours. *J Natl Cancer Inst;* 92:1831–5 (2000)
44. Slattery, M. et al. Associations between dietary intake and Ki-ras mutations in colon tumours: a population-based study. *Cancer Res* 60:6935–41 (2000).
45. Vineis, P. & Perera, F. Molecular epidemiology and biomarkers in etiologic cancer research: the new in light of the old. *Cancer Epidemiol. Biomarkers Prev.* 16, 1954–65 (2007).
46. Chadeau-Hyam, M. et al. Meeting-in-the-middle using metabolic profiling - a strategy for the identification of intermediate biomarkers in cohort studies. *Biomarkers* 16, 83–88 (2011).

47. Lindon, J. C. & Nicholson, J. K. Analytical technologies for metabonomics and metabolomics, and multi-omic information recovery. *TrAC - Trends Anal. Chem.* 27, 194–204 (2008).
48. Pan, Z. & Raftery, D. Comparing and combining NMR spectroscopy and mass spectrometry in metabolomics. *Anal. Bioanal. Chem.* 387, 525–527 (2007).
49. Beckonert, O. et al. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat. Protocol.* 2, 2692–2703 (2007).
50. Spagou, K. et al. HILIC-UPLC-MS for exploratory urinary metabolic profiling in toxicological studies. *Anal. Chem.* 83, 382–390 (2011).
51. Smolinska, A. et al. NMR and pattern recognition methods in metabolomics: From data acquisition to biomarker discovery: A review. *Anal. Chim. Acta* 750, 82–97 (2012).
52. Holmes, E. et al. Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature* 453, 396–400 (2008).
53. Nicholson, G. et al. Human metabolic profiles are stably controlled by genetic and environmental variation. *Mol. Syst. Biol.* 7, 525 (2011).
54. Annesley, T. M. Ion suppression in mass spectrometry. *Clin. Chem.* 49, 1041–1044 (2003).
55. Furey, A. et al. Ion suppression; A critical review on causes, evaluation, prevention and applications. *Talanta* 115, 104–122 (2013).
56. Wilson, I. D. et al. High resolution ‘ultra-performance’ liquid chromatography coupled to a TOF mass spectrometry as a tool for differential metabolic pathway profiling in functional genomic studies. *J. Proteome Res.* 4, 591–598 (2005).
57. Forcisi, S. et al. Liquid chromatography-mass spectrometry in metabolomics research: Mass analyzers in ultra-high-pressure liquid chromatography coupling. *J. Chromatogr. A* 1292, 51–65 (2013).
58. Akul Mehta. The principle of Reversed-Phase Chromatography HPLC/UPLC (with Animation) . PharmaXChange. (2012) Retrieved November 2015.
59. Jose Dimenez. Chromatography: Introductory Theory. 2014 [ONLINE] Available at: <https://teaching.shu.ac.uk/hwb/chemistry/tutorials/chrom/chrom1.htm>.
60. Rick Lake. How do small particle size columns increase sample throughput? 2016 [ONLINE] Available at: <https://teaching.shu.ac.uk/hwb/chemistry/tutorials/chrom/chrom1.htm>.
61. Delmar Larsen. The Van Deemter equation. 2016 [ONLINE] Available at: [https://chem.libretexts.org/Under\\_Construction/Chromedia/01Gas\\_Chromotography\\_\(GC\)/Gas\\_Chromotography%3A\\_Basic\\_Theory/13The\\_Van\\_Deemter\\_equation](https://chem.libretexts.org/Under_Construction/Chromedia/01Gas_Chromotography_(GC)/Gas_Chromotography%3A_Basic_Theory/13The_Van_Deemter_equation).
62. Draper J. et al. Flow infusion electrospray ionisation mass spectrometry for high throughput, non-targeted metabolite fingerprinting: a review. *Metabolomics*, 9, S4-29 (2013).
63. Illig, T. et al. A genome-wide perspective of genetic variation in human metabolism. *Nature Genet.* 42, 137–141 (2010).
64. Suhre, K. & Gieger, C. Genetic variation in metabolic phenotypes: study designs and applications. *Nat. Rev. Genet.* 13, 759–769 (2012).
65. Watson, J. T. & Sparkman, O. D. Introduction to Mass Spectrometry: Instrumentation, Applications, and Strategies for Data Interpretation. *Adv. Cancer Res.* 104, (2007).
66. Tania Portoles, et al. Use of soft and hard ionization techniques for elucidation of unknown compounds by gas chromatography/time of-flight mass spectrometry. *Rapid Commun Mass Spectrom.* 25, 1589-99 (2011).
67. Zhou, B. et al. LC-MS-based metabolomics. *Mol. Biosyst.* 8, 470–81 (2012).
68. Ardrey, R.E. Liquid Chromatography - Mass Spectrometry: An Introduction. John Wiley & Sons, Ltd, 2003.

69. Prasad, B. et al. Metabolite identification by liquid chromatography-mass spectrometry. *TrAC - Trends Anal. Chem.* 30, 360–387 (2011).
70. Murray, K. K. et al. IUPAC standard definitions of terms relating to mass spectrometry. *Pure Appl. Chem.* 85, 1515–1609 (2005)
71. Dudley, E. et al. Targeted metabolomics and mass spectrometry. *Adv. Protein Chem. Struct. Biol.* 80, (2010).
72. Eliassen, A. H. et al. Urinary estrogens and estrogen metabolites and subsequent risk of breast cancer among premenopausal women. *Cancer Res.* 72, 696–706 (2012).
73. Wieling, J. LC-MS-MS experiences with internal standards. *Chromatographia* 55, S107–S113 (2002).
74. Mike Sargent. Guide to achieving reliable quantitative LC-MS measurements. 2013 [ONLINE] Available at: [http://www.rsc.org/images/AMC%20LCMS%20Guide\\_tcm18-240030.pdf](http://www.rsc.org/images/AMC%20LCMS%20Guide_tcm18-240030.pdf).
75. Vicki Barwick. Preparation of Calibration Curves a Guide to Best Practice. 2003 [ONLINE] Available at: <https://www.lgcgroup.com/LGCGroup/media/PDFs/Our%20science/NMI%20landing%20page/Publications%20and%20resources/Guides/Calibration-curve-guide.pdf>.
76. John W. Dolan. When Should an Internal Standard be Used. 2012 [ONLINE] Available at: <http://www.chromatographyonline.com/when-should-internal-standard-be-used-0?id=&sk=&date=&&pageID=1>.
77. Bajad, S. U. et al. Separation and quantitation of water soluble cellular metabolites by hydrophilic interaction chromatography-tandem mass spectrometry. *J. Chromatogr. A* 1125, 76–88 (2006).
78. Buescher, J. M. et al. Ultrahigh performance liquid chromatography-tandem mass spectrometry method for fast and robust quantification of anionic and aromatic metabolites. *Anal. Chem.* 82, 4403–4412 (2010).
79. Sabatine, M. S. et al. Metabolomic identification of novel biomarkers of myocardial ischemia. *Circulation* 112, 3868–3875 (2005).
80. Wang, T. J. et al. Metabolite profiles and the risk of developing diabetes. *Nat. Med.* 17, 448–53 (2011).
81. Vinayavekhin, N. et al. Untargeted metabolomics. *Curr. Protoc. Mol. Biol.* 1–24 (2010).
82. Cho, J.-Y. et al. Metabolomics reveals a novel vitamin E metabolite and attenuated vitamin E metabolism upon PXR activation. *J. Lipid Res.* 50, 924–37 (2009).
83. Dunn, W. B. et al. Serum metabolomics reveals many novel metabolic markers of heart failure, including pseudouridine and 2-oxoglutarate. *Metabolomics* 3, 413–426 (2007).
84. Newgard, C. B. et al. A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. *Cell Metab.* 9, 311–26 (2009).
85. Kolak, M. et al. Adipose tissue inflammation and increased ceramide content characterize subjects with high liver fat content independent of obesity. *Diabetes* 56, 1960–1968 (2007).
86. Rabinowitz, J. D. et al. Metabolomics in drug target discovery. *Cold Spring Harb. Symp. Quant. Biol.* 23 (2011).
87. Dang, L. et al. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* 462, 739–744 (2009).
88. Smith, C. A. et al. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* 78, 779–787 (2006).
89. Katajamaa, M. et al. Toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics* 22, 634–636 (2006).

90. Want, E. J. et al. Global metabolic profiling procedures for urine using UPLC-MS. *Nat. Protocol.* 5, 1005–1018 (2010).
91. Want, E. J. et al. Global metabolic profiling of animal and human tissues via UPLC-MS. *Nat. Protocol.* 8, 17–32 (2013).
92. Kamleh, M. A. et al. Optimizing the use of quality control samples for signal drift correction in large-scale urine metabolic profiling studies. *Anal. Chem.* 84, 2670–2677 (2012).
93. Dunn, W. B. et al. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat. Protocol.* 6, 1060–1083 (2011).
94. Gibson, G. & Muse, S.V. *A primer of genome science* (3rd ed.). Sunderland, MA: Sinauer Associates. (2009).
95. Pevsner J. *Bioinformatics and functional genomics* (2nd ed.). Hoboken, NJ: Wiley-Blackwell. (2009).
96. Bino, R. J. et al. Potential of metabolomics as a functional genomics tool. *Trends Plant Sci.* 9, 418–425 (2004).
97. Buchholz, A. et al. Metabolomics: Quantification of intracellular metabolite dynamics. *Biomol. Eng.* 19, 5–15 (2002).
98. Denkert, C. et al. Metabolomics of human breast cancer: new approaches for tumor typing and biomarker discovery. *Genome Med.* 4, 37 (2012).
99. Griffin, J. L. The Cinderella story of metabolic profiling: does metabolomics get to go to the functional genomics ball? *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 361, 147–61 (2006).
100. Suhre, K. et al. Human metabolic individuality in biomedical and pharmaceutical research. *Nature* 477, 54–60 (2011).
101. Atkinson A.J. et al. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clin. Pharmacol. Ther.* 69, 89–95 (2001).
102. Nordström, A. & Lewensohn, R. Metabolomics: Moving to the clinic. *J. Neuroimmune Pharmacol.* 5, 4–17 (2010).
103. Fedirko, V. et al. Prediagnostic circulating vitamin D levels and risk of hepatocellular carcinoma in European populations: A nested case-control study. *Hepatology* 60, 1222–1230 (2014).
104. Calafat, A. M. et al. Exposure of the U.S. population to Bisphenol A and 4-tertiary-octylphenol: 2003-2004. *Environ. Health Perspect.* 116, 39–44 (2008).
105. Cheng, S. et al. Distinct metabolomic signatures are associated with longevity in humans. *Nat. Commun.* 6, 6791 (2015).
106. Wang, T. J. et al. Metabolite profiles and the risk of developing diabetes. *Nat. Med.* 17, 448–53 (2011).
107. Lind A.M. et al. The use of mass spectrometry for analysing metabolite biomarkers in epidemiology: methodological and statistical considerations for application to large numbers of biological samples. *European Journal of Epidemiology.* 1-17 (2016).
108. Andersen, M. B. S. et al. Untargeted metabolomics as a screening tool for estimating compliance to a dietary pattern. *J. Proteome Res.* 13, 1405–1418 (2014).
109. Scalbert, A. et al. The food metabolome: A window over dietary exposure. *Am. J. Clin. Nutr.* 99, 1286–1308 (2014).
110. Lankinen, Maria, & Ursula Schwab. “Biomarkers of Dairy Fat.” *The American Journal of Clinical Nutrition* 101.5 (2015).
111. Boffetta, P. et al. TCDD and cancer: a critical review of epidemiologic studies. *Crit. Rev. Toxicol.* 41, 622–636 (2011).

112. Sorg, O. et al. 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) poisoning in Victor Yushchenko: identification and measurement of TCDD metabolites. *Lancet* 374, 1179–1185 (2009).
113. Fanos, V. et al. Metabolomics application in maternal-fetal medicine. *Biomed Res. Int.* 720514 (2013).
114. Horgan, R.P. et al. An overview of proteomic and metabolomic technologies and their application to pregnancy research. *BJOG*, 116. 173–181(2009).
115. Syggelou, A. et al. Metabolomics in the Developing Human Being. *Pediatr. Clin. North Am.* 59, 1039–1058 (2012).
116. Athersuch T. Metabolome Analyses in Exposome Studies: Profiling Methods for a Vast Chemical Space. *Arch Biochem Biophys.* 9861(15), 30082-5 (2015).
117. Gruden, K. et al. A ‘crossomics’ study analysing variability of different components in peripheral blood of healthy caucasoid individuals. *PLoS One* 7, (2012).
118. Kruskal, W. H. Ordinal measures of association. *J. Am. Stat. Assoc.* 53, 814–861 (1958).
119. Chen, P. & Popovich, P. Correlation: Parametric and nonparametric measures. *SAGE* 1–105 (2002).
120. Moore, P. C. Statistics Explained: An Introductory Guide for Life Scientists. *Am. Stat.* 61, (2007).
121. Holmes, E. et al. Automatic data reduction and pattern recognition methods for analysis of <sup>1</sup>H nuclear magnetic resonance spectra of human urine from normal and pathological states. *Anal Biochem* 220, 284–296 (1994).
122. Worley, B. & Powers, R. Multivariate Analysis in Metabolomics. *Curr. Metabolomics* 1, 92–107 (2012).
123. Yan, Xin. Linear Regression Analysis: Theory and Computing, World Scientific (2009).
124. Su, X. et al. Linear regression. *Wiley Interdiscip. Rev. Comput. Stat.* 4, 275–294 (2012).
125. Poste, G. Bring on the biomarkers. *Nature* 469, 156–157 (2011).
126. Van, Q. N. & Veenstra, T. D. How close is the bench to the bedside? Metabolic profiling in cancer research. *Genome Med.* 1, 5 (2009).

## **Chapter 2: Application of an LC-MS/MS method for quantitative analysis of fourteen endogenous oestrogens in urine from premenopausal females: application to carriers of the CYP3A7\*1C allele (AC genotype).**

### **2.1 Introduction**

Epidemiological studies have shown a strong positive association between endogenous sex steroid levels (oestrogens, progesterone and testosterone) and breast cancer risk in postmenopausal women. The level of risk increases by two- to three-fold among women with blood or urinary oestrogen levels in the top 20% of the distribution (versus the bottom 20%) [1]. Increased concentrations of circulatory oestradiol ( $E_2$ ) in blood and oestrone ( $E_1$ ) in urine have been shown to be associated with a high risk of breast cancer in postmenopausal women [2]. However, data from laboratory and epidemiological studies are sparse and inconsistent for associations between premenopausal endogenous sex steroid levels and risk of breast cancer [3]. The only consistent finding is a positive association between free testosterone levels and breast cancer risk [1, 3]. To date, both positive and null associations have been observed between oestrogens and breast cancer risk in premenopausal women [3, 4, 5]. Thus, the role of premenopausal oestrogens in breast carcinogenesis in epidemiological studies is not thoroughly understood and remains unclear. This lack of data in premenopausal women is likely due to two reasons. Firstly, cyclical variation in hormone levels during the menstrual cycle, e.g. follicular versus luteal phase, leads to larger intra- and inter-individual variations, making capturing long-term average oestrogen levels in premenopausal women problematic. Secondly, inconsistency and non-standardisation of methods measuring absolute concentrations of oestrogens also contributes uncertainty, especially at lower levels. This also restricts the associations only to the parent or primary oestrogens ( $E_1$ ,  $E_2$  or their conjugate  $E_1$  sulfate), instead of specific patterns of individual oestrogen metabolites (EMs) profiles.

Since premenopausal circulating oestrogen levels are approximately ten times higher for  $E_2$  and around five times for  $E_1$  compared to postmenopausal levels, premenopausal oestrogen levels may be more relevant for assessing the risk of breast cancer [6]. Several studies have demonstrated that both pre- and postmenopausal circulating endogenous oestrogen levels are influenced by genetic factors, such as genetic polymorphisms in cytochrome P450 (CYP 450) genes and enzymes that govern the biosynthesis and catalysis of oestrogen metabolism [6, 7, 8, 9]. Thus, polymorphisms in CYP450 genes could be good candidates for breast cancer predisposition alleles [6, 10] and it is plausible that the genetically determined effects on metabolism may have an impact on individual health outcomes, especially for patients undergoing chemotherapy regimens which include CYP substrates. Recent studies which have

looked at potential genotype-phenotype correlations have shown positive associations between circulating plasma E<sub>2</sub> levels in postmenopausal women [8] and, plasma E<sub>1</sub> and E<sub>2</sub> levels with haplotype SNPs at the aromatase (*CYP19A1*) locus [11]. Another comprehensive study, The National Cancer Institute Breast and Prostate Cancer Cohort Consortium (BPC3) looked at the relationship between polymorphic variants in 36 candidate genes in the sex steroid pathway and serum concentrations of E<sub>1</sub>, E<sub>2</sub>, testosterone, dehydroepiandrosterone (DHEA), androstenedione hormones and SHBG in pre- and postmenopausal women. This study found a significant association between postmenopausal serum levels of E<sub>1</sub> and E<sub>2</sub> and both the *CYP19* and *ESR1* genes [12]. However, none of these studies found any association between hormone levels and breast cancer risk. One study which has looked at associations between genotype (SNPs) and the risk of breast cancer for e.g. SNPs which mapped to the *CYP3A* locus found that a rare *CYP3A4* SNP (rs4986910) was associated with a reduced risk of breast cancer, but found no association between this SNP and urinary oestrogen levels [13]. The BPC3 investigators screened five *CYP3A4* SNPs and found a promoter variant (*CYP3A4\*1B*, rs27404574) that was associated with a predisposition to breast cancer, but again no associations between SNPs, hormone levels and breast cancer risk were reported [12, 14].

The lack of associations between genetic variants (genotype) and phenotypes such as urinary hormone levels and the risk of breast cancer may be attributed to the fact that most research has focused on measuring the levels of the parent oestrogens (E<sub>1</sub> and E<sub>2</sub>) alone, instead of measuring individual EMs and pathways of oestrogen metabolism, which may be more relevant in relation to breast cancer. Individual EMs may have different roles in breast carcinogenesis as they exhibit differential oestrogenic and genotoxic activities. It has also been suggested that the genotoxic potential of oestrogen varies across individuals and may be influenced by genetic variation within the 2-catechol, methylated-catechol or 2- and 16-hydroxylation oestrogen metabolism pathways [15].

The present study used targeted LC-MS/MS metabolic profiling to explore the detailed snapshot profile of urinary endogenous oestrogen metabolism. Specifically, to account for cyclic variation of oestrogens, samples were taken at pre-specified days of the woman's menstrual cycle, and the series was selected such that there were equal numbers of women who were carriers of a rare variant *CYP3A7\*1C* allele and women who were non-carriers. This allele is tagged by a single SNP (rs45446698) mapping to the *CYP3A* locus, such that genotyping this SNP and comparing women with the rare (AC) genotype to women with the common (AA) genotype is equivalent to comparing *CYP3A7\*1C* carriers and non-carriers. The present study is based on an *a priori* hypothesised genotype-phenotype correlation reported by Johnson *et al.* [6, 61] showing an

association between the *CYP3A7\*1C* allele and levels of urinary oestrone glucuronide (E<sub>1</sub>G) (54% reduction).

### **2.1.1 Oestrogen metabolism**

Biosynthesis of all steroid hormones originates from C27 cholesterol and involves a series of enzymatic steps (Figure 2.1). The first step is the conversion of cholesterol to pregnenolone, which acts as a precursor for all steroid hormones. Pregnenolone is further converted to progesterone or androstenedione under the action of the *CYP17* gene. Androstenedione, in turn, is metabolised to other androgens. The biosynthesis of oestrogens from androgens is mediated by the action of a specific rate-limiting CYP 450 enzyme (CYP19) called aromatase. Members of Cytochrome P450 (CYP1, CYP2 and CYP3) family are the major enzymes which play a crucial role in both oestrogen formation and its subsequent oxidation to hydroxylated metabolites (Figure 2.2) [17].

Cytochrome P450s are a family of hemoproteins that are not only responsible for endogenous oestrogen metabolism but also for the metabolism of many drugs and environmental chemicals [17]. The main three families of CYP450s that are involved in the catalysis of endogenous and exogenous hormones are CYP1, CYP2 and CYP3. Of these, the CYP3A subfamily of enzymes is the main CYP discussed in this chapter, due to its relevance in the present study. CYP3A has been established as the most abundant P450 in humans, accounting for approximately 30% of total liver P450 content [18, 19].

In humans, the three major isoforms of CYP3As known to be expressed are CYP3A4, CYP3A5 & CYP3A7. CYP3A4 is the most abundant isoform in adult liver, and intestine and alone contributes to the metabolism of the largest percentage of clinical drugs [19]. Up to 60-fold interindividual variation has been observed in the hepatic expression of CYP3A4 and induction by xenobiotics and endogenous hormones further modulate CYP3A4 expression within individuals [20]. The substrate specificity of CYP3A5 is similar to that of CYP3A4 however, liver expression levels of CYP3A5 are much lower compared to CYP3A4. CYP3A5 is, however, polymorphically expressed in some individuals [21] with approximately 33% of individuals of Caucasian origin and 60% of individuals of African-American origin expressing detectable levels of CYP3A5 [21]. CYP3A7 is the major isoform in the foetal liver where it accounts for 30-50% of total CYP450. CYP3A7 expression was originally reported to decrease sharply, or stop completely shortly after birth. However, a study has recently reported that the CYP3A7 expression has been polymorphically detected between 54% to 88% of the adult livers [22].



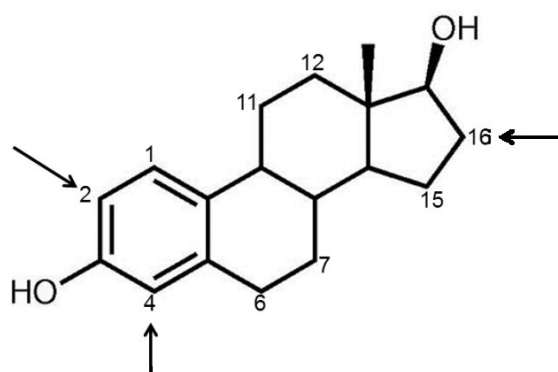


**Figure 2.2 Oxidative metabolism sites in oestradiol (E<sub>2</sub>) by phase I cytochrome P450 enzymes.** Hydroxylation sites of three major metabolic pathways of oestrone metabolism (2-hydroxylation, 4-hydroxylation and 16-hydroxylation) are indicated by the arrow. Diagram ref [16].

The metabolism of oestrogens mainly occurs in the liver where most CYP 450 isoforms are abundantly expressed, and some in specific extra-hepatic tissues (brain, lung and kidney) and endocrine-regulated tissues (breast, uterus and ovary). In terms of overall structure, all oestrogens as shown in Figure 2.2 are C<sub>18</sub> steroids, consisting of one benzene ring, a phenolic group at position C3, and a ketone group in the case of oestrone or a hydroxyl group in the case of oestradiol at position C17. In the human body, the main circulating oestrogens are oestrone (E<sub>1</sub>), oestradiol (E<sub>2</sub>), and oestriol (E<sub>3</sub>). Parent oestrogens (E<sub>1</sub> and E<sub>2</sub>) are reversibly converted under the action of 17 $\beta$ -hydroxysteroid dehydrogenase (17 $\beta$ -HSD). Although E<sub>3</sub> is the most abundant oestrogen in all premenopausal women, and the major oestrone during pregnancy, E<sub>2</sub> is the most biologically active and most important oestrogen in premenopausal women. E<sub>2</sub> is mainly synthesised by ovaries and the adrenal gland. A healthy premenopausal woman of reproductive age produces 700  $\mu$ g of E<sub>2</sub> daily until her menopause [16, 23, 24]. In postmenopausal women, by contrast, E<sub>1</sub> secreted by peripheral tissues is predominant [16].

Irreversible hydroxylation of parent oestrogens (E<sub>1</sub> and E<sub>2</sub>) occurs either at C2, C4, and C16 position to yield three competitive pathways as shown in Figure 2.1: 2-hydroxylation, 4-hydroxylation and 16-hydroxylation are catalysed via phase I CYP 450 enzymes (CYP1A1/2, CYP1B1 and CYP3A4) [16]. Hydroxylation at the C2 and C4 positions results in catechol oestrogens: 2-hydroxyoestrone (2-OHE<sub>1</sub>), 2-hydroxyoestradiol (2-OHE<sub>2</sub>) and 4-hydroxyoestrone (4-OHE<sub>1</sub>). These catechol oestrogens are further metabolised (methylation by the enzyme catechol-O-methyltransferase (COMT)) to five methoxy oestrones: 2-methoxyoestrone (2-MeOE<sub>1</sub>), 2-methoxyoestradiol (2-MeOE<sub>2</sub>), 4-methoxyoestrone (4-MeOE<sub>1</sub>), 4-methoxyoestradiol (4-MeOE<sub>2</sub>) and 2-hydroxyoestrone-3-methyl ether (3-MeOE<sub>1</sub>) [16, 24] (Figure 2.1). Hydroxylation of parent oestrogens at position C16 yields 16 $\alpha$ -hydroxyoestrone (16 $\alpha$ -OHE<sub>1</sub>)

which is further metabolised by reduction and oxidation at position C17 reversibly to form E<sub>3</sub> or its epimer 17-epiestriol (17-epiE<sub>3</sub>), or 16-ketooestradiol (16-ketoE<sub>2</sub>) [16, 24] (Figure 2.1).

In addition to oxidative metabolism (hydroxylation by CYP450s), and methylation (by catechol-O-methyltransferase, COMT); parent and catechol oestrogens are also conjugated by phase II enzymes, particularly glucuronidation by UDP-glucuronosyltransferase and sulfation by sulfotransferase [25, 26]. Conjugations by glucuronide and sulfate moieties represent potential detoxification reactions by which these hormones become more water soluble and are excreted in the urine or faeces [25]. The extended half-life of circulatory oestrogens may be due to sulfate conjugation while glucuronidation is considered as an important excretory pathway. In this chapter, the term 'total oestrogens' (Total EM) refers to the sum of the glucuronidated, sulfated and unconjugated forms of oestrogen metabolites.

#### **2.1.1.1 The 2-hydroxylation pathway**

Quantitatively, 2-hydroxylation is the predominant metabolic pathway compared to the 4- and 16- hydroxylation pathways [16, 23]. In humans, around 50% of endogenous oestrogens are hydroxylated at the C2 position and in the liver, approximately 80% of E<sub>2</sub> is metabolised to 2-hydroxyoestradiol, while around 20% is biotransformed to 4-hydroxyoestradiol [24]. The CYP450 enzymes *CYP1A1*, *CYP1A2* and *CYP3A4*, are a phase I enzymes which exhibit catalytic activity dominated by the C2 hydroxylation of parent oestrogens to their respective catechol oestrogen metabolites (Catechol EM) [24]. Metabolite products of the 2-hydroxylation pathway (2-OHE<sub>1</sub>, 2-OHE<sub>2</sub> and 4-OHE<sub>2</sub>) have been shown to possess reduced hormonal potency and low binding affinity for the oestrogen receptor compared to E<sub>2</sub> in cytosol from human breast cancer [16, 24, 27]. 2-catechol EMs has also been shown to exhibit both non-oestrogenic and anti-oestrogenic activities. *In vitro* studies have shown 2-hydroxylated metabolites, especially 2-hydroxyoestrone, to be associated with normal cell differentiation and apoptosis [28, 29]. Owing to this collective evidence of the protective role of 2-hydroxyoestrone, it is sometimes referred to as 'good oestrone' in the literature [30]. Methylated products of 2-catechol oestrogen, especially 2-methoxyoestradiol have been shown to be an anti-proliferative metabolite [31] and a promising anticancer agent by exhibiting potent apoptotic activity against rapidly growing tumour cells [32].

The protective role of 2-hydroxylated metabolites against tumorigenic activity is attributed to the high rate of O-methylation by *COMT* which may increase the rate of clearance of oestrogens in target tissues, resulting in lower hormonal potency. When the functioning of *COMT* is inhibited, 2-hydroxyoestrogens are shown to cause DNA damage and generate free radicals by redox cycling between catechol oestrogens [16, 33]. The five metabolites of 2-hydroxylation pathway measured

in this study were: 2-hydroxyoestrone (2-OHE<sub>1</sub>), 2-hydroxyoestradiol (2-OHE<sub>2</sub>), 2-hydroxyoestrone-3-methyl ether (3-MeOE<sub>1</sub>), 2-methoxyoestrone (2-MeOE<sub>1</sub>) and 2-methoxyoestradiol (2-MeOE<sub>2</sub>).

#### **2.1.1.2 The 4-hydroxylation pathway**

Quantitatively, metabolites of the 4-hydroxylation pathway are substantially less abundant compared to metabolites of the 2- and 16-hydroxylation pathways. The CYP1B1 in human breast and uterus tissue is the primary enzyme of C4 hydroxylation of oestradiol while CYP3A4/A5 exhibit catalytic activity specifically for the 4-hydroxylation of oestradiol in human liver microsomes [16, 24, 26]. An *in vitro* study has shown that the oxidation of 4-catechol oestrogens (4-hydroxyoestrone and 4-hydroxyoestradiol) to catechol oestrogen-3,4-quinones binds to DNA to form depurination adducts [34]. Depurinating adducts of 4-hydroxylated catechols are considered of carcinogenic potential due to their ability to cause DNA damage and thus promote breast cancer [16, 29]. On the other hand, O-methylation of 2-hydroxyoestradiol and 4-hydroxylated catechol's catalysed by COMT (a Phase II metabolic inactivation pathway for catechol oestrogens) to methylated 4-catechol (4-methoxyoestrone and 4-methoxyoestradiol) has been shown to be protective against oxidative DNA damage in oestradiol-treated MCF-7 cells [35]. This finding agreed with a study in which inhibition of COMT activity in MCF-10 cell lines increased the formation of depurination adducts from 4-catechol oestrogens and blocked the formation of methylated 4-catechol which can be critical factors leading to initiation of breast cancer [36]. The three metabolites of the 4-hydroxylation pathway measured in this study were: 4-hydroxyoestrone (4-OHE<sub>1</sub>), 4-methoxyoestrone (4-MeOE<sub>1</sub>) and 4-methoxyoestradiol (4-MeOE<sub>2</sub>).

#### **2.1.1.3 The 16-hydroxylation pathway**

16-hydroxylation pathway metabolites are catalysed by several CYP450 isoforms (CYP1A1/A2, and CYP3A4/A5/A7). Studies investigating CYP450 enzymes that catalyse metabolism of oestrogen to 16 $\alpha$ -hydroxyoestrone have shown the important role of CYP3A4/7 enzymes in humans [37, 38, 39]. The most important metabolite of the 16-hydroxylation pathway is 16 $\alpha$ -hydroxyoestrone. Several lines of evidence have implicated 16 $\alpha$ -hydroxyoestrone as a potential carcinogenic metabolite of oestrogen. In animal studies, increased urinary concentrations of 16 $\alpha$ -hydroxyoestrone were associated with the higher proliferation of mammary cells [40] and mammary tumour incidence [41]. A study on human breast tissue investigating the levels of 16 $\alpha$ -hydroxyoestrone in relation to risk of breast cancer reported an eight-fold increase in levels in cancerous mammary duct lobular unit compared to mammary fat tissue, suggesting that higher levels of 16 $\alpha$ -hydroxyoestrone may induce breast cancer [42]. The four metabolites of 4-

hydroxylation pathway measured in this study were: 16 $\alpha$ -hydroxyoestrone (16 $\alpha$ -OHE<sub>1</sub>), oestriol (E<sub>3</sub>), 17-epiEstriol (17-epiE<sub>3</sub>) and 16-ketooestradiol (16-ketoE<sub>2</sub>).

It has also been hypothesised that the ratio of urinary 2-hydroxyoestrone to 16 $\alpha$ -hydroxyoestrone could be used as a biomarker of breast cancer risk. This means that the higher the ratio, the higher the risk and that the oestrogen metabolism pathway favouring 2-hydroxylation over 16 $\alpha$ -hydroxylation is associated with a reduced risk. However, the evidence to support this hypothesis has been inconsistent in prospective premenopausal studies [43, 44]. The role of 16 $\alpha$ -hydroxyoestrone in carcinogenesis remains controversial as it is present in abundance during normal pregnancy and the risk of breast cancer is considered low during pregnancy [24, 25].

### **2.1.2 The role of genetic variation in oestrogen metabolism and risk of breast cancer**

Many of the CYP 450 enzymes are involved in the biosynthesis and transformation of oestrogens. It has been postulated that single nucleotide polymorphisms (SNPs) in the genes that code for these enzymes are associated with risk of breast cancer. Some studies have suggested that SNPs mapping to these genes influence oestrogen metabolism, altering the circulating levels of E<sub>1</sub>, E<sub>2</sub>, 2-OHE<sub>1</sub> and 16 $\alpha$ -OHE<sub>1</sub> metabolites in postmenopausal women and, thereby increasing breast cancer susceptibility. Polymorphic variations in genes coding CYP3As have received extensive attention as they play a vital role in the oxidative, peroxidative, and reductive metabolism of diverse range of substrates including not only endogenous steroids (e.g. oestrogens), but also procarcinogens (e.g. 2,3,7,8-Tetrachlorodibenzo-p-dioxin TCDD), and 50% of all clinically used drugs [18, 20, 21, 22, 50].

The nuclear receptors which are known to regulate the induction of CYP3A isoforms are Pregnane X receptors (PXR) and constitutive androstane receptor (CAR) [22, 45]. A study investigating the role of CYP3A7 in relation to total CYP3A activity, identified the *CYP3A7\*1C* allele as a consistent marker of increased *CYP3A7* expression, both in liver and intestine; this allele is present in approximately 11% of subjects [22]. It was shown that this allele could serve as a marker for increased CYP3A activity. One of the possible mechanisms responsible for the high CYP3A7 expression in *CYP3A7\*1C* carriers was attributed to replacement of part of the CYP3A7 promoter by the corresponding region of CYP3A4, which led to the substitution of CYP3A7-ER6 for CYP3A4-ER6. The ER6 motif (everted repeat of hexanucleotide motif with a spacer of 6 base pairs) in the proximal promoter of the CYP3A4 gene was shown to be one of the elements mediating PXR-dependent activation.

For the human CYP3A genes, over 30 CYP3A4 alleles have been described in the literature, including 18 that are associated with non-synonymous mutations [46]. For the CYP3A4 gene, the most extensively studied allele is the *CYP3A4\*1B* allele. This allele carries a mutation in the *CYP3A4* promoter region and has been associated with an increased risk of prostate cancer, however, no association between the *CYP3A4\*1B* allele and the risk of breast or ovarian cancer has been reported [47, 48]. Therefore, the evidence of an association between genetic polymorphisms in the CYP3A4 gene and risk of cancer remains inconsistent. For *CYP3A5*, at least 11 different alleles have been identified, of which the *CYP3A5\*3* allele has been most extensively studied and has been shown to have reduced enzymatic activity due to alternative splicing and protein truncation. However, in the one [26] study that tested for an association with disease directly, no association with risk of developing myeloid leukaemia was reported. For the *CYP3A7* gene, four variant alleles (*CYP3A7\*2*, *CYP3A7\*3*, *CYP3A7\*1B* and *CYP3A7\*1C*) have been described, but none of them (excluding *CYP3A7\*1C*) has been studied in relation to breast cancer risk [49, 6, 61].

Further studies with larger prospective homogeneous populations will be required to substantiate the role of polymorphisms in the CYP3A genes and to characterise some of the functional metabolic effects that they may have on circulating levels of individual oestrogen metabolites, breast cancer risk and, potentially, drug metabolism.

### **2.1.3 A Liquid chromatography-tandem mass spectrometry (LC-MS/MS) method to quantify endogenous oestrogen metabolites (EMs)**

Despite multiple hypotheses in experimental, epidemiological and clinical research suggesting the role of endogenous oestrogens in the aetiology of breast cancer, the contribution of specific oestrogen metabolite profiles and pathways is much more ambiguous and until recently remained unexplored in epidemiological studies. The first eight prospective cancer studies (three premenopausal and five postmenopausal) in urine or serum samples investigated the breast cancer risk with only parent oestrogens ( $E_1$  and  $E_2$ ), the 2-hydroxyoestrone to 16 $\alpha$ -hydroxyoestrone ratio [51]. All these studies used enzyme immune assay (EIA) and the results were inconsistent and generally non-significant, lacking specificity and sensitivity especially at the lower concentrations characteristic of postmenopausal women. To date, commercially available radioimmunoassay (RIA), enzyme-linked immunosorbent assays (ELISA) and EIA are commonly used to measure selective oestrogen metabolites ( $E_1$ ,  $E_2$ , and their conjugated forms, 2-hydroxyoestrone and 16 $\alpha$ -hydroxyoestrone) in both urine and serum samples because of their efficiency, simplicity and low-cost.

However, the use of RIA and EIA assays is shown to suffer from a lack of specificity, accuracy and precision due to cross-reactivity and batch to batch variation of antibodies [52, 53]. Some of the other limitations of using these assays, including ELISA, is that typically each hormone must be assayed independently and the sensitivity of measuring E<sub>2</sub> in postmenopausal levels is low [51, 55]. More importantly, these assays cannot measure a wide range of individual EMs to evaluate more specific hypotheses about the role of oestrogen metabolism in breast cancer causation. A review on standardisation of steroid hormone assays [53] discussed the use of mass-spectrometry-based techniques as the most promising approach for improving accuracy and precision of steroid methods and referred to it as a 'gold standard' against which EIA and ELISA results should be compared.

To facilitate and further explore the contribution of individual oestrogen metabolites to cancer aetiology, Xia Xu et al. [52] developed a method for absolute quantification of 15 endogenous oestrogen metabolites in urine and serum samples using high-performance liquid-chromatography-tandem mass spectrometry (HPLC-MS/MS). This method was later published in Nature Protocols in 2007 [54]. The method involves hydrolysis of oestrogen conjugates (glucuronides and sulfates) from EMs for measurement of total (glucuronidated + sulfated + unconjugated) concentrations of each EM. If the hydrolysis step is omitted, the method is only suitable for measuring unconjugated EMs. This method has been shown to be accurate, precise and sensitive enough to quantify 15 EMs not only in pre-and postmenopausal women but also in men [53]. Since then this LC-MS/MS method was independently applied in multiple laboratories and used in seven prospective epidemiological studies of oestrogen metabolism and breast cancer (six in postmenopausal urine or serum and one in premenopausal urine) [51].

The results from these prospective breast cancer studies were quite consistent. The three postmenopausal prospective breast cancer case-control cohorts (the prostate, Lung, Colorectal, and Ovarian cancer screening Trial (PLCO) [55], Missouri Serum Bank cohort in Columbia [56], Breast and Bone Follow-up to the Fracture Intervention Trial (B-FIT) [57]) and one premenopausal breast cancer case-control cohort (the Nurses Healthy Study II USA (NHS II) [58]) looked at the association of individual, grouped EMs, their pathways and selected ratios in relation to breast cancer risk. The results from these four studies concurred, suggesting the enhanced 2-hydroxylation pathway to be protective against breast cancer risk.

To date, the LC-MS/MS assay capable of measuring 15 endogenous EMs has been used only in case-control breast cancer studies. Therefore, the current study is the first to use the method of Xu et al. [54] to investigate an association between specific oestrogen metabolites and carrier status for the *CYP3A7\*1C* allele in healthy premenopausal women.

### 2.1.4 Hypothesis and Aim

The present study is based on an *a priori* hypothesised genotype-phenotype correlation reported by Johnson *et al.* [6, 61] showing an association between the *CYP3A7\*1C* allele and levels of urinary oestrone glucuronide (E<sub>1</sub>G) (54% reduction). This same allele, tagged by the SNP rs45446698, was also associated with adverse outcomes in breast, lung cancers and chronic lymphatic leukaemia (CLL). The results from Johnson *et al.* were the first to report a functional effect of a *CYP3A* haplotype on urinary E<sub>1</sub>G levels and to further test for an association between *CYP3A* genotype and outcome in breast cancer, lung cancer and CLL patients.

The hypothesis of the present study was that there will be an association between a *CYP3A7\*1C* allele and the urinary levels of specific oestrogen metabolites or oestrogen metabolic pathways. The aim was to determine whether the *CYP3A7\*1C* allele (tagged by the SNP, rs45446698) influences specific urinary EMs and/or their metabolic pathways. As the role of a *CYP3A7\*1C* allele on the circulating urinary oestrogen metabolites was unknown.

The strength and novelty of the current study were that it used targeted LC-MS/MS metabolic profiling to enhance the depth of the profile of urinary endogenous oestrogen metabolism and provide a more detailed snapshot of this intermediate phenotype in a carefully selected series of premenopausal women.

### 2.1.5 Objectives

The objectives of the present study were:

- To implement a precise LC-MS/MS method to quantify endogenous oestrogen metabolite (EM) levels in urine.
- To measure the absolute concentrations and calculate relative concentrations of individual EMs and explore the patterns of grouped oestrogen metabolites, their pathways and selected ratios in follicular phase urine of healthy premenopausal women.
- To investigate the role of the *CYP3A7\*1C* allele, tagged by rs45446698 in regulating levels of individual, grouped oestrogen metabolites and their pathways in healthy premenopausal women.
- To demonstrate that measuring individual EMs and calculating the levels of their pathways provides additional information about the pattern of oestrogen metabolism which is beyond just assessing parent oestrogens and, to explore the correlations between parent EMs and non-parent EMs.

- To understand the wider implications of this work and potential of targeted LC-MS/MS endogenous oestrogen metabolic profiling to improve the power of epidemiological studies to detect genotype-phenotype associations.

## 2.2 Materials and Methods

### 2.2.1 Materials

At the time of this work, one molecular standard used in the original work by Xu *et al.* was not commercially available and therefore the method implemented here sought to measure fourteen oestrogen metabolites (EM) specifically: oestrone ( $E_1$ ), oestradiol ( $E_2$ ), 2-hydroxyoestrone (2-OHE<sub>1</sub>), 2-hydroxyoestradiol (2-OHE<sub>2</sub>), 2-methoxyoestrone (2-MeOE<sub>1</sub>), 2-methoxyoestradiol (2-MeOE<sub>2</sub>), 2-hydroxyoestrone-3-methyl ether (3-MeOE<sub>1</sub>), 4-hydroxyoestrone (4-OHE<sub>1</sub>), 4-methoxyoestrone (4-MeOE<sub>1</sub>), 4-methoxyoestradiol (4-MeOE<sub>2</sub>), 16 $\alpha$ -hydroxyoestrone (16 $\alpha$ -OHE<sub>1</sub>), 17-epiestriol (17-epiE<sub>3</sub>), oestriol ( $E_3$ ) and 16-ketooestradiol (16-ketoE<sub>2</sub>). All EM standards were purchased from Steraloids Inc. (United States). Five deuterium-labelled oestrogen metabolite (d-EM) standards consisting of oestradiol-2,4,16,16-d<sub>4</sub> (d4-E<sub>2</sub>; 98% labelled), estriol-2,4,17-d<sub>3</sub> (d3-E<sub>3</sub>; 98% labelled), 2-hydroxyoestradiol-1,4,16,16,17-d<sub>5</sub> (d5-2-OHE<sub>2</sub>; 98% labelled), 2-methoxyoestradiol-1,4,16,16,17-d<sub>5</sub> (d5-2-MeOE<sub>2</sub>; 98% labelled) and 16 $\beta$ -hydroxy-17 $\beta$ -oestradiol-2,4-d<sub>2</sub> (D<sub>2</sub>-16-epiE<sub>3</sub>; 98% labelled) were obtained from Qmx Laboratories Ltd. (United Kingdom). Reported chemical and isotopic purity of both EM and d-EM standards were >98%. Other solvents and reagents including dichloromethane (99.9%), acetone (99.9%), methanol (99.9%), LC-MS ultra CHROMASOLV® Water, 0.1M sodium bicarbonate (pH 9.0), acetic acid glacial ( $\geq 99.5\%$ ), sodium acetate trihydrate ( $\geq 99.0\%$ ), formic acid (LC-MS Ultra) and ascorbic acid (99.5%) were all HPLC grade and purchased from Sigma-Aldrich. Hydrolytic enzyme  $\beta$ -Glucuronidase from *Helix pomatia* (Type HP-2) and derivatisation reagent dansyl chloride ( $\geq 99.0\%$ , HPLC grade) were also obtained from Sigma-Aldrich. Charcoal stripped human urine with no detectable levels of EM was bought from Golden West Biological, Inc. (US) to prepare calibration standards and quality control samples.

The HPLC-ESI-MS/MS instrument consisted of an AB SCIEX QTRAP®6500 (Applied Biosystems) mass spectrometer coupled with an Agilent 1100 HPLC system (Agilent Technologies). Both the HPLC and the MS were controlled by Analyst® version 1.6.2 Software (Applied Biosystems). The HPLC system consisted of two Agilent 1100 Capillary Pumps and an Agilent 1100 Micro Wellplate autosampler. Chromatographic separation was carried out on a Polar end-capped C18 column (Phenomenex): 150 mm X 2.0 mm packed with 4 mm particle size Synergi Hydro-RP particles. Other equipment required were: Direct heat CO<sub>2</sub> incubator from Lab Implex Research Ltd, Shaker (Denley spiramix 10; Jencons-Pls) and Heating block (Techne DRI-Block®, DB-2A).

### 2.2.2 Enzymatic hydrolysis buffer and derivatisation reagent

The enzymatic hydrolysis buffer (0.15 M sodium acetate, pH 4.6) was prepared fresh for each analytical batch. A stock solution was prepared in 100 mL LC-MS water by adding 0.85g sodium acetate trihydrate plus 500  $\mu$ L of glacial acetic acid and the pH was measured to get the desired pH 4.6. From this stock solution, 20mL was aliquoted into a new scintillation vial and 100 mg of ascorbic acid plus 200  $\mu$ L of  $\beta$ -Glucuronidase were added right before using the buffer. For the derivatisation step pH of ready to use 0.1 M sodium bicarbonate pH 9.0, was measured to confirm the desired pH of 9.0 and dansyl chloride solution (1mg/mL acetone) was prepared fresh before use.

### 2.2.3 Preparation of calibration curve standards and quality control samples

Stock and working standard solutions: Stock solutions of 1 mg/mL for five d-EM were prepared in methanol containing 0.1% (wt/vol) ascorbic acid. Working d-EM internal standard mixture solutions was prepared by taking 100  $\mu$ L of each 1 mg/mL d-EM stock solution (total volume 500  $\mu$ L; concentration 200 $\mu$ g/mL) in a new scintillation vial. 500  $\mu$ L of (200  $\mu$ g/mL) was added to 1.5 mL of methanol +0.1% ascorbic acid to make 50  $\mu$ g/mL. Further serial dilution was carried in methanol + 0.1% ascorbic acid to get the desired 0.05  $\mu$ g/mL d-EM working internal standard mixture solution before the analysis.

Similarly, stock solutions of 1mg/mL for each of the EM non-deuterated standards were prepared in methanol containing 0.1% (wt/vol) ascorbic acid. Working standard mixture solutions were prepared by taking 100  $\mu$ L of stock solution from each of fourteen EM (total volume 1.4mL) and adding 600  $\mu$ L methanol + 0.1% ascorbic acid to obtain a final concentration of 50  $\mu$ g/mL. Further serial dilution was carried out in methanol containing 0.1% ascorbic acid to obtain working standard mixture solutions with following concentrations of 5  $\mu$ g/mL, 0.5  $\mu$ g/mL, 0.05  $\mu$ g/mL and 0.005  $\mu$ g/mL before the sample preparation. All stock and working standard solutions were stored at -20°C.

For matrix matching and to compensate for matrix effects that influence analytical response, charcoal stripped human urine with no detectable levels of EMs + 0.1% (wt/vol) ascorbic acid was used to prepare calibration standards and quality control samples. To correct for the loss of urinary EMs during sample preparation and HPLC-MS/MS analysis, d-EM internal standards were added to all the samples. 20  $\mu$ L of the d-EM working internal standard mixture solution (2.0 ng of each d-EM) was added to 500  $\mu$ L of calibration standards, quality control samples and urine samples. Table 2.1 shows the concentration of 20  $\mu$ L of the d-EM internal standard solution in 500  $\mu$ L of calibration samples, quality control samples and urine samples. For calibration

standards ranging over eight calibration points (0.08 ng/mL to 20 ng/mL of each EM), 20  $\mu$ L of EM working standard solution mixture was added to 500  $\mu$ L of charcoal-stripped human urine + 0.1% ascorbic acid. Table 2.2 shows the concentrations of 20  $\mu$ L of EM working standard mixture solutions in 500  $\mu$ L of calibration samples. Similarly, Quality control samples were also prepared in charcoal stripped human urine + 0.1% ascorbic acid at the concentration of 2 ng/mL of each EM.

| d-EM Internal standards              | Concentration of 20 $\mu$ L of d-EM internal standard solution in 500 $\mu$ L of calibration samples, quality controls and urine samples (ng/mL) | Starting from d-EM working internal standard mix ( $\mu$ g/mL) |
|--------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------|
| d <sub>4</sub> -E <sub>2</sub>       | 2.0                                                                                                                                              | 0.05                                                           |
| d <sub>3</sub> -E <sub>3</sub>       |                                                                                                                                                  |                                                                |
| d <sub>2</sub> -16-epiE <sub>3</sub> |                                                                                                                                                  |                                                                |
| d <sub>5</sub> -2-MeOE <sub>2</sub>  |                                                                                                                                                  |                                                                |
| d <sub>5</sub> -2-OHE <sub>2</sub>   |                                                                                                                                                  |                                                                |

**Table 2.1 Concentration of Deuterated oestrogen metabolites (d-EM) internal standard solution in all the samples.**

| Calibration standard points | Concentration of 20 $\mu$ L EM standard solution mixture in 500 $\mu$ L charcoal stripped urine (ng/mL) | Starting from EM working standard mix ( $\mu$ g/mL) | Volume of EM working standard mix ( $\mu$ L) | Volume of MeOH + 0.1% ascorbic acid ( $\mu$ L) |
|-----------------------------|---------------------------------------------------------------------------------------------------------|-----------------------------------------------------|----------------------------------------------|------------------------------------------------|
| Cal1                        | 0.08                                                                                                    | 0.005                                               | 400                                          | 600                                            |
| Cal2                        | 0.2                                                                                                     | 0.005                                               | 1000                                         | 0                                              |
| Cal3                        | 0.4                                                                                                     | 0.05                                                | 200                                          | 800                                            |
| Cal4                        | 0.8                                                                                                     | 0.05                                                | 400                                          | 600                                            |
| Cal5                        | 2.0                                                                                                     | 0.05                                                | 1000                                         | 0                                              |
| Cal6                        | 4.0                                                                                                     | 0.5                                                 | 200                                          | 800                                            |
| Cal7                        | 8.0                                                                                                     | 0.5                                                 | 400                                          | 600                                            |
| Cal8                        | 20.0                                                                                                    | 0.5                                                 | 1000                                         | 0                                              |

**Table 2.2 Concentrations of oestrogen metabolite (EM) standard solution mixtures in eight calibration samples**

#### 2.2.4 Sample preparation: Hydrolysis, Extraction and Derivatisation

The schematic methodology for quantitative measurement of 14 EM is shown in Figure 2.3 (A). Due to the non-availability of commercial internal standard for one oestrogen metabolite at the time of this work the method implemented measures 14 EM. In each analytical batch, one matrix blank, one matrix blank spiked with internal standards, eight calibration standards, three quality control samples and twenty urine samples were hydrolysed, extracted and derivatised by following the same procedure. Since endogenous oestrogens and their metabolites are mostly excreted in urine as glucuronide and sulfate conjugates [25], an initial hydrolysis step was included. The source of enzyme used for hydrolysis was  $\beta$ -glucuronidase [minimum 100,000

units/mL glucuronidase], extracted from the terrestrial snail *Helix pomatia*, which also has some sulfatase activity [up to 7,500 units/mL sulfatase] as per manufacturer (Sigma) guidelines. In the derivatisation step, the reactive phenolic hydroxyl group of all oestrogen metabolites were attached to large, easily ionisable positively charged dansyl [5-(dimethylamino) naphthalene-1-sulfonyl] (dansyl chloride) moiety. Derivatisation was performed because nonpolar, uncharged, lipophilic compounds such as unconjugated steroids are hardly detectable by Electrospray ionisation (ESI). Dansylation enhances the detection limits significantly and increases the sensitivity of these EMs by improving their ionisation characteristics. Figure 2.3 (B) shows a molecular structure example of oestrone-glucuronide and chemical reactions (hydrolysis and derivatisation) during sample preparation to form and analyse oestrone-dansyl.

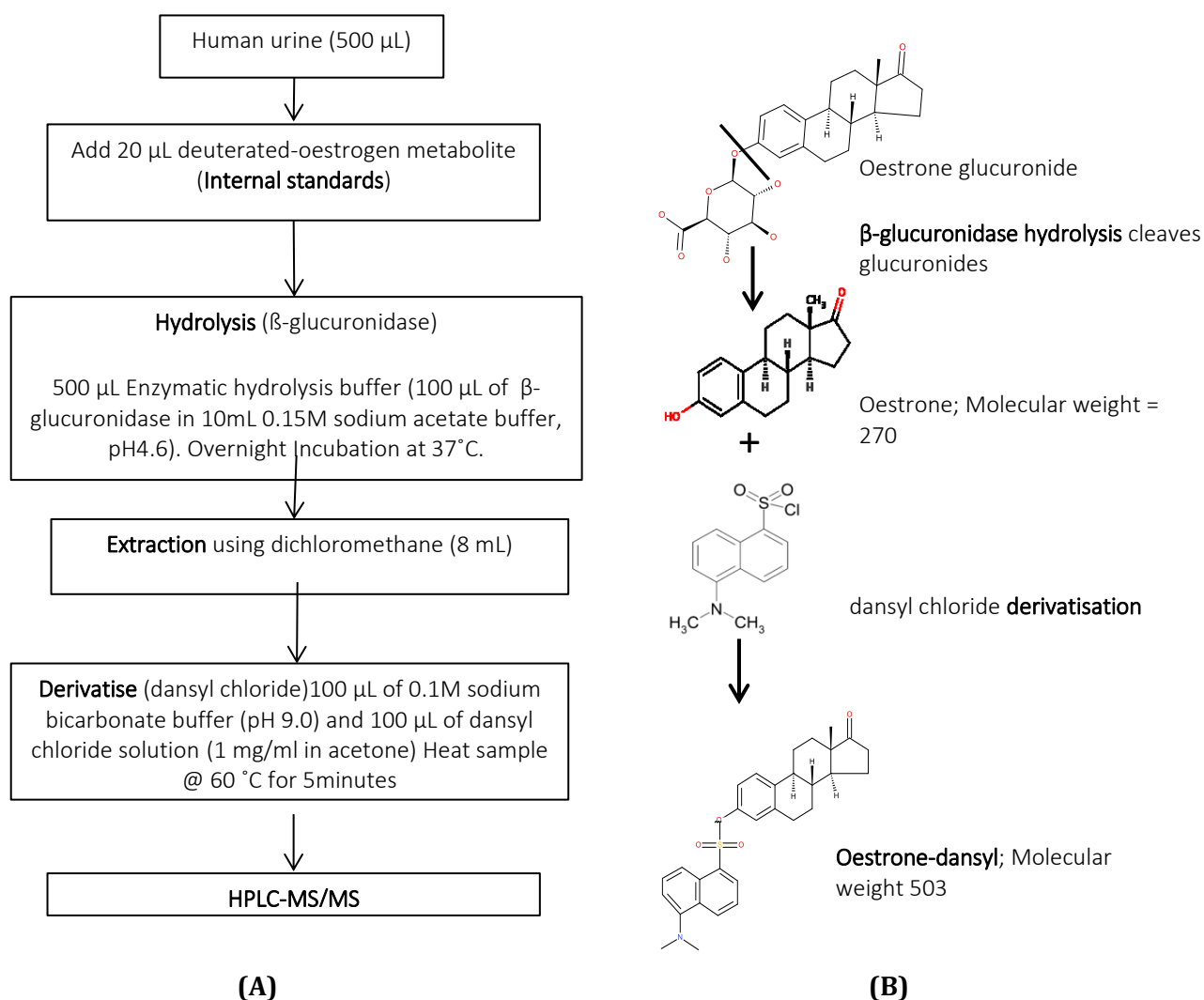


**Figure 2.3 Overall method synopsis for the quantitative measurement of 14 endogenous oestrogen metabolites in urine by HPLC-ESI-MS/MS. (B) Molecular structure example of Oestrone-glucuronide showing hydrolysis and derivatisation reactions to analyse Oestrone-dansyl.**

d-EM). For the hydrolysis step, 500  $\mu$ L of freshly prepared enzymatic hydrolysis buffer (100  $\mu$ L of  $\beta$ -glucuronidase from *Helix pomatia* (Type H-2) in 10 mL 0.15 M sodium acetate buffer, pH4.6 containing 2 mg of ascorbic acid) was added. The samples were then incubated overnight (approximately 20 hours) at 37°C. For the extraction step, 8 mL of dichloromethane was added to the samples to carry out slow and continuous inverse extraction for 30 minutes on a shaker plate. Following extraction, the aqueous layer was discarded and the organic solvent portion (approximately 5 mL) was transferred into a fresh (7 mL) scintillation vial, this organic solvent was then evaporated to dryness under a constant stream of nitrogen gas at 60°C. The dried samples were suspended by adding 100  $\mu$ L of 0.1 M sodium bicarbonate buffer (pH at 9.0) and 100  $\mu$ L of dansyl chloride solution (1 mg/mL in acetone), the samples were vortexed for 10 seconds and heated for 5 minutes at 60°C to form the dansylated derivatives of the non-deuterated EMs and d-EMs (EM-dansyl and d-EM-dansyl, respectively). The final derivatised samples (200  $\mu$ L) were transferred to HPLC vials and analysed using HPLC-MS/MS. To conduct formal proof of performance of this urinary EM method, in total 120 urine samples were analysed as randomised blinded aliquots across six batches. Each batch consisted of a reagent blank (charcoal stripped human urine without d-EM internal standards), blank (charcoal stripped human urine spiked with d-EM internal standards), 8-point calibration curve, 3-Quality controls and 20 urine samples.

### **2.2.5 High-Performance Liquid Chromatography- Tandem Mass Spectrometry (HPLC-MS/MS) analysis**

Quantification of the EMs was performed using electrospray HPLC-MS/MS platform for multiple reaction monitoring (MRM) mode. An AB SCIEX Triple Quad™ 6500 (Triple Quadrupole) mass spectrometer was operated in positive ionisation mode; both the HPLC and MS were controlled with the Analyst version 1.6.2 software. High-Performance Liquid Chromatography was carried using a reversed phase- polar end-capped C<sub>18</sub> column (Phenomenex): 150 mm X 2.0 mm packed with 4  $\mu$ m particle size Synergi Hydro connected to an Agilent 1100 HPLC system. According to manufacturer's instructions, this column is well suited to separate extremely hydrophobic compounds. It is also stable under high percentages of organic mobile phase, resulting in the separation of analytes eluting at a higher percentage of organic mobile phase and consequently improving sensitivity. Furthermore, this column was also recommended by Xia Xu. *et al.* [52] in their protocol to achieve optimal separation for oestrogen metabolites.

The column temperature was maintained at 40°C and injection volume was 10  $\mu$ L. The mobile phases consisted of water + 0.2% formic acid as mobile phase A and methanol + 0.2% formic acid as mobile phase B operating at a flow rate of 250  $\mu$ L/minute. A linear gradient elution was

performed for the analysis of EM-dansyl and d-EM-dansyl, changing the A/B mobile phase ratios from 35:65 to 15:85 in 66 minutes (Table 2.3).

| Step | Total Time(min) | Flow Rate( $\mu$ l/min) | A (%) | B (%) |
|------|-----------------|-------------------------|-------|-------|
| 0    | 10              | 250                     | 35    | 65    |
| 1    | 0               | 250                     | 35    | 65    |
| 2    | 60              | 250                     | 15    | 85    |
| 3    | 61              | 250                     | 15    | 85    |
| 4    | 64              | 250                     | 15    | 85    |
| 5    | 65              | 250                     | 35    | 65    |
| 6    | 66              | 250                     | 35    | 65    |

**Table 2.3 Gradient elution with mobile phase A (Water + 0.2% formic acid) and mobile phase B (methanol +0.2% formic acid), total run time 77 minutes per sample.**

The column was washed with 100% mobile phase A for 5 minutes to remove late-eluting substances; it was then re-equilibrated to A/B composition of 35:65 for 10 minutes prior to the next injection. The total analysis time including equilibration was 77 minutes. Samples were stored at 6°C in the autosampler throughout the analysis.

The ESI source (type: Ion Drive Turbo V source) parameters were set to - TEM 500°C, Curtain Gas 45psi, GS1 50psi, GS2 60psi, Interface heater on, MS parameters were CAD gas Medium. A scheduled detection method was used; the MRM detection window was 240sec with a target scan time of 1sec. The 14 chemical reference standards were directly injected into the MS (individually) and their precursor masses were monitored to get the m/z for the first and second product ions. To receive best signal sensitivity for each EM following parameters were optimised by manipulating values of collision energy (CE) 45 V, entrance potential (EP) 10V, declustering potential (DP) 80V, and collision exit potential (CXP) 5V.

For the protonated [MH]<sup>+</sup> EM-Dansyl and d-EM-dansyl metabolites of interest, the precursor to first and second product ion transitions along with their retention time are shown in Table 2.4.

| Transitions monitored                                   |                                        |                                        |                                        |                          |
|---------------------------------------------------------|----------------------------------------|----------------------------------------|----------------------------------------|--------------------------|
| Oestrogen metabolites-Dansyl                            | Precursor ion [M+H] <sup>+</sup> (m/z) | Product ion 1 [M+H] <sup>+</sup> (m/z) | Product ion 2 [M+H] <sup>+</sup> (m/z) | Retention time (minutes) |
| Oestrone (E <sub>1</sub> )                              | 504                                    | 171                                    | 156                                    | 44.10                    |
| Oestradiol (E <sub>2</sub> )                            | 506                                    | 171                                    | 156                                    | 47.00                    |
| 2-Hydroxyoestrone (2-OHE <sub>1</sub> )                 | 753                                    | 170                                    | 519                                    | 54.80                    |
| 2-Hydroxyoestradiol (2-OHE <sub>2</sub> )               | 755                                    | 170                                    | 521                                    | 56.90                    |
| 2-Methoxyoestrone (2-MeOE <sub>1</sub> )                | 534                                    | 171                                    | 156                                    | 40.60                    |
| 2-Methoxyoestradiol (2-MeOE <sub>2</sub> )              | 536                                    | 171                                    | 156                                    | 43.80                    |
| 2-Hydroxyoestrone-3-methyl ether (3-MeOE <sub>1</sub> ) | 534                                    | 171                                    | 156                                    | 37.10                    |
| 4-Hydroxyoestrone (4-OHE <sub>1</sub> )                 | 753                                    | 170                                    | 519                                    | 57.00                    |
| 4-Methoxyoestrone (4-MeOE <sub>1</sub> )                | 534                                    | 171                                    | 156                                    | 43.40                    |
| 4-Methoxyoestradiol (4-MeOE <sub>2</sub> )              | 536                                    | 171                                    | 156                                    | 45.70                    |
| 16 $\alpha$ -Hydroxyoestrone (16a-OHE <sub>1</sub> )    | 520                                    | 171                                    | 156                                    | 27.30                    |
| 17-Epiestriol (17-epiE <sub>3</sub> )                   | 522                                    | 171                                    | 159                                    | 36.10                    |
| Oestriol (E <sub>3</sub> )                              | 522                                    | 171                                    | 156                                    | 27.40                    |
| 16-Ketooestradiol (16-ketoE <sub>2</sub> )              | 520                                    | 171                                    | 156                                    | 27.30                    |
| D4-E <sub>2</sub>                                       | 510                                    | 171                                    | 156                                    | 47.30                    |
| D3-E <sub>3</sub>                                       | 525                                    | 171                                    | 115                                    | 27.60                    |
| D2-16-epiE <sub>3</sub>                                 | 524                                    | 171                                    | 156                                    | 34.70                    |
| D5-2-MeOE <sub>2</sub>                                  | 541                                    | 171                                    | 156                                    | 43.90                    |
| D5-2-OHE <sub>2</sub>                                   | 760                                    | 170                                    | 526                                    | 57.00                    |

**Table 2.4 Multiple reaction monitoring (MRM) transitions monitored for the protonated [M+H]<sup>+</sup> EM-dansyl and d-EM-dansyl metabolite and their retention times.**

## 2.2.6 Quantification of oestrogen metabolites

Analyst® 1.6.2 software was used for the quantitation of EM in urine. Blank samples (reagent blank and method blank) were checked for any contamination such as contamination in solvents during sample preparation, mobile phase contamination, to check for any carry over any impurity in reference standards. Two transition ions for each metabolite were monitored. The m/z of product ion 1 was used for quantification and m/z of product ion 2 was used as a qualifier. EM peak quantifications were performed by the d-EM internal standard method and constructing matrix matched (charcoal stripped human urine) 8-point calibration curve for each of the six analytical batches. Based on the retention time, chemical and structural similarities following d-EM were used as internal standards for their respected EM: d<sub>4</sub>E<sub>2</sub> for E<sub>2</sub> and E<sub>1</sub>; d<sub>3</sub>E<sub>3</sub> for E<sub>3</sub>, 16-ketoE<sub>2</sub> and 16 $\alpha$ -OHE<sub>1</sub>; d<sub>2</sub>-16-epiE<sub>3</sub> for 17-epiE<sub>3</sub>; d<sub>5</sub>-2-MeOE<sub>2</sub> for 2-MeOE<sub>2</sub>, 4-MeOE<sub>2</sub>, 2-MeOE<sub>1</sub>, 4-MeOE<sub>1</sub> and 3-MeOE<sub>1</sub>; d<sub>5</sub>-2-OHE<sub>2</sub> for 2-OHE<sub>2</sub>, 2-OHE<sub>1</sub> and 4-OHE<sub>1</sub>. The calibration graphs were evaluated by plotting the peak area ratios of dansyl-EM/d-dansyl-EM versus concentration

(ng/mL) of EM in the standard and using linear regression with 1/X weighting to fit the data. Using this linear function, the amount of EM in urine samples were then interpolated.

### **2.2.7 Intra/Inter assay accuracy and precision of the measured oestrogen metabolites**

The intra-batch accuracy and precision of the method were assessed using matrix-matched QC samples prepared as three replicates at the concentration of 2 ng/mL of each EM, while the inter-batch precision was evaluated with six consecutive analytical batches (N=18 QC samples). The accuracy was calculated as relative error % (RE%) = (measured concentration of EM - known concentration of EM / known concentration of EM) \*100. The intra- and inter-batch precision were calculated as a percentage coefficient of variance (CV%).

### **2.2.8 Study population**

Samples for this analysis were drawn from a study of premenopausal hormone levels (E<sub>1</sub>G= oestrone glucuronide and PG=pregnanediol glucuronide) and breast cancer risk [6] in which a total of 729 healthy premenopausal women of European ancestry were genotyped for 642 single nucleotide polymorphisms (SNPs) that tagged 42 genes which might influence endogenous sex steroid levels. Of the 729 healthy premenopausal women who were genotyped, 358 were participants in the Mammography Oestrogens and Growth factors study (MOG), an observational study nested within the intervention arm of a trial of annual mammographic screening [6, full details of this study have been previously published in 59], The other 371 women were first-degree relatives or friends of breast cancer patients participating in the British Breast Cancer study (BBC) [6, full details of this study were previously published in 60]. Eligibility criteria were: regular menstrual cycles (usual cycle length between 21 and 35 days), not taking any oral contraceptives or hormone replacement therapy, not diagnosed with breast cancer at the time of recruitment and having provided serial urine samples at prescribed days during their menstrual cycles. To account for cyclic variation of endogenous hormone (oestrogens and progesterone) levels the date of ovulation for females was estimated based on the date of the first day of their menstrual period and their usual menstrual cycle length. The ovulation date was predicted to be 14 days before the date of the start of their next cycle. Based on this estimate a series of follicular phase early morning urine samples taken around the predicted date of ovulation were collected. Analysis of premenopausal hormone levels (E<sub>1</sub>G and PG), selection of SNPs for genotyping, DNA extraction and genotyping were carried out at The Institute of Cancer Research, London. To account for urine concentration, creatinine levels were measured colorimetrically, these measurements were performed at the MRC Centre for Reproductive Health, The University of Edinburgh.

Out of the 642 SNPs that were genotyped, one SNP (rs10273424) mapping to the cytochrome P4503A gene cluster (CYP3A5, CYP3A7 and CYP3A4 genes) at chromosome 7q22.1 was associated with a 22% reduction in urinary E<sub>1</sub>G levels (95% CI -28% to -15%; P=10<sup>-9</sup>). The association between this SNP (rs10273424) and risk of breast cancer was also examined in relation to breast cancer risk directly, by comparing the frequency of this variant in subsequent work using data from 10,551 breast cancer patients and 17,535 control subjects [6, 61]. The results showed that the minor allele of (rs10235235), a proxy for (rs10273424), was also associated with a moderate reduction in the risk of breast cancer in the case of patients who were diagnosed at, or before, the age of 50. This study implied that the SNP (rs10273424), which maps to the CYP3A locus, may influence breast cancer risk via an effect of CYP3A on the metabolism of oestrogens in premenopausal women. A subsequent study [61], aiming to refine the association between SNP (rs10273424) and premenopausal hormone levels, identified a putative causal variant (rs45446698) to be associated with 54% reduction in E<sub>1</sub>G levels. This putative variant tagging *CYP3A7\*1C* allele was also shown to be associated with adverse outcomes in breast cancer, lung cancer and chronic lymphatic leukaemia.

For this PhD, two of the early morning follicular phase urine samples (taken on the first and second day after ovulation) were analysed, per woman, for each of 60 females of the original 729 healthy premenopausal females of European ancestry. These women were selected on their genotype for the putative causal CYP3A SNP (rs45446698) such that 30 women were carriers of the rare (rs45446698-C allele) and 30 were homozygous for the common (rs45446698-A) allele. 43 women (23 AC-heterozygotes and 20 AA-homozygotes) were from the MOG study and 17 (7 AC- heterozygotes and 10 AA-homozygotes) were from the BBC study. Figure 2.4, shows a summary of the study design and selection of a number of samples analysed in this PhD.

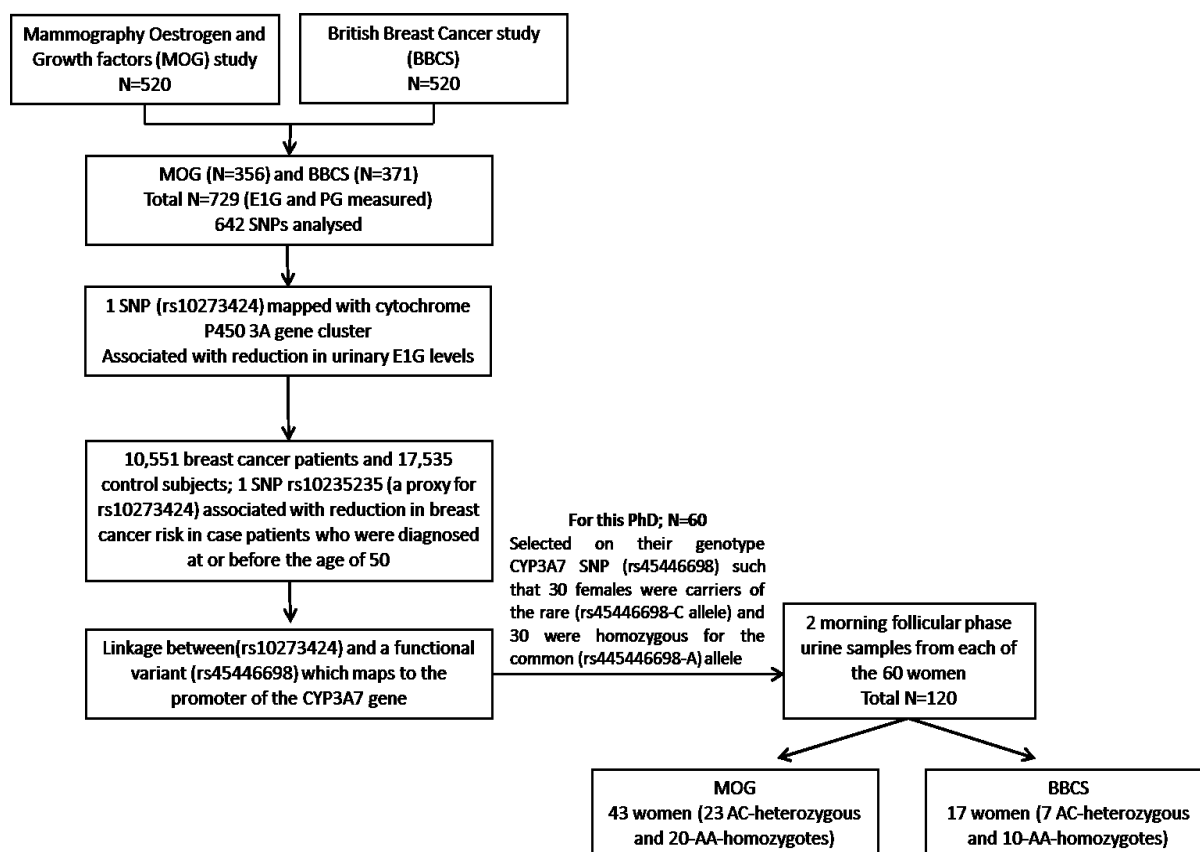


**Figure 2.4 Study design showing selection and number of samples analysed in this PhD.** N= 729 urine samples from healthy premenopausal women in which oestrone glucuronide (E<sub>1</sub>G) and pregnanediol glucuronide (PG) were measured and 642 SNPs were analysed. Out of the original 729 women from MOG and BBCS study, we selected two early morning follicular phase urine samples were selected per woman, for each of 60 women (total N=120) these samples were selected based on their genotype for the *CYP3A7* SNP (rs454466988) such that 30 women were carriers of the rare (rs45446698-C allele) and 30 were homozygous for the common (rs45446698-A) allele.

## 2.2.9 Statistical analysis

All statistical analyses were performed using R (version 3.2.1). Before any statistical analysis, absolute measured concentrations in nanograms per millilitre (ng/mL) of all EM in follicular phase premenopausal urine samples were adjusted for creatinine to convert the data into picomoles per milligram (pmol/mg) of creatinine. The concentrations per woman from each of two follicular phase samples were averaged resulting in a data set of N=60 creatinine adjusted measurements. This data set was then natural log<sub>e</sub> transformed due to the non-normal distribution of the creatinine-adjusted averaged concentrations of the 14 EMs. In addition, from a statistical standpoint, two measurements per woman cannot be used in a linear regression without adjusting for the standard error. To assess the intra-individual variation of the 14 EMs over the two follicular days, intra-class correlations (ICCs) were calculated by dividing the between-women variance by the sum of the within-woman and between women variances and their 95% confidence intervals were also calculated [62].

For descriptive statistics and to express the urinary levels of EMs, geometric means along with 5<sup>th</sup> and 95<sup>th</sup> percentile ranges and percentage measure of total EM were calculated. Some of the observed values were exceptionally large and data was positively skewed to correct for this disproportion, geometric means were calculated to report the central tendency of the measured EM [63]. The advantage of reporting geometric mean over arithmetic mean in the descriptive statistics was that geometric mean is less affected by the distortion of the positively skewed data compared to the arithmetic mean hence provided better statistical inference. This calculation was based on the data from only 30 women with the common AA genotype. The relative (creatinine-adjusted and averaged for follicular days) individual EM concentrations were summed into groups according to their chemical characteristics and also as described in previously published reports from the Nurses' Health study II [64], for e.g. according to catechol (sum of 2-OHE<sub>1</sub>, 2-OHE<sub>2</sub> and 4-OHE<sub>1</sub>), methylated catechol (sum of 2-MeOE<sub>1</sub>, 2-MeOE<sub>2</sub>, 3-MeOE<sub>1</sub>, 4-MeOE<sub>1</sub> and 4-MeOE<sub>2</sub>), 2-catechol (sum of 2-OHE<sub>1</sub> and 2-OHE<sub>2</sub>), 4-catechol (4-OHE<sub>1</sub>, 4-MeOE<sub>1</sub> and 4-MeOE<sub>2</sub>), methylated 2-catechol (sum of 2-MeOE<sub>1</sub>, 2-MeOE<sub>2</sub>, 3-MeOE<sub>1</sub>) and methylated 4-catechol (sum of 4-MeOE<sub>1</sub> and 4-MeOE<sub>2</sub>); according to their pathways e.g. 2-hydroxylation (sum of 2-OHE<sub>1</sub>, 2-OHE<sub>2</sub>, 2-MeOE<sub>1</sub>, 2-MeOE<sub>2</sub> and 3-MeOE<sub>1</sub>); 4-hydroxylation pathway (sum of 4-OHE<sub>1</sub>, 4-MeOE<sub>1</sub> and 4-MeOE<sub>2</sub>); and 16-hydroxylation pathway (sum of E<sub>3</sub>, 16- $\alpha$ -OHE<sub>1</sub>, 17-epiE<sub>3</sub> and 16-KetoE<sub>2</sub>). The sum of oestrone (E<sub>1</sub>) and oestradiol (E<sub>2</sub>) was calculated as Parent EM. Total EM was calculated as the sum of all 14 EMs. Some of the selected ratios between chemical groups and pathways were also calculated for descriptive purposes. Percent of total EM were calculated by dividing individual or grouped EM by the total EM and multiplying by 100. Means and ranges for percent of total EM (percent measures) were arithmetic means and calculated for relative EMs (pmol/mg creatinine, averaged and log<sub>e</sub> transformed) levels.

The associations and the percentage difference in the urinary levels of 14 EMs in 30 premenopausal women who were carriers of the *CYP3A7\*1C* allele (i.e. women with the AC genotype of SNP rs454466988) compared to 30 premenopausal women who were non-carriers (i.e. women with the AA genotype) were calculated using linear regression models of arithmetic relative and log<sub>e</sub>-transformed EM levels with, and without, adjusting for covariates: age, BMI and parity. In linear regression models, AC was coded as 1 and AA as 0, and t-tests of the regression coefficient were used to calculate P-values, while R-squared values were used to estimate the proportion of variance explained by the genotype. In these models' levels of oestrogen metabolites in carriers of the *CYP3A7\*1C* allele (AC) compared to non-carriers (AA) were expressed as a percentage of the levels in non-carriers (AA) such that 100% is equivalent to no difference (the Null hypothesis that women with AA genotype=100% and represents no change). EM in the AC genotype was deemed to be lower compared to AA when they were less than 100%

and deemed to higher than AC genotype when they were more than 100%. Bonferroni-adjusted *P* values were applied to take account of multiple testing. The statistical power of each hypothesis was calculated independently of each EM and the modified Bonferroni *P* value was also reported. Spearman correlation coefficients were also evaluated separately within the datasets of AA and AC genotype to explore and compare the parent EMs with non-parent EMs and 2-, 4- and 16-hydroxylation pathways.

## 2.3 Results

### 2.3.1 HPLC-MS/MS chromatographic profiles of Oestrogen metabolites in quality control and follicular phase premenopausal urine sample

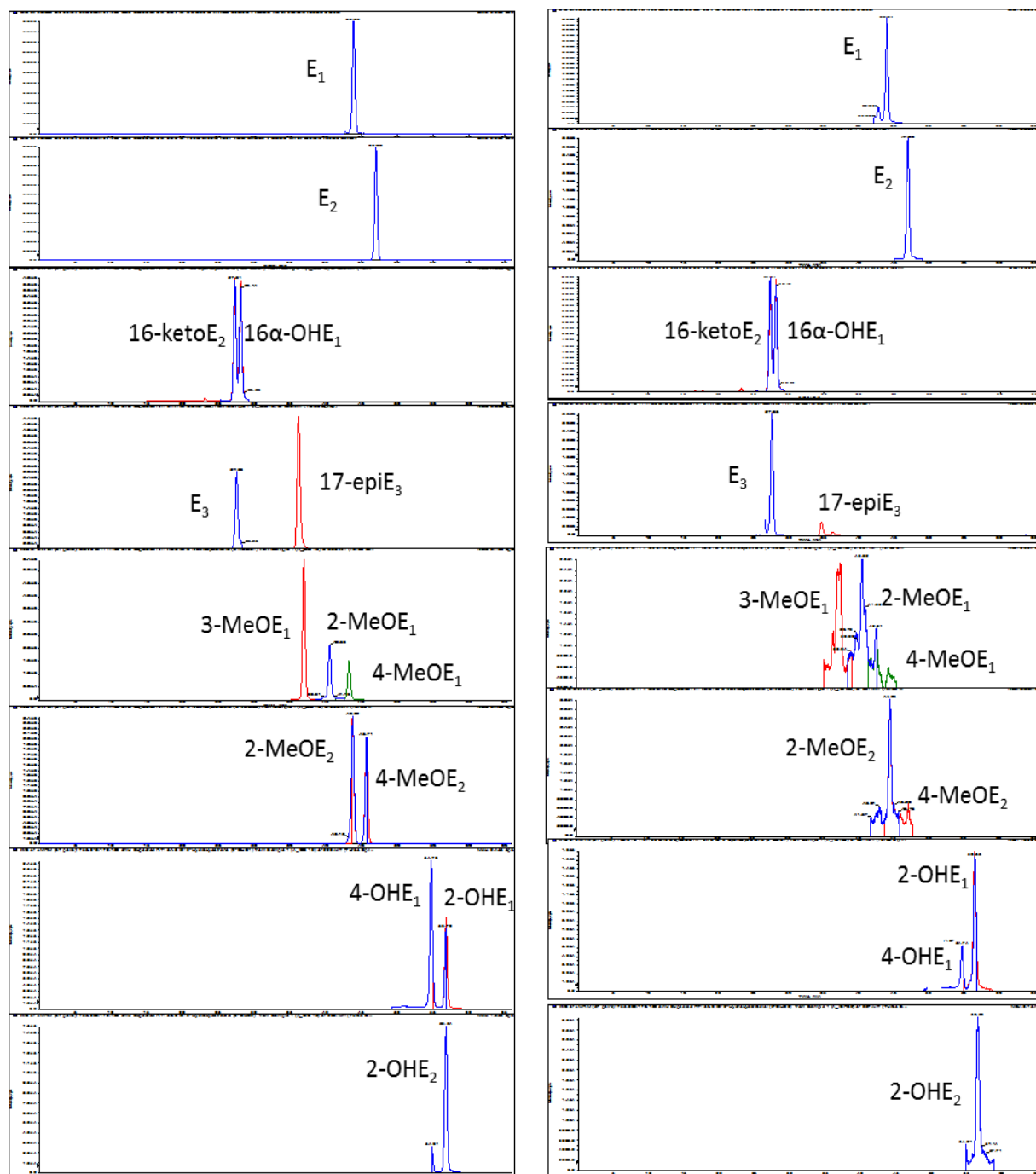


**Figure 2.5 HPLC-MS/MS MRM chromatographic profile of dansylated EMs. (A) QC sample at 2ng/mL and (B) follicular phase premenopausal urine sample. The peaks corresponding to each 14 EMs listed in materials and methods are indicated within the chromatographic profiles.**

Since urinary EM concentrations can range as low as picograms per millilitre (pg/mL) and fluctuate depending upon age, sex, and menopausal status, it was vital to implement a method

which was sensitive, accurate and precise to quantify adequately individual endogenous EM levels. Figure 2.5 shows clear separation within HPLC-MS/MS MRM chromatographic profiles of all 14 dansylated oestrogen metabolites in a QC sample at 2ng/mL concentration and a follicular phase premenopausal urine sample. The observed concentrations of 4-MeOE1 and 4-MeOE2 were very low (below LOD) in premenopausal urine samples.

### **2.3.2 Calibration curve and limit of quantitation (LOQ)**

The 8-point calibration curves prepared (in oestrogen free human urine; charcoal stripped human urine) for the quantitation of each urinary EM were linear and spanned over a broad, approximately  $10^3$  fold concentration range (0.04–10 ng/0.5 mL sample or 0.08-20 ng/mL). The regression correlation coefficients were greater than 0.98 of all in all of six batches.

The limit of detection (LOD) and limit of quantification (LOQ) were determined by calculating standard deviation from a signal to noise ratio (S/N) in 10 replicate blank samples (charcoal stripped urine). Three times the standard deviation of S/N in 10 blank samples were assigned as LOD and ten times the standard deviation of S/N was deemed as LOQ for each EM. Both the LODs and LOQs were determined for the entire study. Although according to FDA guidelines LOD and LOQ should be calculated per analytical batch using calibration curve as explained in Appendix 2 and 3 of ref [84]. In this study, the LOD and LOQ were calculated once for the entire study because of limited resources (time availability of the instrument and finances required to run additional samples, exacerbated by the long LC runtime (77 min/sample)). The LOQ for this method was the lowest concentration of the calibration curve which was set as 0.08 ng/mL. More importantly in this analysis, the LOQ was determined at the concentration (0.08 ng/mL) where the intra- and inter-batch precision of all the EMs were consistently less than 10% and the intra- and inter-batch accuracies were between 97 and 105%.

Although most of the of the EM concentrations measured in this study ranged between 0.08 and 20 ng/mL, measured concentrations of oestriol ( $E_3$ ) were considerably higher, spanning between 20 and 40 ng/mL in 33% (N= 40) and 40 to 76 ng/mL in 8% (N=10) of the premenopausal urine samples. The measured absolute concentration of urinary EMs: 4-MeOE<sub>1</sub>, 4-MeOE<sub>2</sub>, 3MeOE<sub>1</sub> and 2-MeOE<sub>2</sub> were below their LOD values in 78% (N=94), 89% (N=107), 8% (N=10) and 8% (N=10) of urine samples respectively. Accordingly, both 4-MeOE<sub>1</sub>, 4-MeOE<sub>2</sub> were excluded from further analysis. For 3MeOE<sub>1</sub> and 2-MeOE<sub>2</sub> women for whom both samples gave readings below the LOD were excluded from any subsequent analysis; women for whom only one sample was below the LOD were included, and the single measurement was used. Overall, the results demonstrated that this HPLC-MS/MS method had adequate LOQ for measuring endogenous oestrogens in

premenopausal female urine. An approximate concentration of 0.008 ng/sample or 8pg/sample LOD can be readily achieved using this method.

### **2.3.3 Method validation**

This HPLC-MS/MS method was validated to demonstrate its performance regarding selectivity, accuracy and precision.

#### **Selectivity**

To ensure this method could determine EMs in urine samples without any interference from other compounds at the same retention times, the selectivity was examined by comparing chromatograms of a hydrolysed, extracted and derivatised blank sample (charcoal stripped human urine) with chromatograms of a hydrolysed, extracted and derivatised blank sample spiked with EMs and d-EM. The analysis of all 14 EMs using MRM mode was highly selective with no interfering compounds.

#### **Intra/Inter batch Accuracy and Precision of the measured oestrogen metabolites**

A quality control (QC) sample approach was applied to the entire analysis, and premenopausal urine samples bracketed with QC samples (1 QC for every 10 clinical samples) were considered valid only if the QC samples were acceptable on accuracy and precision criteria. The intra-batch accuracy and intra-batch precision were determined within each batch by analysing 3 QC replicates at a concentration of 2 ng/mL. As shown in Table 2.5, intra-batch accuracy (recovery of a known, added amount of EM in percentage, expressed as RE%) for the 2 ng/mL samples ranged from 94% to 108% and intra-day precision ranged between 6% to 10%. The inter-batch precision of the EMs was evaluated from six independent consecutive analytical batches (N=18 QC samples). Figure 2.6 shows that the CV (%) of measured EMs ranged from 6% to 14%. Both intra- and inter-batch precisions were less than 20%, which is deemed acceptable according to the FDA & EMA guidelines for Bioanalytical method validation [65, 66]. These results indicate the adequate reliability and reproducibility of this method.

|                       |                               | 2.0 ng/mL QC urine |                  |
|-----------------------|-------------------------------|--------------------|------------------|
| Oestrogen metabolites |                               | Accuracy (RE %)    | Precision (CV %) |
| Parent EM             | E <sub>1</sub>                | 97                 | 9                |
|                       | E <sub>2</sub>                | 99                 | 9                |
| 2-Pathway             | 2-OHE <sub>1</sub>            | 99                 | 7                |
|                       | 2-OHE <sub>2</sub>            | 103                | 9                |
|                       | 2-MeOE <sub>1</sub>           | 94                 | 6                |
|                       | 2-MeOE <sub>2</sub>           | 102                | 7                |
|                       | 3-MeOE <sub>1</sub>           | 93                 | 9                |
| 4-Pathway             | 4-OHE <sub>1</sub>            | 104                | 10               |
|                       | 4-MeOE <sub>1</sub>           | 108                | 10               |
|                       | 4-MeOE <sub>2</sub>           | 103                | 9                |
| 16-Pathway            | 16- $\alpha$ OHE <sub>1</sub> | 104                | 6                |
|                       | 17-epiE <sub>3</sub>          | 105                | 7                |
|                       | E <sub>3</sub>                | 98                 | 10               |
|                       | 16-ketoE <sub>2</sub>         | 102                | 6                |

**Table 2.5 Inter-batch accuracy (recovery of the known amount of EM in percentage) and precision of EM measurement after hydrolysis, extraction, derivatisation and HPLC-MS/MS steps. The values shown in this table are from one batch and are representative of other five batches.**

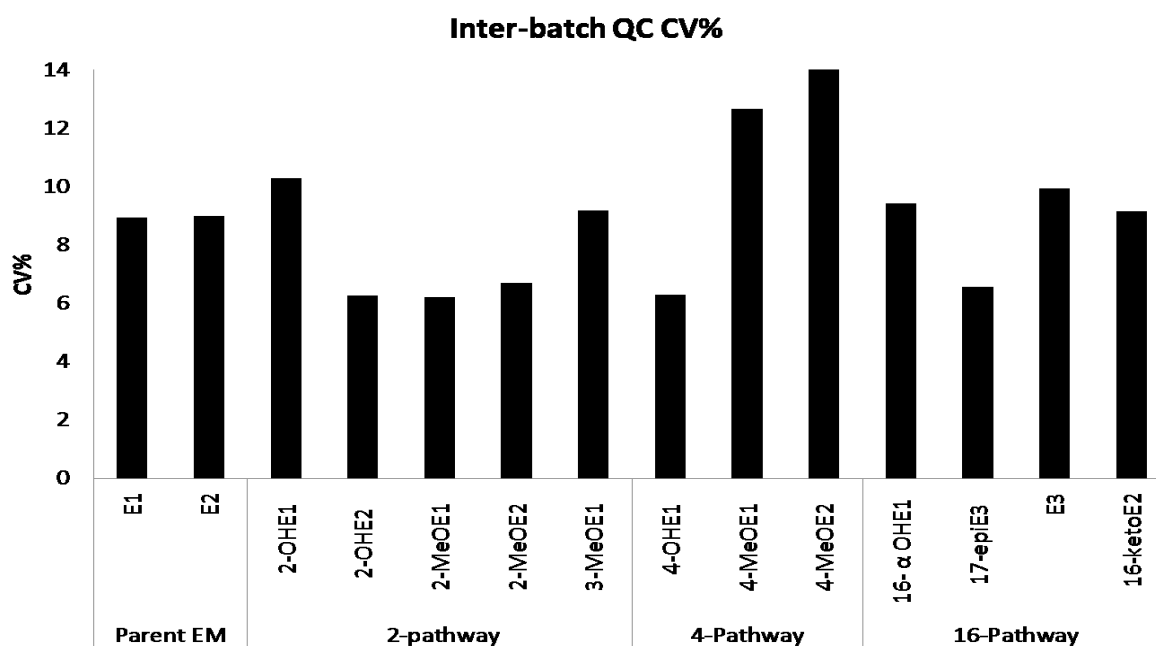


**Figure 2.6 Inter-batch precision calculated as a percentage coefficient of variance (CV %) in QC samples (N=18; concentration 2 ng/mL) of EM measurement after hydrolysis, extraction, derivatisation and HPLC-MS/MS steps.**

### 2.3.4 Intra-class correlations (ICCs) of measured oestrogen metabolites

To assess intra- and inter-individual variation (within the woman and between women) of EMs over two follicular days, Intra-class correlations (ICC) and 95% confidence intervals (95% CI) were calculated using creatinine-adjusted and log<sub>e</sub>-transformed data from the 30 women with the common AA genotype (Table 2.6). Data from these women were used as the AA genotype is common in the population with only 8% of women having the rare (AC) genotype. Thus, restricting these descriptive statistics to this subset of samples provides a reasonable approximation to a random sample from the population. Overall, the ICCs were fairly high for all individual EMs (ICCs 0.62 to 0.83), and the ICC for total EMs was also good (0.65). ICCs for parent EMs were similar to one another (ICC E<sub>1</sub>=0.65, ICC E<sub>2</sub>=0.65).

| Individual and grouped EMs                                    | ICCs (95% CI)    |
|---------------------------------------------------------------|------------------|
| Total EM                                                      | 0.65 (0.41-0.87) |
| Parent EMs                                                    | 0.64 (0.38-0.85) |
| Oestrone (E <sub>1</sub> )                                    | 0.65 (0.39-0.86) |
| Oestradiol (E <sub>2</sub> )                                  | 0.65 (0.38-0.85) |
| Catechol EMs                                                  | 0.60 (0.38-0.81) |
| 2 Catechol EMs                                                | 0.70 (0.49-0.91) |
| 2-Hydroxoestrone (2-OHE <sub>1</sub> )                        | 0.71 (0.51-0.92) |
| 2-Hydroxyoestradiol (2-OHE <sub>2</sub> )                     | 0.64 (0.40-0.79) |
| 4 Catechol EM                                                 |                  |
| 4-Hydroxyoestrone (4-OHE <sub>1</sub> )                       | 0.68 (0.45-0.81) |
| Methylated 2-catechol EMs                                     | 0.85 (0.74-0.97) |
| 2-Methoxyoestrone (2-MeOE <sub>1</sub> )                      | 0.86 (0.75-0.97) |
| 2-Methoxyoestradiol (2-MeOE <sub>2</sub> )                    | 0.62 (0.32-0.83) |
| 2-Hydroxyoestrone-3-methyl ether (3-MeOE <sub>1</sub> )       | 0.68 (0.40-0.86) |
| 2-Hydroxylation pathway EMs                                   | 0.73 (0.54-0.92) |
| 16-Hydroxylation pathway EMs                                  | 0.76 (0.59-0.94) |
| 16 $\alpha$ -Hydroxyoestrone (16 $\alpha$ -OHE <sub>1</sub> ) | 0.70 (0.48-0.91) |
| 17-Epiestriol (17-epiE <sub>3</sub> )                         | 0.83 (0.70-0.96) |
| Oestriol (E <sub>3</sub> )                                    | 0.76 (0.58-0.93) |
| 16-Ketooestradiol (16-ketoE <sub>2</sub> )                    | 0.80 (0.66-0.95) |

**Table 2.6 ICCs (95% confidence intervals) for individual and grouped EMs, expressed as relative to creatinine and log-transformed levels in N=30 women with common (rs45446698-A) allele or AA genotype.**

ICCs for grouped EMs were all  $\geq 0.60$ , with highest ICCs for methylated 2-catechol EMs (0.85) followed by 2 Catechol EMs (ICC 0.70). ICCs for methylated-4-cathechols (4-MeOE<sub>1</sub> and 4-MeOE<sub>2</sub>) were not calculated as they had lowest concentrations and measurements were below LOD in 78% of the samples. For the 2- and 16-hydroxylation pathways, ICCs were high (0.73 and 0.76

respectively), whereas the ICC for the 4-hydroxylation pathway was again not calculated due to low concentrations of metabolites. High ICCs of individual and grouped EMs demonstrated that the measurements of EMs over two follicular days were consistent and reproducible. Therefore, concentrations over two follicular days per woman were averaged to obtain one measurement per woman in both AC and AA genotype population.

### 2.3.5 Urinary levels of oestrogen metabolites

The descriptive statistics, from only the 30 women with the common AA genotype (frequency of AA genotype represents 92% of the population) are presented in Table 2.7. The geometric mean and 5<sup>th</sup>-95<sup>th</sup> percentile ranges of the individual and grouped EMs and selected ratios, along with the percentage of total EMs expressed as arithmetic means with 5<sup>th</sup>-95<sup>th</sup> percentile ranges. Parent EMs ( $E_1$  and  $E_2$ ) constituted around 22% (5<sup>th</sup>-95<sup>th</sup> percentile = 18-26%) of the total urinary EM, whereas the 2-hydroxylation pathway, 4-hydroxylation pathway and 16-hydroxylation pathway were 22% (18-26%), 3% (2- 4%) and 45% (39-51%) respectively.

The geometric mean total EM (all 14 EMs) was 169 pmol/mg creatinine, with the top contributors being  $E_3$  (38% of the total EMs),  $E_1$  (17% of the total EMs) and 2-OHE<sub>1</sub> (12.6% of the total EMs), together accounting for around 67% of the total urinary EM. The EMs with the lowest concentrations were the four methylated catechols: 2- MeOE<sub>2</sub> (0.39% of the total EMs), 3-MeOE<sub>1</sub> (0.43% of the total EM), 4-MeOE<sub>1</sub> and 4-MeOE<sub>2</sub> (were excluded as they were below LOD in the majority of the women); together contributing < 1% of the total EMs. The three catechol oestrogens (2-OHE<sub>1</sub>, 2-OHE<sub>2</sub> and 4-OHE<sub>1</sub>) comprised of ~ 19% of the total EMs, while the five-methylated catechol (2-MeOE<sub>1</sub>, 2-MeOE<sub>2</sub>, 3-MeOE<sub>1</sub>, 4-MeOE<sub>1</sub> and 4-MeOE<sub>2</sub>) oestrogens accounted ~ 6% of the total EMs.

To contrast EMs in different metabolic pathways, selected ratios were calculated. The ratio of catechol/methylated-catechols was 3.0, indicating that the catechol EMs concentrations were higher than methylated-catechols EMs concentrations. The overall concentrations of EMs in the 16- hydroxylation pathway was higher than overall EMs in the 2-hydroxylation pathway (2-Pathway/16-pathway, ratio=0.51), and the overall concentrations of EMs in the 2- hydroxylation pathway were higher than overall EMs in the 4-hydroxylation pathway (4-Pathway/2-pathway, ratio=0.15). The parent oestrogens were less abundant than non-parent oestrogen metabolites (parent-oestrogens/oestrogen-metabolites, ratio 0.22).

| Individual and grouped EM                                    | Geometric mean<br>(5th-95th percentile) | Arithmetic mean<br>%    |
|--------------------------------------------------------------|-----------------------------------------|-------------------------|
|                                                              | pmol/mg creatinine<br>(N=30)            | % of total EM<br>(N=30) |
| Total EMs                                                    | 169 (143-201)                           |                         |
| Parent EMs                                                   | 36.6 (29.5-45.5)                        | 21.6 (18.30-26.3)       |
| Oestrone (E <sub>1</sub> )                                   | 28.6 (23.0-35.5)                        | 16.9 (14.2-20)          |
| Oestradiol (E <sub>2</sub> )                                 | 7.8 (6.2-9.9)                           | 4.6 (3.8-5.5)           |
| Catechol EMs                                                 | 33.4 (25.6-43.5)                        | 19.3 (15.8-23.6)        |
| 2 Catechol EMs                                               | 26.3 (20.6-33.7)                        | 15.3 (12.7-18.5)        |
| 2-Hydroxoestrone (2-OHE <sub>1</sub> )                       | 21.7 (17-27)                            | 12.6 (10.5-15.2)        |
| 2-Hydroxyoestradiol (2-OHE <sub>2</sub> )                    | 4.6 (3.3-5.7)                           | 2.5 (1.9-3.2)           |
| 4 Catechol EMs                                               |                                         |                         |
| 4-Hydroxoestrone (4-OHE <sub>1</sub> )                       | 5.5 (3.7-8.3)                           | 3.2 (2.2-4.5)           |
| Methylated 2-catechol EM                                     | 10.1 (8.0-12.7)                         | 5.9 (4.9-7.0)           |
| 2-Methoxyoestrone (2-MeOE <sub>1</sub> )                     | 9.4 (7.5-11.8)                          | 5.5 (4.5-6.6)           |
| 2-Methoxyoestradiol (2-MeOE <sub>2</sub> )                   | 0.67 (0.51-.087)                        | 0.39 (0.32-0.48)        |
| 2-Hydroxyoestrone-3-methyl ether (3-MeOE <sub>1</sub> )      | 0.74 (0.55-1)                           | 0.43 (0.33-0.55)        |
| Methylated 4-catechol EMs                                    | N/A                                     | N/A                     |
| 4-Methoxyoestrone (4-MeOE <sub>1</sub> )                     | N/A                                     | N/A                     |
| 4-Methoxyoestradiol (4-MeOE <sub>2</sub> )                   | N/A                                     | N/A                     |
| 2-Hydroxylation pathway EMs                                  | 38.0 (30.1-47.9)                        | 22.1 (18.6-26.3)        |
| 4-Hydroxylation pathway EMs                                  | 5.4 (3.6-8.0)                           | 3.1 (2.2-4.4)           |
| 16-Hydroxylation pathway EMs                                 | 75.4 (60.9-93.4)                        | 45.0 (39.7-51)          |
| 16 $\alpha$ -Hydroxoestrone (16 $\alpha$ -OHE <sub>1</sub> ) | 6.6 (4.6-9.4)                           | 3.9 (3.0-5.0)           |
| 17-Epiestriol (17-epiE <sub>3</sub> )                        | 0.31 (0.20-0.46)                        | 0.18 (0.12-0.26)        |
| Oestriol (E <sub>3</sub> )                                   | 63.5 (51.5-78.4)                        | 38.0 (33.1-43.6)        |
| 16-Ketooestradiol (16-ketoE <sub>2</sub> )                   | 2.6 (1.9-3.6)                           | 1.6 (1.2-2.1)           |
| <b>Ratios</b>                                                |                                         |                         |
| 4-Catechol/2-catechols                                       | 0.22 (0.16-0.29)                        |                         |
| 2-Catechols/16-pathway                                       | 0.35 (0.26-0.47)                        |                         |
| Catechols/16-pathway                                         | 0.44 (0.32-0.61)                        |                         |
| 4-Pathway/2-pathway                                          | 0.15 (0.11-0.20)                        |                         |
| 2-Pathway/16-pathway                                         | 0.51 (0.38-0.67)                        |                         |
| 4-Pathway/16-pathway                                         | 0.07 (0.05-0.11)                        |                         |
| 2-Catechols/methylated-2-catechols                           | 2.63 (2.2-3.1)                          |                         |
| 4-Catechol/methylated-4-catechols                            | 7.3 (5.4-10.0)                          |                         |
| Catechols/methylated-catechols                               | 3.0 (2.5-3.6)                           |                         |
| Parent-oestrogens/oestrogen-metabolites                      | 0.22 (0.18-0.26)                        |                         |

**Table 2.7** Geometric mean and 5<sup>th</sup>-95<sup>th</sup> percentile ranges of individual and grouped urinary EM expressed as relative (creatinine-adjusted and follicular days averaged) levels in N=30 women with common (rs45446698-A) allele or AA genotype. Selected EM ratios are also reported. Percentage of the total EM expressed as arithmetic mean (creatinine adjusted, follicular days averaged and log<sub>e</sub> transformed) and their 5<sup>th</sup>-95<sup>th</sup> percentile ranges.

### 2.3.6 Associations between urinary oestrogen metabolite levels and carriers of the CYP3A7\*1C allele (AC genotype) compared to non-carriers (AA genotype)

The percent difference in EM levels in women who were carriers of the rare *CYP3A7\*1C* allele (AC genotype) compared to non-carriers (AA genotype) of SNP (rs45446698) from unadjusted linear regression models and that were adjusted for age, BMI and parity are shown in Table 2.8. R-squared values represent the proportion of variance explained by “genotype” in unadjusted models and by “genotype, BMI, age and parity” in adjusted models. After adjusting for age, BMI and parity, the estimate of the effect size (P-values) remain similar for most of the EM and results from the adjusted models are described below.

Based on the Bonferroni *P* values, there was a strong evidence of a difference in the levels of individual and grouped EMs in carriers of the C-allele of rs45446698 (AC genotype) compared to non-carriers (AA genotype). The biggest percentage difference was observed in the levels of 2-OHE<sub>1</sub> ( $P=6.6 \times 10^{-13}$ ) followed by 2-MeOE<sub>1</sub> ( $P=1.9 \times 10^{-12}$ ), 2-OHE<sub>2</sub> ( $P=1.8 \times 10^{-7}$ ) and E<sub>1</sub> ( $P=0.001$ ). Women with the AC genotype had 20.7% (95% CI 14.7%-29%) 2-OHE<sub>1</sub> levels compared to women with the AA genotype and “genotype” accounted for 59% of the variability in 2-OHE<sub>1</sub> levels and “genotype, BMI, age and parity” accounted for 64% of the observed variability. This meant that the three additional covariates ‘age, BMI and parity’ only explained the 5% of the variance and ‘genotype’ was far more important in accounting for the variances. Similarly, AC genotype women had levels of 18.8% (95% CI 12.4%-26.4%) 2-MeOE<sub>1</sub> and 30.7% (95% CI 20.7%-45.6%) 2-OHE<sub>2</sub> compared to women with the AA genotype, and the genotype accounted for 58% and 37% variability in their levels respectively. Among the grouped EMs the largest difference comparing AC genotype women to those with the AA genotype was found in 2-catechol EMs 22.7% (95% CI 19.3%-40.9%,  $P=3.6 \times 10^{-12}$ ). Overall, women with the AC-genotype also had lower levels of 2-hydroxylation pathway EMs (24.1%, 95% CI 17.7%-32.8%,  $P=6.6 \times 10^{-13}$ ) and methylated-2-catechol EMs (20.7%, 95% CI 14.2%-30.3%,  $P=2.7 \times 10^{-11}$ ). Compared to AA-homozygotes, the levels of both parent EMs (E<sub>1</sub> and E<sub>2</sub>) were 54.6% and 74.3% in carriers of the C-allele of rs45446698, although the association with E<sub>2</sub> was non-significant. For the 16-hydroxylation pathway, however, the opposite was observed; levels of 16 $\alpha$ -OHE<sub>1</sub> were 200.9% (95% CI 123.6%-328.6%,  $P=0.005$ ) and levels of 17-epiE<sub>3</sub> were 260% (95% CI 164%-528%,  $P=0.0005$ ) in women with the AC-genotype of the SNP rs45446698, compared to women with the AA genotype. Overall levels of 16-hydroxylation pathway EMs were significantly higher in women who were carriers of the rare AC genotype.

| Individual and grouped EMs  | Percentage (%) change in EMs in carriers* (95% CI), |                        |           | Percentage (%) change in EMs** in carriers* (95% CI), |                        |           | Bonferroni adjusted P-value |
|-----------------------------|-----------------------------------------------------|------------------------|-----------|-------------------------------------------------------|------------------------|-----------|-----------------------------|
|                             | AC genotype                                         | P value                | R squared | AC genotype.                                          | P value                | R-squared |                             |
| Total EM                    | 84.8 (67.3-107)                                     | 0.16                   | 0.03      | 84.6 (66.1-108.2)                                     | 0.18                   | 0.04      | 1                           |
| Parent EM                   | 59.05 (42.7-81.5)                                   | 0.001                  | 0.16      | 59.1 (41.9-83.5)                                      | 0.003                  | 0.15      | 0.13                        |
| E <sub>1</sub>              | 54.7 (39.4-75.9)                                    | 0.0005                 | 0.19      | 54.6 (38.5-77.4)                                      | 0.001                  | 0.19      | 0.04                        |
| E <sub>2</sub>              | 73.2 (52.5-102.2)                                   | 0.06                   | 0.06      | 74.3 (52.1-106)                                       | 0.10                   | 0.06      | 1                           |
| Catechol EM                 | 28.9 (19.9-42.1)                                    | 1.2 X10 <sup>-08</sup> | 0.43      | 28.1 (19.3-40.9)                                      | 9.0 X10 <sup>-09</sup> | 0.50      | 3.3 X10 <sup>-07</sup>      |
| 2 Catechol EM               | 23.87 (17-33.4)                                     | 8.3 X10 <sup>-12</sup> | 0.56      | 22.7 (16.2-31.7)                                      | 3.6 X10 <sup>-12</sup> | 0.62      | 1.3 X10 <sup>-10</sup>      |
| 2-OHE <sub>1</sub>          | 21.7 (15.4-30.4)                                    | 1.0 X10 <sup>-12</sup> | 0.59      | 20.7 (14.7-29)                                        | 6.6 X10 <sup>-13</sup> | 0.64      | 2.4 X10 <sup>-11</sup>      |
| 2-OHE <sub>2</sub>          | 32.0 (21.6-47.4)                                    | 2.7X10 <sup>-07</sup>  | 0.37      | 30.7 (20.7-45.6)                                      | 1.8 X10 <sup>-07</sup> | 0.44      | 6.6 X10 <sup>-05</sup>      |
| 4 Catechol EM               |                                                     |                        |           |                                                       |                        |           |                             |
| 4-OHE <sub>1</sub>          | 48.1 (27-85.7)                                      | 0.01                   | 0.1       | 49.9 (26.2-83.8)                                      | 0.01                   | 0.20      | 0.43                        |
| Methylated 2 catechol EM    | 21.3 (14.8-30.7)                                    | 1.0 X10 <sup>-11</sup> | 0.55      | 20.7 (14.2-30.3)                                      | 2.7 X10 <sup>-11</sup> | 0.58      | 1.0 X10 <sup>-09</sup>      |
| 2-MeOE <sub>1</sub>         | 18.8 (12.9-27.4)                                    | 1.9 X10 <sup>-12</sup> | 0.58      | 18.1 (12.4-26.4)                                      | 1.9 X10 <sup>-12</sup> | 0.62      | 7.2 X10 <sup>-11</sup>      |
| 2-MeOE <sub>2</sub>         | 37.2 (21.1 to 65.3)                                 | 0.0009                 | 0.18      | 36.2 (19.8-66.2)                                      | 0.001                  | 0.18      | 0.05                        |
| 3-MeOE <sub>1</sub>         | 79.9 (50.6 to 126.2)                                | 0.28                   | 0.02      | 79.0 (49.8-125.3)                                     | 0.311                  | 0.11      | 1                           |
| 2 Hydroxylation pathway EM  | 25 (18.4-34.1)                                      | 1.6 X10 <sup>-12</sup> | 0.58      | 24.1 (17.7-32.88)                                     | 6.6 X10 <sup>-13</sup> | 0.64      | 2.4 X10 <sup>-11</sup>      |
| 4 Hydroxylation pathway EM  | 47.6 (27-83.9)                                      | 0.01                   | 0.11      | 47.6 (26.8-84.4)                                      | 0.01                   | 0.20      | 0.43                        |
| 16 Hydroxylation pathway EM | 134.7 (101.6-178.5)                                 | 0.04                   | 0.07      | 136.6 (102.2-182.7)                                   | 0.03                   | 0.14      | 1                           |
| 16α-OHE <sub>1</sub>        | 191.5 (120.2-305.0)                                 | 0.007                  | 0.12      | 200.9 (123.6-326.5)                                   | 0.005                  | 0.14      | 0.21                        |
| 17-epiE <sub>3</sub>        | 260 (145.9-463.4)                                   | 0.002                  | 0.16      | 295.5 (164.6-528.7)                                   | 0.0005                 | 0.22      | 0.02                        |
| E <sub>3</sub>              | 125.5 (94.8-166)                                    | 0.11                   | 0.04      | 126.1 (94.7-168)                                      | 0.10                   | 0.13      | 1                           |
| 16-ketoE <sub>2</sub>       | 162.3 (108.6-242.6)                                 | 0.02                   | 0.09      | 171.6 (114.7-256.8)                                   | 0.009                  | 0.18      | 0.35                        |

**Table 2.8 Percentage difference in EM levels in 30 premenopausal women who were carriers of the *CYP3A7\*1C* allele (AC genotype) compared to women who were non-carriers (AA genotype) of rs45446698. The results show the output of linear regression analysis with and without adjusting for covariates\*\* age, BMI and parity. The percentage difference of each EM in carriers (AC genotype) was compared to wild type (AA genotype) with the "AA" women's EMs set to 100%. R-squared values represent the proportion of variance explained by genotype in the unadjusted analysis and the proportion of variance explained by all four variables (genotype, BMI, parity and age) in the adjusted model. Methylated-4-catechol EMs (4-MeOE<sub>1</sub> and 4-MeOE<sub>2</sub>) concentrations were below LOD in more than 78% of the samples, therefore, these metabolites were excluded from this analysis.**

Spearman correlation coefficients were also evaluated separately within the datasets of women with the AA and AC genotypes, to explore and compare the associations of individual parent EMs with non-parent EMs and 2-, 4- and 16-hydroxylation pathways. Figure 2.7 shows  $r_s$  values of the correlation within each box of the heat map.

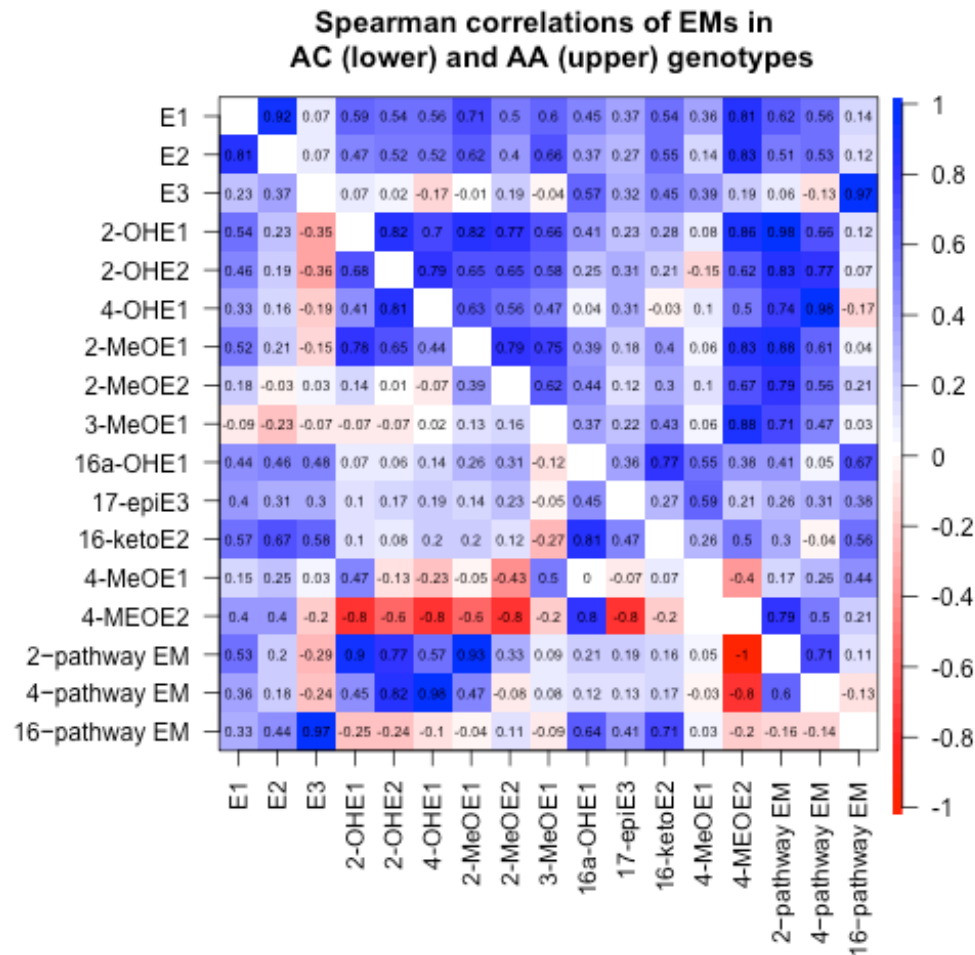


**Figure 2.7 Heatmap of Spearman correlation coefficients for individual EMs and 2-, 4- and 16-hydroxylation pathways, in women (N=30) who were carriers of the CYP3A7\*1C allele (AC genotype) and in women (N=30) who were non-carriers (AA genotype) of rs45446698. The lower left side represents correlations in women with the AC-genotype and the upper right side represents correlations in women with the AA-genotype. Numbers inside the heatmap boxes represents  $r_s$  values.**

In both the AC and AA genotype individuals, the correlations between the parent EMs E<sub>1</sub> and E<sub>2</sub> were quite strong (AC  $r_s$  =0.81; AA  $r_s$  =0.92). The correlations between parent EMs and non-parent EMs in women with both the AC and the AA genotype were weak to modest (AC  $r_s$  -0.23 to 0.67, AA  $r_s$  0.12 to 0.81). In women with the AC genotype, the correlations between E<sub>1</sub> and other non-parent EMs were weak  $\leq$ 0.57 (strongest with 16-ketoE<sub>2</sub>), and the correlations between parent E<sub>2</sub> and non-parent EMs were even weaker,  $r_s$  for most metabolites was  $\leq$ 0.46 (except for 16-KetoE<sub>2</sub>  $r_s$  0.67). E<sub>3</sub>, the most abundant EM in premenopausal urine was weakly inversely

correlated with all EMs except for the 16-hydroxylation pathway and parent EMs. E<sub>3</sub> was also weakly inversely correlated with 2- and 4- hydroxylation pathways ( $r_s$  -0.29 and -0.24 respectively) and all the individual metabolites of the 2- and 4- hydroxylation pathways. Correlations of individual EMs within each pathway in the AC population varied substantially (five EMs within the 2-hydroxylation pathway  $r_s$ = -0.07 to 0.78, lowest for 3-MeOE<sub>1</sub>; three EMs in 4-hydroxylation pathway were inversely correlated  $r_s$  -0.23 to -0.8; four EMs in 16-hydroxylation pathway was modest to strong  $r_s$  0.3 to 0.81). The 2- and 4-hydroxylation pathways were strongly correlated with a  $r_s$  0.6, but weakly inversely correlated with 16-hydroxylation pathway ( $r_s$  -0.14). 4-MeOE<sub>2</sub> was negatively correlated with most individual metabolites except for parent oestrogens and 16 $\alpha$ -OHE<sub>1</sub> and strongly inversely correlated with 2- and 4- pathways  $r_s$  -1 and -0.8. We can rule out any effect of correlations of 4-MeOE<sub>1</sub> and 4-MeOE<sub>2</sub> given that their concentration was below LOD in 80% of AC genotype population.

In the AA genotype population, the correlations between E<sub>1</sub> and most of the other non-parent EMs were modest >0.56 (weakest with E<sub>3</sub>  $r_s$  0.07 and 4-MeOE<sub>1</sub>  $r_s$  0.14), and the correlations between parent E<sub>2</sub> and most of the non-parent EMs were also modest > 0.52 (again weakest for E<sub>3</sub>  $r_s$  0.07 and 4-MeOE<sub>1</sub>  $r_s$  0.12). E<sub>3</sub> was weakly to modestly correlated with most EMs and weakly inversely correlated with 4-OHE<sub>1</sub> and the 4-pathway. Individual EMs within each pathway were modest to strongly correlated (five EMs within the 2-hydroxylation pathway  $r_s$ = 0.58 to 0.82; three EMs in the 4-hydroxylation pathway were weak to modest  $r_s$  0.1 to 0.5; four EMs in 16-hydroxylation pathway were modest to strong  $r_s$  0.3 to 0.77). Similar to as seen in AC genotype population, the AA genotype population also showed that the 2- and 4-hydroxylation pathways were strongly correlated  $r_s$  0.71 but weakly to inversely correlated with 16-hydroxylation pathway  $r_s$  0.11 and -0.13 respectively.

Overall correlations within both AC and AA genotype population showed a similar trend. However, the correlations within individual EMs according to their pathways in AC genotype population were weaker compared to AA genotype population.

## 2.4 Discussion

This study focused on applying a specific and accurate targeted LC-MS/MS method to quantify 14 endogenous oestrogen metabolites, to measure these in follicular phase urine samples from 30 premenopausal women who were carriers of the rare *CYP3A7\*1C* allele (AC genotype) and 30 premenopausal women who were non-carriers (AA genotype) and to explore their association with genotype.

Since urinary EMs concentrations can be as low as picogram per millilitre (pg/mL), the LC-MS/MS method developed for this study provided linear calibration standard curves of each of the 14 EMs over an approximately  $10^3$ -fold concentration range (0.08-20 ng/mL). The LOQ for each EM was 40 pg/0.5 mL urine samples. The linearity range of this method was broad enough to reliably measure most of the EMs concentrations which ranged between 0.04 and 10 ng/0.5 mL urine samples, however, concentrations of  $E_3$  were considerably higher spanning between 20 and 40 ng/mL in 33% (N= 40) and 40 to 76 ng/mL in 8% (N=10) of the samples; the higher concentration of  $E_3$  was expected as it is the most abundant EM in premenopausal urine. On the contrary, measured absolute concentrations were below their LOD (8 pg/0.5 mL urine sample) for methylated-4-catechol EMs: 4-MeOE<sub>1</sub>, 4-MeOE<sub>2</sub> in 78% (N=93), 89% (N=107) urine samples.

Metabolites of the 4-hydroxylation pathway, specifically methylated-4-catechols are reported as the least abundant EMs in the literature [64]. The intra and inter-batch precision (CV%) of the method including the hydrolysis, extraction and derivatisation steps were below 10% except for the two methylated-4-catechol EMs (4-MeOE<sub>1</sub> and 4-MeOE<sub>2</sub>) which had the lowest concentrations and %CVs of between 10–14%. The intra- and inter-batch accuracy measured in the QC samples at a concentration of 2 ng/mL was 94-105%. Overall, the method was accurate and precise enough to be applied in population-based studies of both pre- and postmenopausal women, to determine levels of EMs.

One of the challenges of measuring premenopausal endogenous oestrogen levels is whether a single premenopausal urine/blood measurement of oestrogens can adequately reflect long-term exposure to these hormones (as long-term exposure is of primary interest in the aetiology of hormone-related cancers). In the present study, the premenopausal urine ICCs (Table 2.6) calculated over two follicular days of the one menstrual cycle in N=30 women to assess the intra-individual variability of individual and grouped EMs were very high ranging between 0.64-0.86 (except for methylated-4-catechols). These data suggest reliability and consistency of EMs in follicular phase urine and that one measurement may adequately represent reproducibility over short periods of time. To the best of my knowledge, there are no previously published data which evaluated the reproducibility of follicular phase urinary EMs over longer periods of time.

However, a nested study within the Nurses health study II (NHS II) examined the reproducibility of 15 EMs measured (using the same LC-MS/MS analytical method as used in the present study) in premenopausal women with three mid-luteal urine samples collected over a 3 year period. This showed good intra-individual ICCs  $>0.57$  [64]. In the NHS II study, intra-individual ICCs for 2-, 4- and 16-hydroxylation pathways were high to modest (0.72, 0.57 and 0.52 respectively). Among grouped EMs, ICCs for 2-catechols and the individual catechols were (0.58-0.72). ICCs for grouped and individual methylated catechols were also modest (0.51-0.62) except for methylated-4-catechol where ICCs were low (0.14-0.27). Another study which looked at the reproducibility of plasma (both follicular and luteal:  $E_1$ ,  $E_2$  and  $E_1$  sulfate) and urinary (luteal phase: 2-OHE<sub>1</sub> and 16 $\alpha$ -OHE<sub>1</sub>) oestrogen levels in premenopausal women over a one-year period, showed ICCs for follicular phase plasma oestrogens (ICC 0.53-0.67) were better than ICCs for luteal phase plasma oestrogens (ICCs 0.19 – 0.53). ICCs for luteal urinary 2-OHE<sub>1</sub> and 16 $\alpha$ -OHE<sub>1</sub> were 0.58 and 0.56 respectively [64]. ICCs from these studies compare favourably with several year reproducibility of some of the established clinical biomarkers such as postmenopausal plasma  $E_2$  (ICC=0.68) [61], serum cholesterol (ICC=0.65) [68], blood glucose (ICC=0.52) [64] and blood pressure (ICC= 0.60-0.64) [69]. Results from these studies indicate that a single sample measurement of premenopausal blood or urine from a specific phase (i.e., follicular or luteal) of the menstrual cycle can reasonably represent average oestrogen levels over long-term (1- to 3-years) in that phase.

In the present study of premenopausal follicular phase urinary EMs levels in 30 healthy women with the common AA genotype, the most abundant pathway was the 16-hydroxylation pathway (45 % of total EMs) followed by the 2-hydroxylation pathway (22.1% of total EMs) and parent oestrogens (21.6 % of total EMs), while the 4-hydroxylation pathway was the least abundant (Table 2.7). Within the 16-hydroxylation pathway,  $E_3$  alone contributed to around 38% of total EMs. Apart from  $E_3$ , among the other individual EMs, the most abundant were  $E_1$  (16.9% of total EMs), 2-OHE<sub>1</sub> (12.6% of total EMs), 2-OHE<sub>2</sub> (5.5% of total EMs),  $E_2$  (4.2% of total EMs) and 16 $\alpha$ -OHE<sub>1</sub> (3.9% of total EMs). To the best of my knowledge there is only one previously published study (the Nurses Health Study II) which has reported geometric mean and 5<sup>th</sup>-95<sup>th</sup> percentile ranges (absolute levels of EMs) and arithmetic mean (percent of total EMs), measured using same analytical method as used in this study, in healthy premenopausal women (N=110), in urine collected during the mid-luteal phase of the menstrual cycle [64]. The descriptive statistics (Table 2.7) of most of the individual and grouped EMs presented in this study complements and agrees with the pathway abundance results from the Nurses Health Study II, even though their results show measurements in luteal phase urine compared to the follicular phase urine shown in the present study. The values of the selected ratios of the EMs calculated in the present study and the

Nurses Health Study II [64] also approximately matches with each other. The only differences observed were the values for 2-Catechol EM (2-OHE<sub>1</sub> and 2-OHE<sub>2</sub>) and E<sub>3</sub> were quite different from those reported in the luteal phase (higher in luteal phase). These differences could potentially be driven by the different phases of the menstrual cycle, i.e. luteal versus follicular. In future analysis, luteal phase urine from the women analysed in the present study will provide a direct comparison between luteal versus follicular phase EM differences. The potential use of these data could be to provide a baseline or reference dataset to be compared against specific conditions related to endogenous oestrogen metabolism (e.g. breast cancer). However, the range of baseline or what is considered as average or normal EMs profile could be quite large. The inter-individual levels of EMs could vary up to 10-100 fold depending on age, gender and menstrual cycle status [64].

The main findings of this study were strong significant differences in premenopausal urinary EMs levels in women who were the carriers of the rare *CYP3A7\*1C* allele (AC genotype) compared to those to who were non-carriers (AA genotype) (Table 2.8). The results from the present study indeed support the *a priori* hypothesised genotype-phenotype correlation based on the work of Johnson *et al.* [6, 61], which demonstrated that *CYP3A7\*1C*, is the putative functional allele that is associated with lower circulating urinary E<sub>1</sub>G levels. The current study showed decreased levels of circulating E<sub>1</sub> (54.6%) in carriers of the *CYP3A7\*1C* allele. However, the differences were more pronounced for 2-OHE<sub>1</sub>, 2-MeOE<sub>1</sub>, 2-OHE<sub>2</sub>, 2-MeOE<sub>2</sub> and the 2-hydroxylation pathway where these EMs/pathways were between 18% and 36% in carriers of the *CYP3A7\*1C* allele compared to non-carriers. The opposite effect was observed for metabolites of the 16-hydroxylation pathway, where the difference in 16 $\alpha$ -OHE<sub>1</sub>, 17-epiE<sub>3</sub> and 16-ketoE<sub>2</sub> were between 170% to 296% in carriers of the *CYP3A7\*1C* allele compared to non-carriers. The difference in parent EMs, E<sub>1</sub> and E<sub>2</sub>, was 54% and 74 % respectively, though the difference in E<sub>2</sub> levels was non-significant. Also, R-squared values demonstrated that “genotype” explained a large proportion of the variability of EMs levels in this analysis. The AC genotype explained 64% variability in 2-OHE<sub>1</sub> levels, 62% in 2-MeOE<sub>1</sub> levels and 44% in 2-OHE<sub>2</sub> levels. However, replication of this work is required to validate the findings.

The current data are also supported by another independent study where the *CYP3A7\*1C* allele has been shown to be associated with significantly reduced circulating serum E<sub>1</sub> and dehydroepiandrosterone sulfate (DHEAS) [70]. The results strongly support a role for the *CYP3A7\*1C* allele in determining urinary individual endogenous EMs and supports my observation that the pattern of endogenous oestrogen metabolism may vary substantially among individuals according to their genetic polymorphisms. Also, relatively low correlations between parent oestrogens and non-parent EMs and within pathways suggests that urinary

concentrations of the parent EMs, do not accurately represent the concentrations of individual non-parent EMs. Measuring individual EMs provides information about the pattern of oestrogen metabolism which is not available when only assessing exposure with urinary parent oestrogens. The lack of strong correlations between parent and non-parent EMs and more pronounced associations between SNP (rs45446698) with levels of 2-OHE<sub>1</sub>, 2-MeOE<sub>1</sub> and 2-OHE<sub>2</sub> suggests that these EMs may provide additional information for identifying promising new candidate SNPs and provide new insight into the functional background of these associations.

Interest in EMs derives not only from the fact that selected circulating oestrogens (E<sub>1</sub>, E<sub>2</sub>, ratio of 2-OHE<sub>1</sub> to 16 $\alpha$ -OHE<sub>1</sub>) are established risk factors for breast cancer in postmenopausal women [1, 2, 71, 72], and may be associated with risk among premenopausal women [3, 4, 5, 73], but also from the fact that the laboratory studies have shown that 2-catechol, 4- catechol EMs and 16 $\alpha$ -OHE<sub>1</sub> exhibit differential oestrogenic and genotoxic activities which may play a role in breast carcinogenesis. The carcinogenicity of endogenous oestrogens may be related to unbalanced oestrogen metabolism leading to excess oestrogen quinones and formation of depurination DNA adducts. The individual EMs vary substantially in their ability to induce DNA damage; catechol oestrogens can be oxidised to quinones to induce DNA damage either via direct formation of DNA adducts or indirectly via redox cycling, generating reactive oxygen species (ROS) [33, 74]. 2-catechols are oxidised to form stable and reversible DNA adducts, while 4-catechols form unstable adducts leading to depurination [75, 76]. 16 $\alpha$ OHE<sub>1</sub> has also been shown to increase unscheduled DNA synthesis in an *in vitro* study and can be genotoxic [77]. On the other hand, the methoxy- catechol EMs are hypothesised to reduce the risk of breast cancer either via directly inhibiting tumour growth or decreasing circulating levels of catechol EMs by phase II enzymatic activity of COMT resulting in reduced oestrogenicity [32, 36, 78]. Therefore, the balance between the activities of phase I (oxidation) enzymes CYP 450s and phase II (methylation) enzyme COMT of oestrogen metabolism can be important in hormonally related cancer development. In four prospective studies circulating urinary and serum enhanced 2-hydroxylation pathway activity was shown to be protective against breast cancer risk [51]. A breast cancer case-control study nested within NHS II showed higher 17-epiE<sub>3</sub> levels (a metabolite of 16-hydroxylation pathway) were significantly associated with increased risk of breast cancer [58].

The implication of this work is also very relevant from a clinical and therapeutic strategies standpoint. The CYP3A genes, which encode the enzymes responsible for oestrogen metabolism, are also responsible for the metabolism of other CYP3A substrates including drugs used in chemotherapy treatment of cancer, e.g. Tamoxifen and Docetaxel. Thus, polymorphisms that influence the expression of these genes may influence clinical outcomes [79]. There is reported variability in treatment outcomes among patients receiving adjuvant breast cancer therapy, and

high inter-individual genetic variability in the metabolism of therapeutic agents also plays a role in the prediction of survival [80]. Treatments such as chemotherapy can alter a metabolic phenotype by selective mechanisms, based on the pharmacology of the drugs irrespective of genotype [81]. The frequency of a *CYP3A7\*1C* allele is just 4%, with 8% carriers in the original MOG and BBCs study groups [61], which is in agreement with frequencies observed in previous studies (3.4% to 4.2%) [70]. It is plausible that women who are carriers of the *CYP3A7\*1C* allele, with low levels of 2-OHE<sub>1</sub> and 2-MeOE<sub>1</sub> and high levels of 16 $\alpha$ -OHE<sub>1</sub> and 17-epiE<sub>3</sub> or altered 2- and 16-hydroxylation pathways are also more or less responsive to particular classes of drugs (especially anti-cancer drugs that are CYP3A substrates). Therefore, for the approximately 8% of cancer patients who are *CYP3A7\*1C* carriers, standard chemotherapy regimens that include CYP3A substrates and for which maximum tolerated doses are calculated for the “average” patient, may not be optimal [61]. For these reasons, it is essential to develop a greater understanding of the functional phenotypic implications of genotype.

## 2.5 Strengths and limitations

The strength of this study includes the potential of targeted LC-MS/MS metabolic profiling of oestrogen metabolism to bridge the gap between carefully selected, genotyped population and to identify their functional urinary intermediate phenotypes. Another strength of the study was that measurement of urinary EMs carefully accounted for cyclic variation in all the samples and the results provide descriptive statistics of the follicular phase of the menstrual cycle, avoiding any inconsistency in EMs levels related to different phases of the menstrual cycle.

Limitations of this study were: firstly, clearly, the findings from the current analysis need to be validated in larger epidemiological studies. Secondly, urinary EMs are partially correlated with oestrogenic activity in the breast tissue and may not provide exact oestrogen exposure to the most relevant tissue. However, strong associations between genotype and urinary EMs provides confidence in this approach. Thirdly, associations with luteal phase concentrations may or may not be similar to what I observed in this current study of follicular phase urine and genotype. Finally, although the targeted LC-MS/MS method was shown to be precise, accurate and fit for the purpose, it is limited by the long chromatographic gradient that it requires (77 minutes per sample) which results in the analysis of only 20 urine samples per batch over two days. The long gradient was related to the fact that the mode of chromatography was HPLC, compared to UPLC which would have halved the runtime. Also given the complexity of urine, the longer gradient is required to attain adequate resolution to avoid false identification of EM from co-eluting peaks. In future studies, packed column supercritical fluid chromatography (pSFC) could be used to shorten the gradient runtime to approximately 10 minutes compared to 77 minutes [82].

## 2.6 Conclusion

Hormone-related cancer may be associated with metabolic phenotypes which could be picked up in carefully genotyped populations by measuring related genetically influenced metabolites. The present study is the first to explore the potential of targeted LC-MS/MS metabolic profiling of endogenous oestrogen metabolism, to bridge the gap between potentially functional genetic variations (SNPs) and urinary intermediate metabolic phenotypes which could potentially be risk factors for the breast cancer. The results from the present study confirmed our *a priori* hypothesised genotype-phenotype correlation. The *CYP3A7\*1C* allele, tagged by the SNP rs45446698, and mapping to the *CYP3A* locus has previously been associated with levels of urinary oestrone glucuronide (E<sub>1</sub>G) (54% reduction). In the present study, differences in levels of 2-OHE<sub>1</sub>, 2-MeOE<sub>1</sub>, 2-OHE<sub>2</sub>, 2-MeOE<sub>2</sub> and 2-hydroxylation pathway, in women who were carriers of the *CYP3A7\*1C* allele compared to non-carriers, were more pronounced. The observations from the present study clarified the role of this potentially functional allele in determining urinary individual endogenous EMs by demonstrating altered 2- and 16-hydroxylation pathways in carriers of *CYP3A7\*1C*. This allele is also associated with adverse breast cancer outcomes. Data also showed low correlations between parent and non-parent EMs, suggesting that measuring individual EMs provides information about the pattern of oestrogen metabolism which is not available when only assessing exposure with urinary parent oestrogens.

The wider implications of this work are relevant for identifying promising new candidate SNPs and from a clinical and therapeutic strategies standpoint. It is plausible that women who are carriers of the *CYP3A7\*1C* allele (and who have low levels of 2-OHE<sub>1</sub> and 2-MeOE<sub>1</sub> and high levels of 16 $\alpha$ -OHE<sub>1</sub> and 17-epiE<sub>3</sub> are also more or less responsive to particular classes of drugs (especially anti-cancer drugs that are CYP3A substrates). This suggests that for approximately 8% of cancer patients who are *CYP3A7\*1C* carriers, the standard chemotherapy regimens may not be optimal. These EMs can be used in clinical studies for association with response to drug treatment or when specific complications arise during treatment. Furthermore, the knowledge of the genetic variation in human metabolic individuality will provide a fresh perspective for starting personalised gene-metabolome based healthcare strategies.

## 2.7 References

1. Eliassen, A. H. & Hankinson, S. E. Endogenous hormone levels and risk of breast, endometrial and ovarian cancers: prospective studies. *Adv Exp Med Biol* 630, 148–165 (2008).
2. Thomas, H. V. *et al.* Endogenous estrogen and postmenopausal breast cancer: A quantitative review. *Cancer Causes Control* 8, 922–928 (1997).
3. Hankinson, S.E. & Eliassen, A.H. Circulating sex steroids and breast cancer risk in premenopausal women. *Hormones & cancer*.1 (1), 2-10 (2010).
4. Kaaks, R. *et al.* Serum sex steroids in premenopausal women and breast cancer risk within the European Prospective Investigation into Cancer and Nutrition (EPIC). *J. Natl. Cancer Inst.* 97, 755–765 (2005).
5. Kabuto, M. *et al.* A prospective study of estradiol and breast cancer in Japanese women. *Cancer Epidemiol. Biomarkers Prev.* 9, 575–9 (2000).
6. Johnson, N. *et al.* CYP3A variation, premenopausal estrone levels, and breast cancer risk. *J. Natl. Cancer Inst.* 104, 657–69 (2012).
7. Tsuchiya, Y. *et al.* Cytochrome P450-mediated metabolism of estrogens and its regulation in human. *Cancer Lett.* 227, 115–124 (2005).
8. Dunning, A. M. *et al.* Polymorphisms Associated With Circulating Sex Hormone Levels in Postmenopausal Women. *JNCI J. Natl. Cancer Inst.* 96, 936–945 (2004).
9. Huang, C. S. *et al.* Breast cancer risk associated with genotype polymorphism of the estrogen-metabolizing genes CYP17, CYP1A1, and COMT: A multigenic study on cancer susceptibility. *Cancer Res.* 59, 4870–4875 (1999).
10. Hunter, D. J. *et al.* A candidate gene approach to searching for low-penetrance breast and prostate cancer genes. *Nat Rev Cancer* 5, 977–985 (2005).
11. Haiman, C. A. *et al.* Genetic variation at the CYP19A1 locus predicts circulating estrogen levels but not breast cancer risk in postmenopausal women. *Cancer Res* 67, 1893–1897 (2007).
12. Beckmann, L. *et al.* Comprehensive analysis of hormone and genetic variation in 36 genes related to steroid hormone metabolism in pre- and postmenopausal women from the breast and prostate cancer cohort consortium (BPC3). *J Clin Endocrinol Metab.* 96, 360-367 (2011).
13. Sigurdson, A. J. *et al.* Polymorphisms in estrogen biosynthesis and metabolism-related genes, ionizing radiation exposure, and risk of breast cancer among US radiologic technologists. *Breast Cancer Res. Treat.* 118, 177–184 (2009).
14. Canzian, F. *et al.* Comprehensive analysis of common genetic variation in 61 genes related to steroid hormone and insulin-like growth factor-I metabolism and breast cancer risk in the NCI breast and prostate cancer consortium. *Hum Molecular Genetics.* 19, 387303884 (2010).
15. Doherty, J. A. *et al.* Genetic factors in catechol estrogen metabolism in relation to the risk of endometrial cancer. *Cancer Epidemiol. Biomarkers Prev.* 14, 357–66 (2005).
16. Samavat, H. & Kurzer, M. S. Estrogen metabolism and breast cancer. *Cancer Letter.* 356. 231-243 (2015).
17. Nelson, D. R. *et al.* P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics* 6, 1–42 (1996).
18. Schuetz, J. D. *et al.* Selective expression of cytochrome P450 CYP3A mRNAs in embryonic and adult human liver. *Pharmacogenetics* 4, 11–20 (1994).
19. Slaughter, R. L. & Edwards, D. J. Recent advances: the cytochrome P450 enzymes. *Ann. Pharmacother.* 29, 619–24 (1995).

20. Williams, E. T. *et al.* Estrogen regulation of the cytochrome P450 3A subfamily in humans. *J. Pharmacol. Exp. Ther.* 311, 728–735 (2004).
21. Kuehl, P. *et al.* Sequence diversity in CYP3A promoters and characterization of the genetic basis of polymorphic CYP3A5 expression. *Nat. Genet.* 27, 383–391 (2001).
22. Burk, O. *et al.* Molecular mechanisms of polymorphic CYP3A7 expression in adult human liver and intestine. *J. Biol. Chem.* 277, 24280–24288 (2002).
23. Martucci, C. P. & Fishman, J. P450 enzymes of estrogen metabolism. *Pharmacol. Ther.* 57, 237–57 (1993).
24. Zhu, B. T.. Functional role of estrogen metabolism in target cells: review and perspectives. [Review]. *Carcinogenesis* 19, 1–27 (1998).
25. Raftogianis, R. *et al.* Estrogen metabolism by conjugation. *J. Natl. Cancer Inst. Monogr.* 113–124 (2000).
26. Tsuchiya, Y. *et al.* Cytochrome P450-mediated metabolism of estrogens and its regulation in human. *Cancer Lett.* 227, 115–124 (2005).
27. Van Casper H. *et al.* Binding of 2-hydroxyestradiol and 4-hydroxyestradiol to estrogen receptors from human breast cancers. *J. Steroid Biochem.* 32, 485–492 (1989).
28. Vandewalle, B. & Lefebvre, J. Opposite effects of estrogen and catecholesterogen on hormone-sensitive breast cancer cell growth and differentiation. *Molecular Cell Endocrinology.* 61,239–46 (1989).
29. Telang, N. T. *et al.* Estradiol metabolism: an endocrine biomarker for modulation of human mammary carcinogenesis. *Environ. Health Perspect.* 105 Suppl , 559–64 (1997).
30. Bradlow, H. L. *et al.* 2-hydroxyestrone: the ‘good’ estrogen. *J Endocrinol* 150 Suppl, S259–65 (1996).
31. Dawling, S. *et al.* Methoxyestrogens exert feedback inhibition on cytochrome P450 1A1 and 1B1. *Cancer Res.* 63, 3127–3132 (2003).
32. Lahani, N. J. *et al.* 2-Methoxyestradiol, a promising anticanceragent. *Pharmacotherapy* 23, 165–72 (2003).
33. Liehr, J. G. & Roy, D. Free radical generation by redox cycling of estrogens. *Free Radic. Biol. Med.* 8, 415–423 (1990).
34. Cavalieri, E. L. *et al.* Molecular origin of cancer: catechol estrogen-3,4-quinones as endogenous tumor initiators. *Proc. Natl. Acad. Sci. U. S. A.* 94, 10937–10942 (1997).
35. Lavigne, J. *et al.* The effects of catechol-O-methyltransferase inhibition on estrogen metabolite and oxidative DNA damage levels in estradiol-treated MCF-7 cells. *Cancer Res.* 61, 7488–7494 (2001).
36. Lu F. *et al.* Estrogen metabolism and formation of estrogen-DNA adducts in estradiol-treated MCF-10F cells. The effects of 2,3,7,8-tetrachlorodibenzo-p-dioxin induction and catechol-O-methyltransferase inhibition. *J Steroid Biochem Mol Biol.* 105(1-5),150-158 (2007).
37. Huang Z. *et al.* 16Alpha-hydroxylation of estrone by human cytochrome P4503A4/5. *Carcinogenesis.* 19(5), 867-72 (1998).
38. Badawi, F. *et al.* Role of human cytochrome P450 1A1, 1A2, 1B1, and 3A4 in the 2-, 4-, and 16alpha-hydroxylation of 17beta-estradiol. *Metabolism.* 50, 1001–3 (2001).
39. Anthony J. Lee. *et al.* Human Cytochrome P450 3A7 Has a Distinct High Catalytic Activity for the 16-Hydroxylation of Estrone but not 17-Estradiol. *CANCER RESEARCH.* 63, 6532–6536 (2003).
40. Suto, A. *et al.* Experimental down-regulation of intermediate biomarkers of carcinogenesis in mouse mammary epithelial cells. *Breast Cancer Res. Treat.* 27, 193–202 (1993).

41. Bradlow, H.L. *et al.* Estradiol 16 alpha-hydroxylation in the mouse correlates with mammary tumor incidence and presence of murine mammary tumor virus: a possible model for the hormonal etiology of breast cancer in humans. *Proc Natl AcadSci U S A.* 82(18),6295-9 (1985).
42. Osborne, M. P. *et al.* Upregulation of estradiol C16 alpha-hydroxylation in human breast tissue: a potential biomarker of breast cancer risk. *J Natl Cancer Inst* 85, 1917–1920 (1993).
43. Ursin G. *et al.* Urinary 2-hydroxyestrone/16alpha-hydroxyestrone ratio and family history of breast cancer in premenopausal women. *Breast Cancer Res Treat.* 72(2),139-43 (2002).
44. Muti, P. *et al.* Estrogen metabolism and risk of breast cancer: a prospective study of the 2:16alpha-hydroxyestrone ratio in premenopausal and postmenopausal women. *Epidemiology* 11, 635–640 (2000).
45. Goodwin, B. *et al.* Regulation of cyp3a gene transcription by the pregnane x receptor. *Annu. Rev. Pharmacol. Toxicol.* 42, 1–23 (2002).
46. Seredina, T. A. *et al.* Association of cytochrome P450 genetic polymorphisms with neoadjuvant chemotherapy efficacy in breast cancer patients. *BMC Med. Genet.* 13, 45 (2012).
47. Tayeb, M.T. *et al.* CYP3A4 and VDR gene polymorphisms and the risk of prostate cancer in men with benign prostate hyperplasia. *Br. J. Cancer.* 88, 9280932 (2003).
48. Spurdle, A. B. *et al.* The CYP3A4\*1B polymorphism has no functional significance and is not associated with risk of breast or ovarian cancer. *Pharmacogenetics* 12, 355–366 (2002).
49. Lee, S. S. *et al.* Identification of a null allele of cytochrome P450 3A7: CYP3A7 polymorphism in a Korean population. *Mol. Biol. Rep.* 37, 213–7 (2010).
50. Zhou, S.-F. Drugs behave as substrates, inhibitors and inducers of human cytochrome P450 3A4. *Curr. Drug Metab.* 9, 310–22 (2008).
51. Ziegler, R. G. *et al.* Review: Epidemiological studies of estrogen metabolism and breast cancer. *Steroids.* 99, 67-75 (2015).
52. Xu, X. *et al.* Measuring fifteen endogenous estrogens simultaneously in human urine by high-performance liquid chromatography-mass spectrometry. *Anal. Chem.* 77, 6646–6654 (2005).
53. Stanczyk, F. Z. *et al.* Standardization of steroid hormone assays: Why, how, and when? *Cancer Epidemiol. Biomarkers Prev.* 16, 1713–1719 (2007).
54. Xu, X. *et al.* A liquid chromatography-mass spectrometry method for the quantitative analysis of urinary endogenous estrogen metabolites. *Nat. Protocol.* 2, 1350–1355 (2007).
55. Fuhrman, B. J. *et al.* Estrogen metabolism and risk of breast cancer in postmenopausal women. *J. Natl. Cancer Inst.* 104, 326–339 (2012).
56. Falk, R. T. *et al.* Relationship of serum estrogens and estrogen metabolites to postmenopausal breast cancer risk: a nested case-control study. *Breast Cancer Res.* 15, R34 (2013).
57. Dallal, C. M. *et al.* Estrogen metabolism and breast cancer risk among postmenopausal women: a case-cohort study within B~FIT. *Carcinogenesis* 35, 346–355 (2014).
58. Eliassen, A. H. *et al.* Urinary estrogens and estrogen metabolites and subsequent risk of breast cancer among premenopausal women. *Cancer Res.* 72, 696–706 (2012).
59. Walker, K. *et al.* Premenopausal mammographic density in relation to cyclic variations in endogenous sex hormone levels, prolactin, and insulin-like growth factors. *Cancer Res.* 69, 6490–6499 (2009).
60. Johnson, N. *et al.* Interaction between CHEK2\*1100delC and other low-penetrance breast-cancer susceptibility genes: A familial study. *Lancet* 366, 1554–1557 (2005).
61. Johnson, N. *et al.* CYP3A7 and cancer outcome. The CYP3A7\*1C allele is associated with adverse outcome in breast and lung cancers and chronic lymphocytic leukemia. *Cancer Res.* 76(6):1485-1493 (2016).

62. Hankinson, S. E. *et al.* Reproducibility of plasma hormone levels in postmenopausal women over a 2-3-year period. *Cancer Epidemiol. Biomarkers Prev.* 4, 649–54 (1995).
63. Olivier, Jake. *et al.* The logarithmic transformation and the geometric mean in reporting experimental IgE results: what are they and when and why to use them. *Annals of Allergy, Asthma & Immunology* vol. 100 (4) p. 333-337 (2208).
64. Eliassen, A. H. *et al.* Reproducibility of fifteen urinary estrogens and estrogen metabolites over a 2- to 3-year period in premenopausal women. *Cancer Epidemiol. Biomarkers Prev.* 18, 2860–2868 (2009).
65. U.S. Department of Health and Human Services, Food and Drug Administration. Guidance for Industry, Bioanalytical Method Validation (2001). at <http://www.fda.gov/downloads/Drugs/Guidances/ucm070107.pdf>
66. European medicine agency. Guideline to Bioanalytical method validation (2011). At [http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Scientific\\_guideline/2011/08/W\\_C500109686.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2011/08/W_C500109686.pdf)
67. Michaud, D. S. *et al.* Reproducibility of plasma and urinary sex hormone levels in premenopausal women over a one-year period. *Cancer Epidemiol. Biomarkers Prev.* 8, 1059–1064 (1999).
68. Shekelle, R. B. *et al.* Diet, Serum Cholesterol, and Death from Coronary Heart Disease. *N. Engl. J. Med.* 304, 65–70 (1981).
69. Rosner, B. *et al.* Age-specific correlation analysis of longitudinal blood pressure data. *Amer J Epidemiol* 106, 306–13 (1977).
70. Smit, P. *et al.* A common polymorphism in the CYP3A7 gene is associated with a nearly 50% reduction in serum dehydroepiandrosterone sulfate levels. *J. Clin. Endocrinol. Metab.* 90, 5313–6 (2005).
71. Key, T. *et al.* Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. *J. Natl. Cancer Inst.* 94, 606–16 (2002).
72. Missmer, S.A. *et al.* Endogenous estrogen, androgen, and progesterone concentrations and breast cancer risk among postmenopausal women. *J. Natl. Cancer Inst.* 96, 1856–1865 (2004).
73. Thomas, H. V *et al.* A prospective study of endogenous serum hormone concentrations and breast cancer risk in pre-menopausal women on the island of Guernsey. *Br. J. Cancer* 76, 401–405 (1997).
74. Yager, J. D. & Liehr, J. G. Molecular mechanisms of estrogen carcinogenesis. *Annu. Rev. Pharmacol. Toxicol.* 36, 203–232 (1996).
75. Cavalieri, E. *et al.* Estrogens as endogenous genotoxic agents--DNA adducts and mutations. *J Natl Cancer Inst Monogr* 75–93 (2000).
76. Yue, W. *et al.* Genotoxic metabolites of estradiol in breast: Potential mechanism of estradiol induced carcinogenesis. in *J. Steroid Biochem. Mol. Biol.* 86, 477–486 (2003).
77. Telang, N. T. *et al.* Induction by estrogen metabolite 16 alpha-hydroxyestrone of genotoxic damage and aberrant proliferation in mouse mammary epithelial cells. *J. Natl. Cancer Inst.* 84, 634–638 (1992).
78. Fukui, M. & Zhu, B. T. Mechanism of 2-methoxyestradiol-induced apoptosis and growth arrest in human breast cancer cells. *Mol. Carcinog.* 48, 66–78 (2009).
79. Rodriguez-Antona, C. & Ingelman-Sundberg, M. Cytochrome P450 pharmacogenetics and cancer. *Oncogene* 25, 1679–1691 (2006).
80. Choi, J.-Y. *et al.* The role of genetic variability in drug metabolism pathways in breast cancer prognosis. *Pharmacogenomics* 7, 613–24 (2006).

81. Bathe, O. F. & Farshidfar, F. From genotype to functional phenotype: unravelling the metabolomic features of colorectal cancer. *Genes (Basel)*. **5**, 536–60 (2014).
82. Xu, X. *et al.* Analysis of fifteen estrogen metabolites using packed column supercritical fluid chromatography-mass spectrometry. *Anal. Chem.* **78**, 1553–8 (2006).

## **Chapter 3: Targeted mass spectrometry (LC-MS/MS and FIA-MS/MS) based metabolic profiling exploring the biochemical effects of persistent organic pollutants exposure in maternal and umbilical cord serum within the RHEA mother–child cohort study**

### **3.1 Introduction**

#### **3.1.1 Persistent organic pollutants (POPs)**

Persistent organic pollutants POPs are man-made chemicals comprised of organic, environmental pollutants such as polychlorinated biphenyls (PCBs) and organochlorine pesticides including hexachlorobenzene (HCB) and dichlorodiphenyldichloroethylene (DDE). They are heterogeneous in structure and are composed mainly of carbon, hydrogen and several chlorine atoms per molecule. These chemicals share many characteristics in common including the ability to resist biological, chemical and photolytic degradation. Thus, once released into the environment they persist in the environment for a very long time and are subject to long-transboundary transport via humans, animals, birds and atmospheric transport [1, 2].

The major route of POPs exposure in humans is via the food chain. Around 90% of POPs total body burden comes from an animal-based diet such as consumption of fish, meat and dairy products. Other forms of exposure are air inhalation and skin contact. Once inside the body they are rapidly absorbed by the small intestine, diffuse across cell membranes, enter blood vessels and the circulatory system and are eventually stored in soft tissue compartments rich in lipids, mainly adipose tissues. Persistent chemicals have high lipid affinity in part due to the presence of multiple chlorine atoms in their structure which make them lipophilic. Once bioaccumulated in adipose tissue, they can re-enter the bloodstream (via plasma or serum lipids) and are also actively secreted in breast milk [1, 2, 3]. These chemicals have been easily detected and quantified in human biofluids and tissues such as blood (cord blood, placenta, maternal plasma and serum), adipose tissue (especially breast), breast milk, and urine [1, 2].

Since POPs bioaccumulate potential adverse effects on birth outcomes and children's health are a matter of public health concern. There is emerging evidence that background exposure to low-dose of these persistent chemicals is linked with adverse health effects in humans such as obesity [3, 4], endocrine and reproductive defects [4, 5, 6], neurobehavioral abnormalities [7, 8, 9] and breast and testicular cancer [10, 11]. Although numbers of studies are going on to understand the molecular pathways involved in POP toxicity, little is known about the relationship between low-dose exposure and diseases caused by their toxicity in humans. The research from *in-vivo* and *in-vitro* studies shows that the mechanism of organochlorine POP toxicity is believed to be related to their ability to disrupt genes, hormones and enzymes involved in steroid and xenobiotic

metabolism [1, 2, 3]. A better understanding of how early life (pre-and postnatal) exposure to the chemicals influences the burden of childhood disease is an important step forward in improving public health.

### 3.1.1.1 PCBs (Polychlorinated biphenyls)

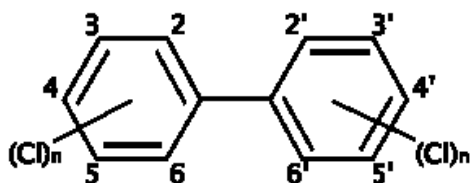


**Figure 3.1 Molecular structure of PCB (Polychlorinated biphenyls)**

Since 1929 around 2 million tonnes of PCBs have been produced and even after its ban in 1970's around 10% of PCBs still exist in the environment. Major uses included dielectric isolators in electrical equipment and heat transfer fluids and lubricants. Figure 3.1 shows the molecular structure of PCBs which consist of two connected phenyl rings. In PCBs hydrogen atoms attached to carbon on these phenyl rings are replaced by one of several chlorine atoms. Carbon numbering represents the different position of chlorine atoms where  $n$  ranges from 1-10. PCBs occur in 209 different configurations or congeners, but only 130 have been identified in commercial products. There are 12 coplanar PCB congeners which are dioxin-like and can either be toxic or non-toxic. PCBs are named either according to the position of the chlorine atom (e.g. 3,4,4',5-tetrachlorobiphenyl) or by simply assigning a number from 1 to 209 for each specific congener. Since different PCB congeners exhibit different toxic or non-toxic profiles, therefore, it is important to consider specific individual PCB congeners when assessing the health risk of PCBs in humans. Although PCBs enter the environment as a commercial mixture, the accumulation and retention of specific PCBs do not reflect the PCB profile of commercial formulations [12].

PCBs are metabolised by microsomal cytochrome P450 (CYP 450) enzymes in the liver to less lipophilic metabolites and conjugated with glucuronic acid to be excreted in urine or faeces. The rate of metabolism is dependent upon the chlorine pattern on the phenyl ring in the specific congener. Many studies provide strong evidence that the mechanism of toxicity of dioxin-like PCB congeners involves initial binding to the aryl hydrocarbon receptor (AhR), a ligand-activated transcription factor involved in biological detoxification of exogenous chemicals, in the cytosol of the target cell. This results in subsequent changes in gene expression including activation of CYP1A1/1A2, leading to toxic responses which are found to be similar to those caused by 2,3,7,8-TCDD and other related halogenated aromatic hydrocarbons [3, 12]. The neurological effects of certain PCB congeners which are not effective AhR agonists have been hypothesised to be linked

to their ability to interfere with calcium homeostasis [13]. However, the mechanisms of other PCB congeners which are associated with estrogenic or carcinogenic activity have not been fully elucidated [12, 13].

#### 3.1.1.2 HCB (Hexachlorobenzene)

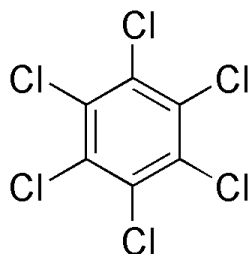


**Figure 3.2 Molecular structure of HCB (Hexachlorobenzene)**

HCB was first introduced in 1945 and was widely used as fungicide especially for seed treatment of wheat. As of 2003, it has been banned from use in 23 countries but its occurrence as an industrial by-product of perchloroethylene, trichloroethylene and pentachlorobenzene are less controlled, and it is also a known low-level impurity of several other fungicides [14]. Its half-life in humans is estimated to be around 9 years, and the most sensitive target organs are liver, ovary and the central nervous system [1, 14, 16]. The current HCB emission is approximated at 23,000 kg/yr with a substantial amount coming from volatilisation from 'old soil' from past contamination along with unidentified sources [16]. Its molecular structure is composed of fully chlorinated aromatic hydrocarbon as shown in Figure 3.2.

Limited research data is available on the mechanism of HCB toxicity in endocrine, and immunological effects and the main focus has been on hexachlorobenzene-induced porphyria- a disorder of heme biosynthesis. The most noticeable health effect of HCB was in eastern Turkey between 1954 and 1959 where consumption of HCB-treated seed grain induced porphyria in approximately 3000 people. Children of pregnant women who ingested the seed developed "pembe yara" or pink sore, with a reported mortality rate of approximately 95% of children under the age of 2 [15]. A 20-30-year follow-up study of 200 people showed that symptoms of porphyria still existed [17]. Porphyrin urine excretion was found to be significantly elevated. Porphyrin metabolism in the liver is considered as the main target of hexachlorobenzene in short- or long-term exposure. Although not only specific to HCB, abnormalities in heme biosynthesis and porphyria can be considered an effective primary marker of HCB exposure in humans.

HCB is shown to be a weak agonist for AhR. However HCB does have a strong affinity for the oestrogen receptor (ER) [1]. An *in vivo* study has also shown that HCB can elicit tumorigenic activity by inducing alterations in insulin-like growth factors (IGFs) [18].

### 3.1.1.3 DDE (Dichlorodiphenyldichloroethylene)

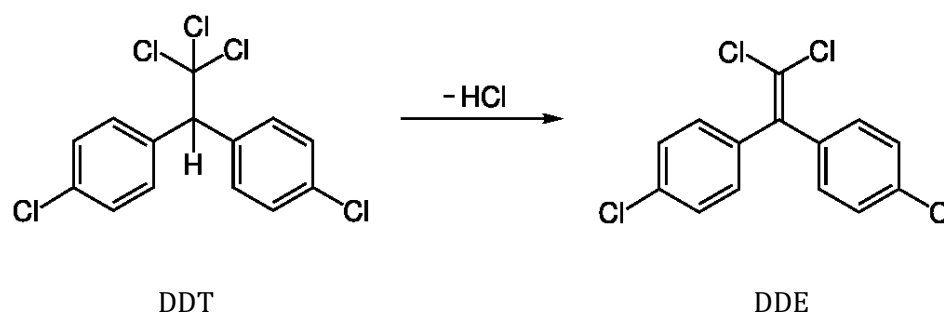


**Figure 3.3 Molecular structure of DDT (dichlorodiphenyltrichloroethane) & DDE (dichlorodiphenyldichloroethylene)**

DDT was extensively used during the Second World War against the spread of malaria and typhus and afterwards it was widely used as a pesticide in agriculture. The global production of DDT was approximately 1.8 million tonnes in the mid-nineties, and the current global production stands at 3314 tonnes (2009) [19]. Figure 3.3 shows the molecular structures of DDT and DDE. Due to growing environmental concerns and the adverse impact on wild birds, its use was heavily restricted in the 1970s, but it is still used for vector control for malaria. DDT and its metabolite DDE are very persistent and are known to remain in the human body for 50 years while 50% of it can remain in the soil for 10-15 years [1, 19]. DDE is more lipophilic than DDT, and it is also the main bioaccumulative chemical in the DDT family which is easily found in human tissues.

DDT and its metabolite DDE are known endocrine disruptors. They exhibit hormonal activity by influencing endocrine and reproductive mechanisms through their interaction with the steroidogenic pathway for oestrogens and androgens. Their endocrine effect results from a combination of mechanisms. They possess the ability to mimic endogenous oestrogens (17- $\beta$  oestradiol), and thus their intracellular mechanism of action is mediated through the oestrogen receptor (ER) or androgen receptor (AR). Receptor binding blocks the endogenous hormone hence DDT/DDE act to antagonise endogenous androgenic hormones, modifying ER and AhR levels and interfering with the metabolism of endogenous hormones further perturbing normal functioning of cytochrome P450 enzymes [1, 19].

Neurological effects of DDT and DDE have been shown to be related to their ability to interfere with the movement of ions through neuronal membranes in the central nervous system [19]. The reduced level of the neurotransmitter serotonin is also being linked to exposure to DDT and DDE, and it is suggested that it is due to the effect of DDT on membrane transport which influences neurotransmitter release [19].

### 3.1.2 Prenatal exposure to persistent organic pollutants and pregnancy outcomes

Persistent organic pollutants (POPs) are developmental toxicants, and epidemiological studies have suggested that fetuses are particularly vulnerable to the effects of these environmental hazards. Human exposure to POPs including PCBs, HCB and DDE begins *In utero*, and three main factors affect the vulnerability of the developing fetus. First is that the active metabolites of POPs, by virtue of their low molecular weight, lipid affinity, non-polarity, and low protein binding properties can cross the placenta with ease and evidence suggests that environmental chemicals with high molecular weight can also reach the fetus [20, 21]. The second factor is the dose of these environmental chemicals and the third and most important is the time of exposure [22, 23]. Compared to adults, the fetus and infants are more vulnerable to the adverse effects of environmental chemicals due to the immaturity of their metabolism, immune system and their detoxification mechanisms which are not fully functional [23]. Foetal exposure to POPs occurs primarily through the maternal diet. Studies have demonstrated that the diet of pregnant women is associated with the levels of POPs found in maternal and cord blood as well as in placenta. Postnatal exposure continues by transmission of these chemicals via breastfeeding, as breast milk is one of the main elimination pathways for POPs [24].

*In utero* exposure to high concentrations of POPs causes foetal toxicity but the evidence at low exposure levels remains unclear and limited. A meta-analysis including 12 European birth cohorts showed that low-level exposure to PCB impairs foetal growth where birth weight decreased with increased PCB concentration in cord serum [25]. In some studies, maternal and cord serum levels of PCBs and DDE has been associated with preterm births, and an increase in “small for gestational age” (SGA) babies [26, 27, 28]. POPs including PCBs, HCB and DDE have also been hypothesised to be obesogenic [29]; a recent study has shown a positive association between 1<sup>st</sup>-trimester maternal serum HCB and DDE levels and rapid growth and becoming overweight in infancy [30]. Another study showed a strong positive association between prenatal PCBs and DDE exposure and being overweight in girls at the age of 6.5 years; the association in overweight boys was DDT and diet-dependent [31]. *In utero* exposures to PCBs and organochlorine pesticides (HCB and DDE) has been long linked to neurological deficits in neonates and young children [7, 8, 9, 32]. Various mother-child cohort epidemiological studies have linked prenatal PCBs and DDE exposure to impaired cognitive, quantitative, verbal, and memory components of child neurodevelopment. [33, 34]. *In utero* exposures, PCBs and organochlorine pesticides have also been linked to changes in male reproductive health, including congenital malformations, shorter anogenital distance in boys, reduced male fertility related to sperm quality and testicular cancer in adult life [11, 35, 36].

### 3.1.3 Metabolic profiling to identify biomarkers of exposure to POPs

To understand how the 'early-life exposome' affects the development of children and influences the risk of disease in later life, metabolic profiling can be used as an integrative tool to increase our understanding of biological disruption and mode of biochemical action induced by chemical exposures.

Since samples from prospective studies are in principle not influenced by the inherent metabolic changes due to the disease itself. Therefore, this setting is the more suitable for research of disease aetiology and to identify any early metabolic change leading to biological effects. The most common biofluids to assess prenatal exposure to POPs are placental, maternal and cord serum. In recent years, the number of studies reporting prenatal concentrations of POPs has increased. However, to date, there are only three previously published prospective population-based studies which explored the associations between environmental pollutants and their influence on body fluid metabolite profiles. The first study looked at  $^1\text{H}$  NMR-based urinary metabolite profiles and exposure to cadmium in 178 people living close to the Avonmouth zinc smelter [37]. In this study, six urinary metabolites involved in mitochondrial metabolism (citrate, 3-hydroxyisovalerate, 4-deoxyerythronic acid) and three metabolites in one-carbon metabolism (dimethylglycine, creatinine and creatine) were associated with cadmium exposure. The second study explored the associations between  $^1\text{H}$  NMR spectroscopic based serum metabolite profiles in 51 male workers exposed to welding fumes (chromium, nickel, and manganese particles) in Taiwan. Results showed increased levels of glycine, taurine, betaine/TMAO, serine, S-sulfocysteine, hippurate, gluconate, creatinine, and acetone among workers exposed to welding fumes. These metabolites were shown to be involved in the inflammatory and oxidative tissue injury processes [38]. There is only one pregnancy-based prospective epidemiological study which looked at the modifications in the 1<sup>st</sup>-trimester urine metabolite profile of pregnant women and its association with prenatal environmental pesticides using  $^1\text{H}$  NMR spectroscopic based metabolic profiling [39]. This study reported a positive association between glycine, threonine, lactate and glycerophosphocholine and a negative association with citrate and 1<sup>st</sup>-trimester parental pesticide exposure.

The present study is the first to use targeted MS-based hypothesis on 1<sup>st</sup>-trimester pregnancy serum and cord serum with contrasted prenatal exposure to POPs (PCBs, HCB and DDE) with the aim to discriminate metabolite signature of exposures and link these to a possible mechanism of toxicity.

### 3.1.4 Biocrates AbsoluteIDQ P180 assay

In this study, the Biocrates AbsoluteIDQ p180 assay was used to identify and quantify 185 metabolites across four metabolite classes via LC and FIA tandem mass spectrometry. The decision to choose Biocrates AbsoluteIDQ p180 assay was based on the following reasons:

1. The need for measurements of identified metabolites rather than unidentified metabolite features or  $m/z$  values. The RHEA cohort is part of an exposome project where data are shared across many researchers and integrated with datasets generated by other omics platforms. Metabolite identification is a laborious and difficult process. A targeted method allows ready interpretation by non-analytical scientists and makes global data integration with other omics immediately possible. However, this targeted approach comes at the cost of metabolome coverage, i.e. losing any novel associations which may be uncovered using untargeted approach. Although a significant compromise in terms of the number of metabolites that can be detected (~5% of the serum metabolome [ Ref 17, 21 from Chapter 1]), a targeted panel of ~180 measurements would still provide more insight into the relationships between abundant serum metabolites and POP exposure than had ever been reported to date.
2. The Biocrates assay required low sample volume, which was important as we were limited to minimal volumes of a precious sample bank, and was relatively high throughput (80 samples per day).
3. Several metabolites were fully quantified. Measurements quantified in absolute terms can in principle be directly compared and integrated at the whole profile level with quantitative data from other platforms. Untargeted metabolomics data is not amenable to such meta-analysis.
4. The platform is highly reproducible across laboratories and LC-MS instrumentation (7.6% average interlaboratory CV,) [40].

A major limitation for the platform centres on the FIA-MS/MS part of the assay which is used for acylcarnitines and lipid profiling. Most quantification in this part of the assay is not absolute but 'semi-quantitative' because only a few internal standards were used and one-point calibration was implemented for their quantification. A disadvantage of using one-point calibration is that measured results/concentrations can be unreliable if the instrument response is distant from calibration sample or the detector response is not linear. Another limitation that is more common to direct infusion methods is that lipids are identified in by total fatty acid carbons and unsaturated bonds, eg. PC acyl 40:5. Dozens of possible lipid species, arising from different combinations of the length of the fatty acid moiety and exact position of double bonds, could

contribute to a given signal. Hence precise chemical structure cannot be determined without further analysis, either further fragmentation, on-line or off-line separation (LC or GC). However, even with these limitations given the rich lipid content of the serum, an enormous amount of intact lipids are measured using Biocrates assay.

### **3.1.5 Hypothesis and Aim**

This study was based on the hypothesis that exposure to low-dose persistent environmental toxicants could alter the metabolome and these changes could impact molecular pathways which could, in turn, affect the functioning of the entire organism [41, Ref 7, 8 Chapter 1]. The aim of the study was to measure LC-MS/MS metabolic profiles in the mother-child cohort study RHEA and explore the association of these metabolic profiles with pre-natal persistent organic pollutants.

### **3.1.6 Objectives**

The objectives of this study were to:

- Use targeted mass spectrometry-based metabolic profiling which allows wide coverage of five classes of metabolites (185 metabolites) to explore and compare the metabolome of healthy maternal-foetal (cord) serum.
- To explore whether the prenatal low-dose exposure of persistent organic pollutants are associated with metabolite profiles in 1<sup>st</sup>-trimester maternal serum and cord serum of healthy pregnancies.
- To explore the links of discriminant metabolites of exposure to molecular pathways that could influence growth, health and development of children.

## 3.2 Materials and Methods

### 3.2.1 Materials

The Biocrates Absolute*IDQ* p180 assay kit was purchased from BIOCRATES Life Science AG, Austria. Each kit contained reagents and solvents needed to prepare a 96-well plate assay including a 96-deep-well Absolute*IDQ* kit plate attached to a filter plate, seven vials of Biocrates calibration standards (lyophilised), one vial of Biocrates internal standards (lyophilised) and 3 vials of Biocrates Quality control samples (lyophilised). Biocrates Met*IDQ*<sup>TM</sup> software was also included which allowed automation of the assay workflow, from sample registration to data processing. Other solvents and reagents including ethanol (99.9%), methanol (99.9%), pyridine (99.8%), water (99.9%), acetonitrile (99.9%), ammonium acetate ( $\geq 99\%$ ), formic acid (LC-MS Ultra) and phosphate buffered saline were all HPLC grade and purchased from Sigma-Aldrich. The derivatisation reagent phenylisothiocyanate ( $\geq 99.0\%$ ) was also bought from Sigma-Aldrich.

The UPLC-ESI-MS/MS instrument consisted of Waters Xevo® tandem quadrupole stepWave<sup>TM</sup> (TQ-S) mass spectrometer coupled with a Waters ACQUITY UPLC® system. Both the UPLC and the MS were controlled by Waters MassLynx<sup>TM</sup> MS Software. Chromatographic separation was carried out using a Waters ACQUITY UPLC BEH C18 1.7 $\mu$ M 2.1 x 50mm column, the pre-column was Waters ACQUITY UPLC BEH C17 1.7  $\mu$ M VANGUARD. Other equipment required were: Nitrogen evaporator, Eppendorf MixMate® shaker and a centrifuge for 96-well plate microtiter plates of 5cm height at 100 x g.

### 3.2.2 Biocrates Absolute*IDQ* p180 assay

The analytical platform employed in this study was the Biocrates Absolute*IDQ* p180 assay designed to be used with Waters tandem quadrupole stepWave<sup>TM</sup> (TQ-S) mass spectrometer. This method allowed simultaneous identification and quantification of up to 185 endogenous metabolites in 10  $\mu$ L of serum sample using combined electrospray liquid chromatography-tandem mass spectrometry (ESI LC-MS/MS) and direct infusion flow injection analysis-tandem mass spectrometry (FIA-MS/MS). The LC method allows for the absolute quantification of 21 amino acids and 19 biogenic amines, whereas the FIA method allows the semi-quantitative measurement of 40 acylcarnitines (Cx:y), 90 glycerophosphatidylcholines (PCs, aa = diacyl, ae = acyl-alkyl), 14 sphingolipids (SMx:y) and hexose sugar. The lipid side chain composition is abbreviated as Cx:y, where x denotes the number of carbons in the side chain and y denotes the number of double bonds. The exact position of the double bonds and carbon atoms distribution in different fatty acid chains cannot be determined in this method. A list of the five classes of

metabolites measured using this assay in both LC and FIA mode is shown in Table 3.1, and lists of individual metabolites measured in each class are shown in Appendix Table 1 & Table 2.

| Metabolite class          | Mode of measurement | Absolute quantification/<br>semi quantitative                                                     | Appendix (list 1 & 2 of individual metabolites)                     |
|---------------------------|---------------------|---------------------------------------------------------------------------------------------------|---------------------------------------------------------------------|
| Amino acids (21)          | LC-MS/MS            | Absolute quantification using seven-point calibration curve and stable isotope internal standards | 1.List of metabolites measured using Biocrates p180 LC-MS/MS assay  |
| Biogenic amines (18)      | LC-MS/MS            | Absolute quantification using seven-point calibration curve and stable isotope internal standards |                                                                     |
| Acylcarnitines (40)       | FIA-MS/MS           | Semi quantitative, one-point calibration                                                          | 2.List of metabolites measured using Biocrates p180 FIA-MS/MS assay |
| Glycerophospholipids (89) | FIA-MS/MS           | Semi-quantitative, one-point calibration                                                          |                                                                     |
| Sphingolipids (14)        | FIA-MS/MS           | Semi-quantitative, one-point calibration                                                          |                                                                     |
| Hexose (1)                | FIA-MS/MS           | Semi-quantitative, one-point calibration                                                          |                                                                     |

**Table 3.1 List of five classes of metabolites measured using Biocrates AbsoluteIDQ p180 assay, in parenthesis number of metabolites measured in each class, mode of measurement (LC or FIA), quantification (absolute or semi) and appendix list of individual metabolites measured.**

### 3.2.2.1 Internal standards

The kit comes with a mixture of internal standards for quantification of metabolites as follows: for FIA method, chemical homologous internal standards for acylcarnitines, glycerophospholipids, sphingolipids and hexose were already implemented on the plate filter. For the FIA method, internal standards consisted of 9 isotopes labelled acylcarnitines, 1 isotope labelled hexose, 1 non-labelled lyso-PC, 2 non-labelled PCs and 1 non-labelled SM. Whereas 20 stable isotope-labelled internal standards were used in the LC method. These standards were prepared by dissolving lyophilised internal standard vial with 1200 µL of HPLC grade water and 15 minutes at 1000 g before using. Same volume (10 µL) of internals standards were added to each well on the plate except the first well which was treated as a blank.

### 3.2.2.2 Calibration standards and Quality control samples

For the LC method, the seven-point serial dilution calibrators each containing 38 different standard chemicals (all amino acids and biogenic amines) were used to generate the calibration curve. Seven lyophilised calibration standard vials were prepared by adding 100 µL of HPLC water to each vial and shaking for 15 minutes at 1220 rpm before using. Whereas in the FIA

method 1-point calibration was used containing 17 standard chemicals (12 acylcarnitines, 4 lipid standards and 1 hexose). These standards were already implemented on the plate filter.

Biocrates quality control (QC) samples were comprised of human plasma samples spiked with metabolites at defined concentrations. There were three different Biocrates QC samples on each plate at low, medium and high concentrations of the spiked metabolites. These QCs were used to verify the accuracy of the method and the mass spectrometer per plate. Three lyophilised QC vials were reconstituted in 100  $\mu$ L water and shaken for 15 minutes at 1000 x g. It was centrifuged at 4°C for 5 minutes at 2750 x g before loading on the plate.

### **3.2.2.3 Zero samples (Phosphate saline buffer) to calculate limit of detection (LOD), derivatisation reagent and extraction solvent**

The limit of detection (LOD) of all metabolites was calculated per plate by using the median value of three phosphate buffer saline (PBS) solutions referred to as zero samples. The median signal of all three samples was calculated as an approximation of background noise, 3 times this median value was listed as LOD for each metabolite by Met*IDQ* software. Internal standard intensities of PBS samples and other samples on the plate were compared for automated calculation of ion suppression effect. Biocrates recommended the use of PBS buffer to calculate LOD when analysing serum /plasma samples because its salt content is similar to serum/plasma resulting in comparable ion suppression for most metabolites. Derivatisation reagent 5% v/v Phenylisothiocyanate (PTIC) was prepared by adding 150  $\mu$ L of 100% PTIC to a mixture of 950  $\mu$ L ethanol, 950  $\mu$ L water and 950  $\mu$ L pyridine. The extraction solvent was prepared by dissolving 19 mg ammonium acetate in 50 mL methanol. Due to the polar nature of the amino acids derivatisation using PTIC was carried out to make them hydrophobic. PTIC reacts with the N-terminal amino group of amino acids to form phenylcarbamoyl derivatives (nonpolar moieties) which are easily separated on the RP column and increase their sensitivity.

### **3.2.2.4 Sample preparation workflow**

The overall sample preparation workflow using Biocrates Absolute*IDQ* p180 assay was the same for both LC and FIA mode and is shown in Figure 3.4. Biocrates p180 kit plate comes with FIA method internal standards already placed on the plate filter. An additional 10  $\mu$ L of internal standards for the LC method were added to all wells except the first well which is used as blank to calculate background level and check the system for any contaminants. Serum samples were left to thaw on ice and then vortexed and centrifuged at 10,000 x g. 10  $\mu$ L of supernatant from each serum sample was carefully loaded in the middle of the filter and was allowed to dry under a stream of nitrogen for 30 minutes. 50  $\mu$ L of Phenylisothiocyanate was added for derivatisation and after a 20-minute incubation, the filter spots on the plate were dried under nitrogen for 45-60 minutes to evaporate excess liquid. 300  $\mu$ L of extraction solvent was added to each well and

plate was shaken on an Eppendorf plate shaker for 30 minutes at 450 rpm at room temperature. The plate was then centrifuged for 2 minutes at 100 x g to filter the samples. After filtering, 50  $\mu$ L of the extracted samples were transferred to a fresh 96-well plate and diluted with 450  $\mu$ L of organic solvent (60% MeOH and 40% ammonium acetate in water) and covered with silicon mats before the LC-MS/MS run. For FIA-MS/MS, 20  $\mu$ L of extracted samples were transferred to a fresh 96-well plate and diluted with 380  $\mu$ L of FIA running solvent (solvent provided by Biocrates). The LC-MS/MS method was run first as most Biogenic amines are not stable for a longer period after adding water (stability 48 hours at 4°C), whereas FIA plate samples were stable for a week at 4°C as per Biocrates instructions.

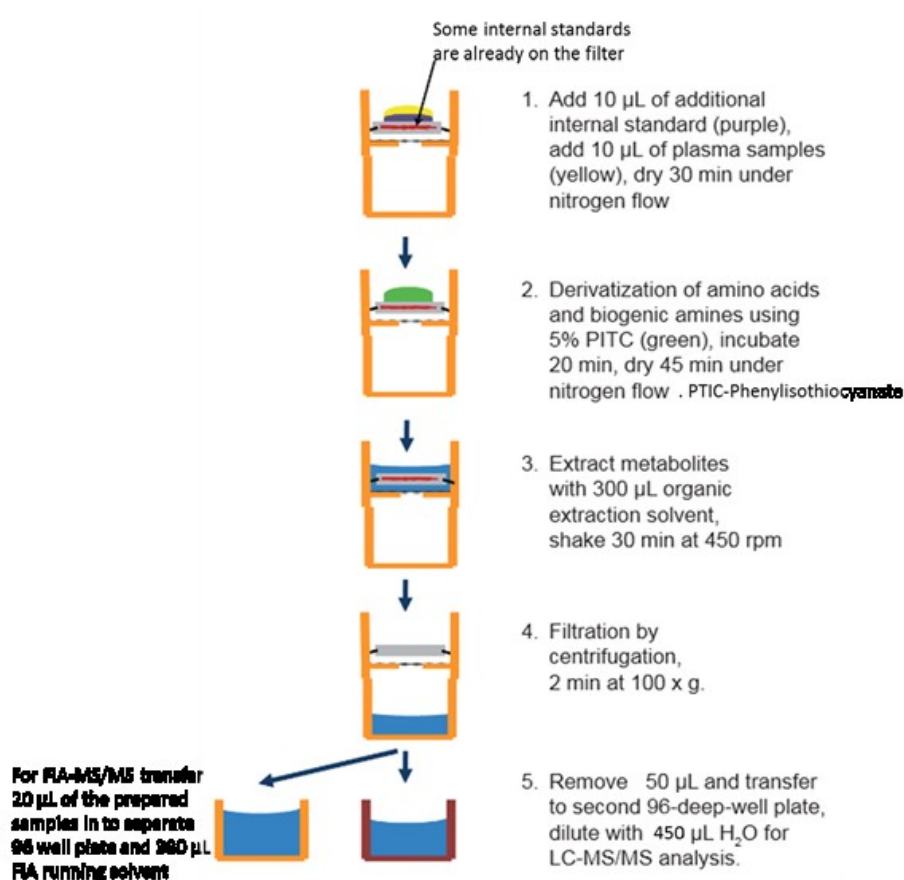


**Figure 3.4 Biocrates Absolute/DQ p180 sample preparation workflow [Figure Ref Biocrates website]**

The number of samples analysed per plate were: 1 blank (3 injections/sample), 3 phosphate saline buffer (PBS) to calculate the limit of detection (LOD), for the UPLC-MS/MS method seven different concentration calibration standard curves (7 samples)/ for FIA no calibration curve, 3 different concentration (low, medium, high) quality control samples and 82 experimental samples including experimental QCs. Total time taken from start to end including sample registration to data processing per plate, for both LC and FIA was approximately 28 hours.

### 3.2.2.5 Ultra performance liquid chromatography tandem mass spectrometry (UPLC-MS/MS) method

Amino acids and Biogenic amines were analysed by the LC-MS/MS method. Liquid Chromatography was carried using Waters ACQUITY UPLC BEH C18 1.7 $\mu$ M 2.1 x 50mm column connected to Waters ACQUITY UPLC® system. The column temperature was maintained at 50°C and injection volume was 5  $\mu$ L. The mobile phase A consisted of water + 0.2% formic acid and mobile phase B as acetonitrile + 0.2% formic acid at a flow rate of 900  $\mu$ L/minute. A gradient elution was performed for the analysis, changing the A/B mobile phase ratios from 100:0 to 40:60 in 3.75 minutes (Table 3.2). Total LC-MS/MS analysis time per sample was 4.15 minutes/sample, per plate was approximately 7 hours. Samples were stored at 10°C in the auto-sampler throughout the analysis.

| Step | Time (minutes) | Flow rate ( $\mu$ L/min) | A (%)<br>Water +0.2% Formic acid | B (%)<br>Acetonitrile +0.2% Formic acid |
|------|----------------|--------------------------|----------------------------------|-----------------------------------------|
| 0    | 0.00           | 900                      | 100                              | 0                                       |
| 1    | 0.25           | 900                      | 40                               | 60                                      |
| 2    | 3.75           | 900                      | 100                              | 0                                       |
| 3    | 4.05           | 900                      | 100                              | 0                                       |
| 4    | 4.15           | 900                      | 100                              | 0                                       |

**Table 3.2 Gradient elution with Mobile phase A (Water + 0.2% formic acid) and Mobile phase B (Acetonitrile + 0.2% formic acid), total run time 4.15 minutes per sample.**

The mass spectrometer parameters were set to polarity positive mode, capillary voltage 3.20kV, cone voltage 27, ESI source temperature 150°C, desolvation gas temperature 600 °C, cone gas 250, desolvation gas 1000 and collision energy 2kv.

### 3.2.2.6 Flow injection analysis tandem mass spectrometry (FIA-MS/MS) method

For FIA-MS/MS method no chromatographic column was used and the autosampler was directly connected to the MS ion source. An isocratic elution method (Table 3.3) was used for the FIA analysis, using Biocrates provided solvent mixed with methanol. The injection volume was 20  $\mu$ L and column temperature set at 25 °C. Even though no column was used during this analysis, the column chamber was set at 25 °C so that it has no effect on the system with fluctuations in the room temperature. Sample loop option was set to full loop. Total FIA-MS/MS analysis time per sample was 3 minutes/ sample, per plate approximately 5 hours. Samples were stored at 10°C in the auto-sampler throughout the analysis. Samples were stored at 10°C because as per manufacturer's instructions lipids tend to stick to the sides of the 96-well plate at a temperature below 6°C.

| Step | Time (minutes) | Flow rate (μL/min) | A (%) | B (%)<br>FIA running solvent + Methanol |
|------|----------------|--------------------|-------|-----------------------------------------|
| 0    | 0.00           | 30                 | 0     | 100                                     |
| 1    | 1.60           | 30                 | 0     | 100                                     |
| 2    | 2.40           | 200                | 0     | 100                                     |
| 3    | 2.80           | 200                | 0     | 100                                     |
| 4    | 3.00           | 30                 | 0     | 100                                     |

**Table 3.3 FIA-MS/MS isocratic elution method**

The mass spectrometer parameters were set to capillary voltage 3.30kV, cone voltage 22, ESI source temperature 150°C, desolvation gas temperature 350 °C, desolvation gas flow rate 650 litres/hour. Acylcarnitines, glycerophospholipids and sphingolipids were measured in positive mode. The isocratic (homogenous) mobile phase started at a flow rate of 30 μL/min up until 1.60 minutes, from 1.60 to 2.40 minutes' flow rate gradually increased to 200 μL/min and then decreased to 30 μL/min in last 60 seconds. The total run time per sample was 3 minutes.

### 3.2.2.7 Data processing

The MetIDQ™ software package (part of Biocrates AbsoluteIDQ p180 kit) was used for managing the entire kit workflow and for automated calculation of metabolite concentrations expressed as μM of serum. Identification and quantification of the metabolites were achieved by using internal standards and multiple reaction monitoring (MRM) detection. For the LC method data were first processed in Waters MassLynx™ software application TargetLynx for quality assessment of the blank signal, limit of detection or noise signal from PBS samples, evaluating calibration curves and manually checking individual metabolite peak integration. These results were imported in MetIDQ™ for further processing and calculating metabolite concentrations. The FIA raw data files from Mass Lynx were directly imported to MetIDQ™ which used automated algorithms to read and convert the data files into metabolite concentrations.

Plate validation (data quality assessment) was carried out in the MetVal application which performed an automated quality assessment per plate to specifically check the measured values of blanks, standards and QC samples were within the set pre-defined ranges. LODs were also automatically calculated per plate for LC-MS/MS and FIA-MS/MS by using noise signal from PBS samples (3 times the median value of all 3 PBS samples was calculated as LODs). LODs for each metabolite were reported and metabolites with measured concentrations below their LOD values were highlighted. Measured metabolite concentration data were exported as excel format for further statistical analysis. In addition to the quantification of 180 individual metabolites, 28 ratios or sums (Appendix Table 3) of selected metabolites with the significance to provide detailed insight into a wide range of disorders of the metabolic diseases were also reported. All the metabolite peak integrations measured via LC-MS/MS method were double checked manually.

### **3.2.3 Study population: the RHEA mother-child cohort in Crete, Greece**

Samples for this study were drawn from the 'RHEA mother-child cohort' study based in the prefecture of Heraklion, Crete, Greece. It prospectively examined a population-based study of pregnant females and their children. The detailed study of RHEA cohort is explained elsewhere [42]. In brief, female residents (Greek and immigrants) who became pregnant over the course of one year, starting in February 2007, were asked to participate in the study at four maternity clinics (two private and two public), the same protocol and joint training were followed in all four locations. First time point of contact: face to face interview, urine and fasting blood collection was made during the 1<sup>st</sup> trimester of pregnancy (around 12 weeks of gestation mean  $\pm$  SD 11.96  $\pm$  1.49 weeks), followed by contacts at 6th month of pregnancy, at the time of birth, postnatal for child's health follow-up at 6 months, 18 months and 4 years of age. N=1610 females agreed to participate in the study, N=1317 were followed up until delivery and N= 1135 females provided 1<sup>st</sup>-trimester fasting blood samples. A detailed questionnaire including information on their lifestyle (education, smoking, alcohol consumption, socioeconomic status, physical activity etc.), age, diet during pregnancy along with medical records and other health characteristics such as reproductive history, maternal anthropometric examination and environmental exposure were recorded by trained interviewers at the time of enrolment. Female participants who provided incomplete questionnaire information or had pre-eclampsia pregnancies, multiple pregnancies, spontaneous or induced abortions were excluded from the present study.

At the time of the birth, the neonates' umbilical cord blood was collected in 10 mL BD gel separator vacutainers. The samples were centrifuged within 2 hours at 2000 x g for 10 minutes after centrifugation serum was separated into 0.5mL aliquots and stored at -80°C until assayed. The information on birth outcomes such as birth weight, preterm births, small or large for gestational age, length and head circumference were recorded for clinical records and anogenital distances were measured in neonates [36, 42]. All methods and procedures used in this study were approved by the ethical committee of the University Hospital in Heraklion, Crete, Greece. All the participants in the study provided written informed consent concerning themselves and their children after complete description of the study.

#### **3.2.3.1 Organochlorine POPs exposure assessment in 1st-trimester maternal serum**

In the RHEA cohort study, persistent organic pollutants including six individual polychlorinatedbiphenyl (PCB) congeners (IUPAC name 118, 156, 138, 153, 170 and 180), hexachlorobenzene (HCB), dichlorodiphenyltrichloroethane (DDT) and dichlorodiphenyldichloroethylene (DDE) were measured in fasting maternal serum (N=1117) collected during 1<sup>st</sup> trimester in females with only singleton pregnancies [43]. The concentrations

of these POPs were measured using an Agilent 7000B gas chromatography triple quad mass spectrometer (GC-MS/MS) at the National Institute of Health and Welfare, Chemical exposure unit, Kuopio, Finland [43]. Sample preparation and other details of the method are explained elsewhere [44]. All the reported concentrations were on wet weight expressed in pg/mL serum. The limit of quantification (LOQ) determined for PCB118 was 6pg/mL and 10 pg/mL for PCB 152, PCB138, PCB 170, PCB 180, HCB and DDE; 100% of samples were above their LOQ values. For PCB 156 and DDT only 40% of the samples were above the value of LOQ [43] and due to the high percentage of the sample below the LOQ, these POPs were excluded from any further analysis.

### **3.2.3.2 RHEA cohort sample sets analysed**

The study design and the selection criteria for the number of samples included in the final statistical analysis for the present study are shown in Figure 3.5. In the present study, metabolic profiling of N= 148 1<sup>st</sup> trimester fasting maternal serum from females in the RHEA cohort was analysed using Biocrates Absolute*IDQ* p180 assay across two kit plates. Out of these, females N=47 with preterm births, small or large gestational age birth outcomes, and ones with metabolic syndrome were excluded from the data analysis. The rationale behind excluding these samples was to observe the effect of POPs on healthy pregnancies. A further N=34 samples were excluded due to unavailability of organochlorine POPs data and only the remaining N=67 were used in the statistical analysis.

Similarly, metabolic profiling of N=114 umbilical cord serum samples from RHEA cohort was analysed using Biocrates Absolute*IDQ* p180 assay across four kit plates. Out of these, N=28 samples where organochlorine POPs data were not available were excluded from any subsequent analysis. The samples without POPs data were initially analysed because at the time of the analysis, it was unclear for which samples POPs data would become available, as these analyses were conducted at a later stage. The initial expectation that all available samples for metabolic profiling would have POPs data was not possible due to resource constraints.

In this sample set, 1<sup>st</sup>-trimester maternal serum and cord serum samples (mother and cord) were not paired and were conducted as independent analyses. The selection of cord blood was based on those children for whom both matched POPs exposure data and DNA samples were expected to be available, whereas selection and analysis of mothers' samples for metabolic profiling was made before this information was available.

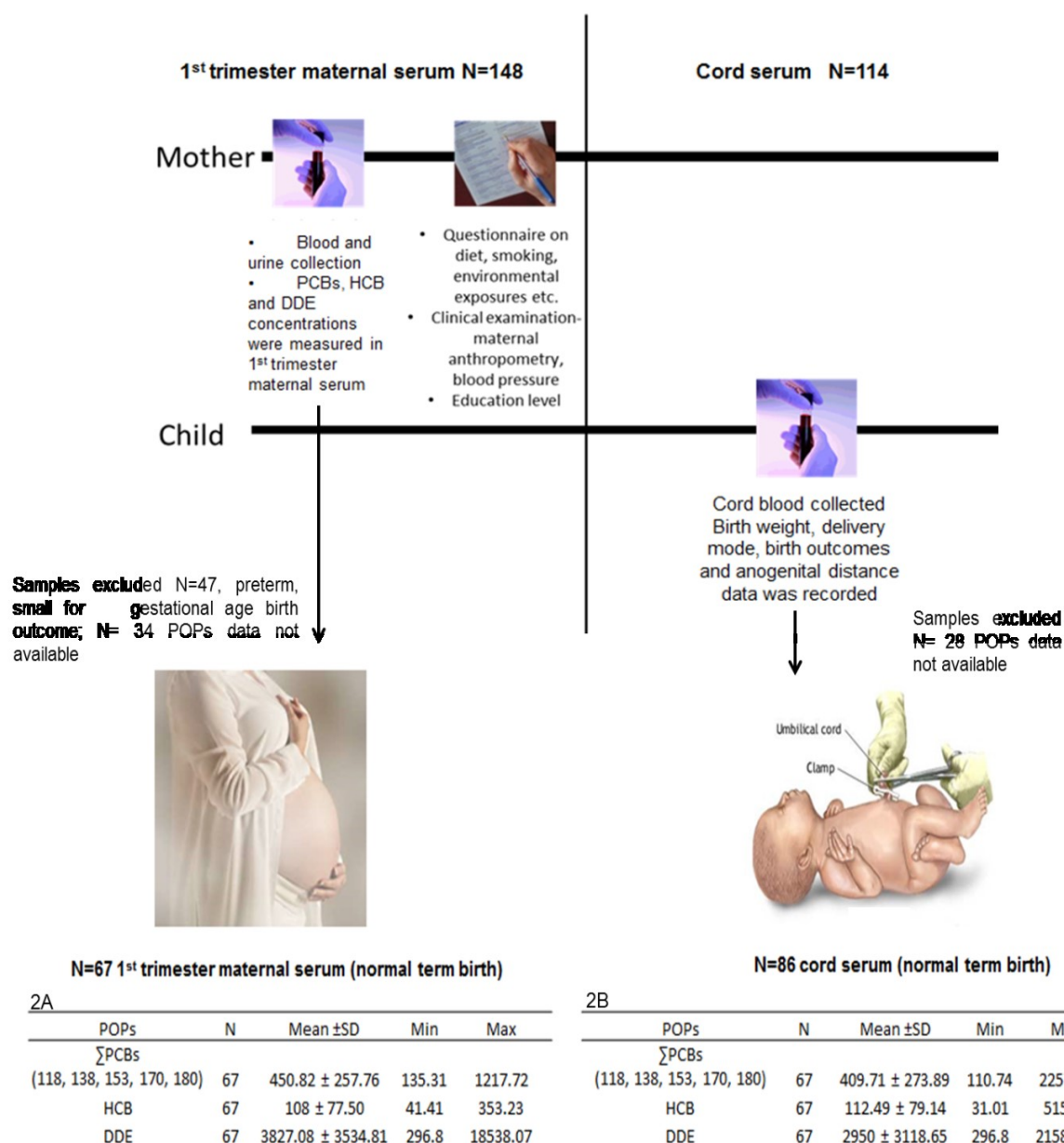


**Figure 3.5 Study design of the 1st-trimester maternal serum and umbilical cord serum samples analysed in this PhD from RHEA mother-child cohort using Biocrates AbsoluteIDQ P180 (LC-MS/MS and FIA-MS/MS) assay. 2A shows concentration (pg/mL) of 1<sup>st</sup>-trimester maternal serum POPs in N=67 samples where metabolite profile data was measured, and 2B shows the concentration of 1<sup>st</sup>-trimester maternal serum POPs and their paired N=86 umbilical cord serum samples or which metabolite profile data was analysed. 1<sup>st</sup>-trimester maternal serum and umbilical cord serum samples analysed in this study were not paired.**

The objective of this study was to examine the potential associations between 1<sup>st</sup> trimester maternal concentrations of POPs: HCB, DDE, and the total PCBs ( $\Sigma$ PCBs) which was calculated by summing the concentration of five individual PCB congeners (118, 153, 138, 170 and 180) and the metabolic profile data in N=67 1<sup>st</sup> trimester maternal serum and N=86 umbilical cord serum. To account for potential biases associated with automatic lipid adjustment (serum lipids) [45] PCB concentrations were lipid-adjusted. Lipid-adjusted concentrations were calculated by

dividing the serum concentration of PCBs, HCB and DDE with the serum concentration of lipids (triglycerides and cholesterol); the resulting units were nanograms of PCB per gram of lipid (ng/g lipid) in serum. The unadjusted and lipid-adjusted PCBs were strongly correlated with each other and the same results were observed in statistical analysis with metabolic profile data in 1<sup>st</sup>-trimester maternal serum and cord serum. However, cholesterol and triglyceride data for N=25 samples were unavailable (since cholesterol and triglycerides were not measured in these samples) all subsequent analysis were performed using unadjusted PCBs concentrations. The concentrations (wet weight expressed in pg/mL serum) of 1<sup>st</sup>-trimester maternal serum POPs ( $\Sigma$ PCBs, HCB and DDE) are shown in Figure 2 for both sets of the population. The highest concentration was found for DDE followed by  $\Sigma$ PCBs and HCB.

### 3.2.4 Quality control assessment criteria for the measured metabolites

Biocrates p180 kit used with the Water's triple quadrupole StepWave™ (TQ-S) MS platform is a highly optimised method which has been demonstrated to be fit-for-purpose for metabolomics studies [46, 47, Ref 21 from Chapter 1]. This method has also been proven to be in accordance with the FDA-guidelines for Bioanalytical method validation, which shows proof of reproducibility within given error range. However, quality control assessment was performed in order to assess its robustness within our laboratory. RHEA quality controls (RQC) samples were provided along with RHEA cohort samples and were treated as experimental QC's throughout the analysis. As a measure of quality assurance and to determine the precision of the quantified metabolites, RQC samples were repeatedly introduced into the analysis after every 28 samples. This resulted in 15 experimental quality control samples across 5 plates (3 RQCs on each plate). RQCs were used to calculate intra and inter-day precision for each metabolite measured using LC and FIA method. The inter-batch precision was calculated as a percentage coefficient of variance (CV%) for each metabolite across five classes of metabolites. As an additional measure of quality assurance of the analytical method and to compare the reproducibility of the measured metabolites over the long term, a commercially bought serum was also used as a quality control sample (CQC) which was repeatedly analysed 4 times per plate (Total=20, 4 CQCs every plate).

A quality control assessment criterion was set according to the precision acceptance criterion of intra and inter-batch CV%  $\leq 20\%$  in conformance to FDA & EMA guidelines for Bioanalytical method validation [Ref 65, 66 from Chapter 2]. Furthermore, metabolites with measured concentrations below the limit of detection (LOD) in more than 30% of the samples were excluded from any subsequent analysis. The CV% of RQCs samples were calculated and only metabolites with RQC CV%  $\leq 20\%$  were deemed acceptable whereas metabolites with RQC CV%  $> 20\%$  were excluded.

### 3.2.5 Statistical analysis

All statistical analyses were performed using SPSS version 20 and R version 3.2.1. The concentrations of POPs were first converted to  $\mu\text{M}$  units to match with metabolite concentration units. A four-step analysis including Spearman correlation to screen/shortlist metabolites of interest followed by multiple linear regression models controlling for major covariates was conducted to identify metabolites associated in 1<sup>st</sup> trimester maternal and umbilical cord serum with 1<sup>st</sup>-trimester prenatal organochlorine POPs exposure ( $\Sigma\text{PCBs}$ , HCB, and DDE). The threshold statistical significance was set at  $p \leq 0.05$ . In the first step, to identify metabolites associated with POPs exposure, a non-parametric test (Spearman correlation, due to non-normal distribution of metabolites) was used to calculate associations between concentrations of 145 metabolites and 21 intra-class metabolite ratios (number of metabolites that passed quality assessment criteria) measured in 1<sup>st</sup> trimester maternal and umbilical cord serum separately; metabolites with significant association were shortlisted as 'metabolites of interest'. In step two, potential covariates for  $\Sigma\text{PCBs}$ , HCB and DDE exposure were identified by separate spearman correlation for 100 continuous variables (such as mother's age, pre-pregnancy BMI, diet etc.) and ANOVA analysis for 10 categorical variables (parity, maternal education, smoking status, delivery type etc.) within this dataset. Step three was to  $\log_e$  transform POP exposure and previously shortlisted metabolites of interest due to the non-normal distribution of the data. In a fourth step, multiple linear regression models were used to further investigate whether POP concentrations were associated with metabolite identified as of interest in step one in 1<sup>st</sup> trimester maternal and umbilical cord serum. The models contained the log-transformed metabolite concentrations as the dependent variable and individual log-transformed POPs ( $\Sigma\text{PCBs}$ , HCB, and DDE) as the independent/explanatory variable with and without adjusting for maternal age, pre-pregnancy BMI and maternal education as covariates. No false discovery rate (FDR) correction was applied on the selected metabolites due to the small number of samples. However, the robustness of the significant ( $p \leq 0.05$ ) adjusted linear regression models were validated by resampling 1000 bootstrap models and their p values were calculated [48].

## 3.3 Results

### 3.3.1 Inter-batch precision of the reliably measured metabolites

Overall 145 metabolites out of 180 measured metabolites passed the quality control assessment criteria and were deemed acceptable to be used for the subsequent analysis including: 21 amino acids, 6 biogenic amines, 1 free carnitine and 23 acylcarnitines (Cx:y), 13 lyso-glycerophosphatidylcholines (lysoPCs), 67 glycerophosphatidylcholines (30 PCs, aa = diacyl, 34 ae = acyl-alkyl) and 14 sphingolipids (SMx:y). The lipid side chain composition is abbreviated as Cx:y, where x denotes the number of carbons in the side chain and y denotes the number of double bonds. The exact position of the double bonds and carbon atoms distribution in different fatty acid chains cannot be determined in this method.

RHEA QCs (QC samples provided along with RHEA cohort samples) were used to calculate the intra- and inter-batch precision of the metabolites which were reliably and reproducibly measured. Table 3.4 shows that inter-batch precision (across five plates) calculated as a percentage coefficient of variance (CV%) was  $\leq 20\%$  of each metabolite in five metabolite classes. Average CV% represents the average of all metabolites in that specific metabolite class along with minimum and maximum CV% reported. Intra-batch CV% (result not shown) of RHEA QC samples (within each plate of 82 samples) was also less than 20 % for 131 metabolites. Similarly, 21 metabolite sums and ratios out of 28 routinely evaluated using the Biocrates platform passed the quality control assessment criteria of intra and inter-batch precision.

| Compound class                                         | LC-MS/MS                 |                              |                             | FIA-MS/MS                         |                            |
|--------------------------------------------------------|--------------------------|------------------------------|-----------------------------|-----------------------------------|----------------------------|
|                                                        | Amino acids <sup>a</sup> | Biogenic amines <sup>b</sup> | Acylcarnitines <sup>c</sup> | Glycerophospholipids <sup>d</sup> | Sphingolipids <sup>e</sup> |
| Average CV % across metabolites in each compound class | 5.96                     | 6.46                         | 7.42                        | 6.72                              | 7.95                       |
| Minimum CV %                                           | 4.63                     | 5.5                          | 3.4                         | 3.79                              | 6.99                       |
| Maximum CV %                                           | 8.76                     | 8.76                         | 11.73                       | 17.69                             | 10.63                      |

**Table 3.4 Inter-batch precision of RHEA QCs average CV% across five Biocrates P180 plates (excluding metabolites below LOD in more than 30% of samples and metabolites with inter-day CV% >20%).**

a=all 21 measured amino acids passed the strict quality control assessment criteria; b= 6 out 20 biogenic amines passed QC acceptance criteria and the rest were excluded from subsequent analysis; c=24 out of 40 acylcarnitines metabolites passed QC criteria; d=7 out of 87 glycerophospholipids were excluded and e= all 14 sphingolipids passed the QC acceptance criteria.

A cluster map analysis was performed (Figure 3.6) which showed the arrangement of clusters formed by two separate quality control serum samples RHEA QCs and commercial serum QCs

(analysed across five plates) compared to all the other samples (in total N=350). QC samples forming clusters were RHEA QCs N= 15 and commercial serum QC (CQC) N=20. These clusters show the similarity of the serum sets within each other, as well as dissimilarity with other serum, sets according to their metabolite profiles. This illustrated that the different types of samples could be separated according to their metabolome and provides confidence in the analytical methodology. This cluster analysis also shows the good quality of the data as the QCs were tightly clustered among other samples.

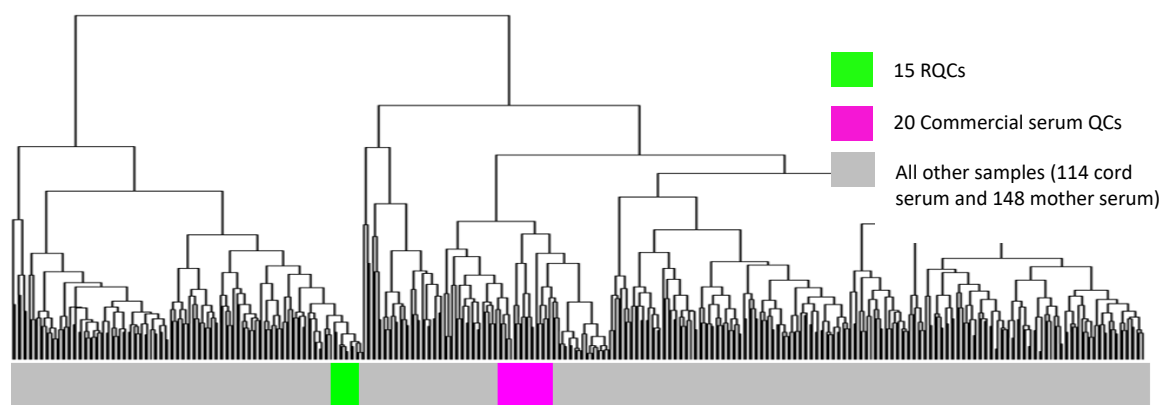


**Figure 3.6** Cluster analyses of QC serum sets according to their metabolite profiles (RHEA QCs and commercial serum QCs). A green cluster represents 15 RHEA QCs and pink cluster represents 20 commercial serum QCs while other samples including maternal and cord serum are shown in grey.

### 3.3.2 Metabolite profile comparison between maternal serum and cord serum

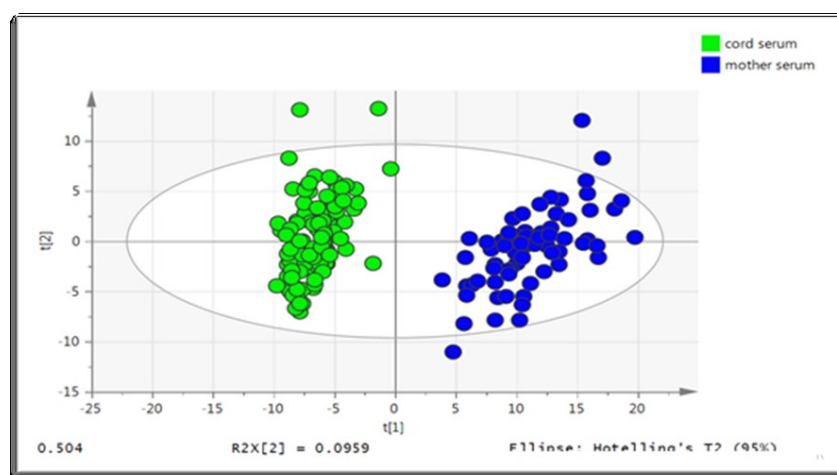


**Figure 3.7(A)** Principal components analysis (PCA) scores plot discriminating the metabolic profiles in N=67 1st-trimester maternal serum (blue cluster) and N=86 umbilical cord serum samples (green cluster). Each green dot represents one healthy cord sample, while each blue dot represents one healthy 1st-trimester maternal sample.



A total of 145 metabolites and 21 metabolites sums and ratios were reproducibility measured in serum obtained from the 1<sup>st</sup> trimester maternal and umbilical samples. Using these metabolites, an unsupervised principal component analysis (PCA) was first performed as shown in Figure 3.7(A). A clear distinction between metabolite profiles of N=67, 1<sup>st</sup>-trimester maternal serum and N=86 umbilical cord serum was observed with all the cord samples in the left half and all the maternal samples in the right half of the plot. The serum samples were mapped in the space by first two principal components PC1 versus PC2, which accounted for 54% of the total variation in the sample sets, the second component which accounted for 12% of the total variance was related to inter-individual differences. The unsupervised PCA model revealed the general metabolic information and differences between 1<sup>st</sup> trimester maternal and umbilical cord serum. The metabolites responsible for these differences were assessed from the loading plot from PCA (Figure 3.7B) and by bar graphs (In Appendix Figure 1A-1G) comparing mean concentrations of both maternal and cord serum samples. Compared with cord serum, maternal serum had significantly higher levels of lipids and lower levels of essential amino acids.

were marginally higher in maternal serum. Comparison of mean concentrations of 21 individual amino acids between maternal and cord serum are shown in Appendix Figure 1A. The results show that compared to maternal serum individual amino acids: glutamine, glycine, lysine, methionine, ornithine, threonine, tryptophan and tyrosine were significantly (T-test,  $P \leq 0.0005$ ) higher in cord serum. Similarly, mean concentration levels of individual biogenic amines metabolites: Kynurenine, total dimethylarginine (total DMA) and symmetric dimethylarginine (SDMA) were significantly ( $P \leq 0.0005$ ) higher in cord serum compared to maternal serum Appendix Figure 1B.

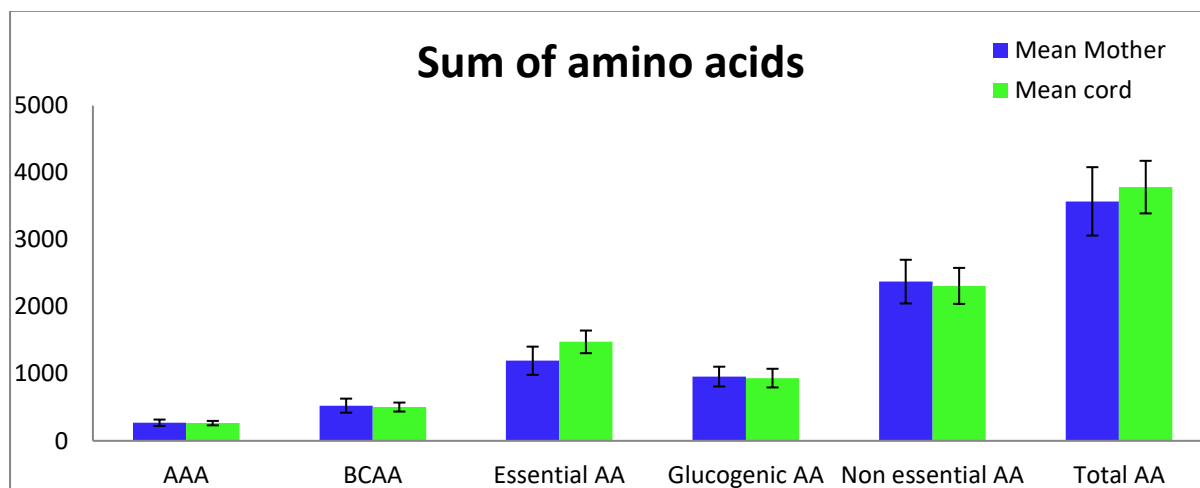


**Figure 3.4 Bar graph comparing mean concentrations of the sum of amino acids (AA) in 1st trimester maternal and cord serum amino acids. The x-axis represents the sum of amino acids (AAA= aromatic amino acids and BCAA= Branched chain amino acids) and Y-axis represents their mean concentration in (μM). Error bar indicates mean ± standard deviation.**

Overall the concentrations of lipids including: lyso-glycerophosphocholines (lysoPC) except for Lyso PCa C16:1, Lyso PCa C20:3 and Lyso PCa C20:4 Appendix Figure 1C; diacyl-phosphatidylcholine (PC aa) Appendix Figure 1D; phosphatidylcholine plasmalogens (PC ae) Appendix Figure 1E; and sphingomyelins (SM) Appendix Figure 1F were higher in 1<sup>st</sup> trimester maternal serum compared to cord serum. Similarly, the acylcarnitines average concentrations were higher in maternal serum compared to cord serum (except for short chain carnitines C2 C3, C4 and C5; long chain C16, C16-OH and C16:1), Appendix Figure 1G.

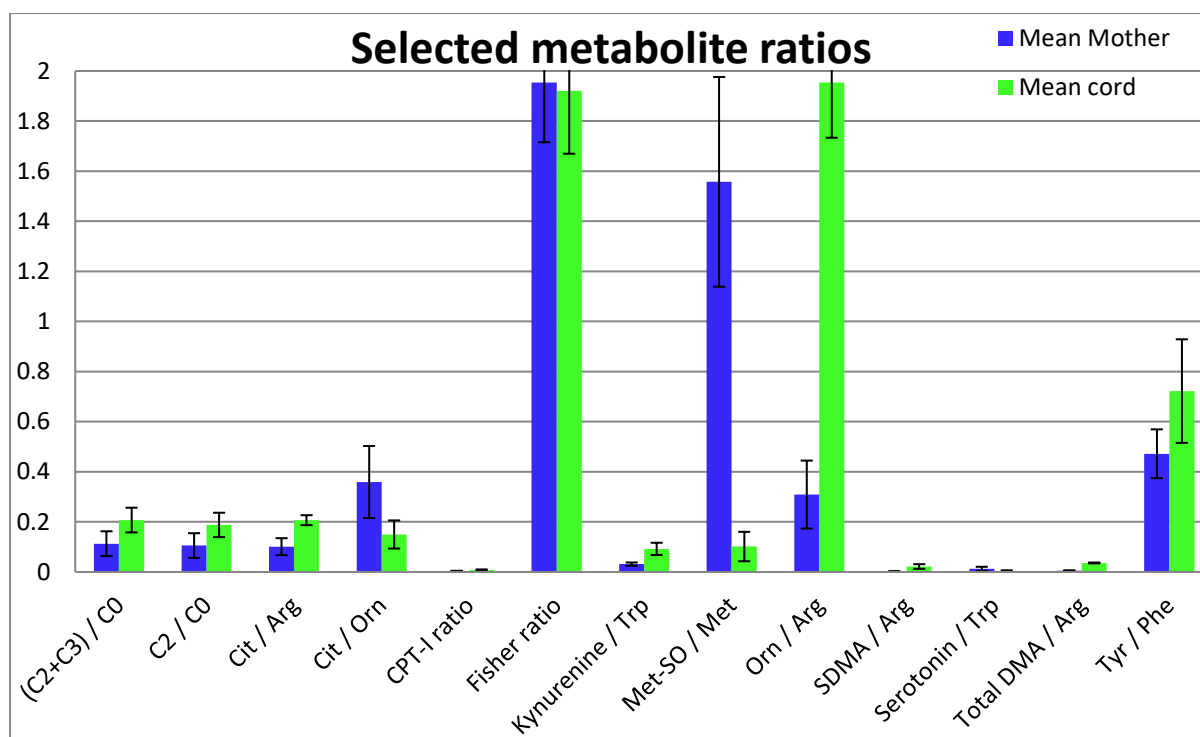


**Figure 3.5 Bar graph comparing mean concentrations of selected metabolite ratios in the 1st trimester maternal and cord serum. The x-axis represents selected intra-class metabolite ratios and Y-axis represents their concentration in (μM). Error bar indicates mean ± standard deviation.**

Intra-metabolite class ratios were also calculated in maternal and cord serum and their comparison is shown in Figure 3.9. These intra-metabolite classes provide detailed insight of the enzymatic reactions and better understanding into some of the metabolic pathways associated with wide a range of disorders of the metabolic diseases. Ratios observed to be higher in cord serum were: C2/C0, ratio of acetylcarnitine to free carnitine; (C2/C3)/C0 ratio of short chain to free carnitine representing measure of  $\beta$ -oxidation of even-numbered fatty acids and measure of overall  $\beta$ -oxidation activity respectively; ratio of citrulline/arginine (Cit/Arg) indicates activity of enzyme nitric oxide synthase (NOS); ratio of kynurenine/tryptophan (kyr/Trp) indicates the rate of Trp degradation to niacin, an indicator of indole dioxygenase activity (immunosuppression/tolerance); the ratio of ornithine to arginine(Orn/Arg) indicates activity of arginase; ratio of total dimethylated arginine to total unmodified arginine (Total DMA/Arg) shows activity of protein arginine methyltransferases (PRMT) and ratio of Ratio of tyrosine to phenylalanine (Tyr/Phe) indicates enzymatic activity of phenylalanine hydroxylase (PAH). Whereas ratios higher in maternal serum were: Ratio of citrulline to ornithine (Cit/Orn) which shows the activity of enzyme ornithine carbamoyl transferase (OCT); Ratio of sulfoxidized methionine to total unmodified methionine (Met-SO/Met) is used as a measure of systemic oxidative stress and Fisher ratio (Ratio of BCAA to aromatic AA) is used to indicate liver damage.

### 3.3.3 Associations between 1st-trimester organochlorine POPs exposure and metabolite profile of 1st-trimester maternal serum

The aim of this analysis was to investigate potential links between current low levels of different POPs ( $\Sigma$ PCBs, HCB and DDE) found in 1<sup>st</sup>-trimester maternal serum and the metabolite profile of 1<sup>st</sup>-trimester maternal serum in N=67 women with healthy pregnancy from the RHEA pregnancy cohort. Spearman correlation coefficients were calculated to short list 'metabolites of interest' with significant associations ( $P \leq 0.05$ ), 32 metabolites and 1 metabolite ratio was identified to be linked with POPs exposures (Appendix Table 4). Out of which in total 27 metabolites and 1 metabolite ratio were associated with the sum of all PCB congeners ( $\Sigma$ PCBs), 20 metabolites and 1 metabolite ratio were associated with HCB, and 2 metabolites and 1 metabolite ratio were associated with DDE.

However, the observed significance for most of the metabolites was low and could be influenced by the covariate factors for the POPs exposures. For the metabolites identified as of interest by the above analysis, their association with POPs exposures was tested in unadjusted and adjusted multiple regression models (Table 3.5), adjusting for POPs exposure covariates identified within this dataset such as mothers' age, pre-pregnancy BMI, and maternal education. Table 3.5 shows Beta coefficients, 95% confidence intervals and their p values of unadjusted and adjusted multiple regression models for 1<sup>st</sup>-trimester maternal serum POPs exposures and metabolites of interest in 1<sup>st</sup>-trimester maternal serum. None of the associations passed (or remained significant) after considering false discovery rate calculated using the Benjamini-Hochberg method [50]. Robustness of the significant ( $P \leq 0.05$ ) adjusted models were validated by bootstrapping. Bootstrapping generates confidence intervals for a given statistic via random resampling with replacement [48, 49]. Bootstrapping here was used to estimate the confidence intervals of p values estimated from the regression. *P*-value distributions were derived by bootstrap resampling of 1000 models. The bootstrap p-value was then defined by the 95% CI by taking the 2.5 & 97.5 percentiles of the ranked list of the 1000 values. Though bootstrapping does not account for multiple testing, it is an alternative statistical inference to measure the accuracy of the statistical test when the stability of the result and its assumption is in doubt.

|                                                        | <u>Unadjusted model</u>   |                | <u>Adjusted model<sup>b</sup></u> |                | <u>Bootstrapping<sup>c</sup></u> |
|--------------------------------------------------------|---------------------------|----------------|-----------------------------------|----------------|----------------------------------|
|                                                        | $\beta$ (95% CI)          | <i>P</i> value | $\beta$ (95% CI)                  | <i>P</i> value | <i>P</i> value (1000 models)     |
| <b><u>Log<sub>10</sub> <math>\Sigma</math>PCBs</u></b> |                           |                |                                   |                |                                  |
| Log <sub>10</sub> Citrulline                           | -.22 (-1.13, .03)         | <b>0.06</b>    | -.29 (-1.137, -.268)              | <b>0.002</b>   | <b>0.01</b>                      |
| Log <sub>10</sub> Kynurenine                           | -.35 (-1.38, -.29)        | <b>0.003</b>   | -.220 (-.967, -.064)              | <b>0.02</b>    | <b>0.05</b>                      |
| Log <sub>10</sub> lysoPC a C16:0                       | .18 (-.131, .890)         | 0.14           | .155 (-.083, .735)                | 0.11           | -                                |
| Log <sub>10</sub> PC aa C28:1                          | .22 (-.051, 1.123)        | 0.07           | .121 (-.179, .765)                | 0.21           | -                                |
| Log <sub>10</sub> PC aa C30:0                          | .20 (-.072, .753)         | 0.10           | .100 (-.161, .501)                | 0.30           | -                                |
| Log <sub>10</sub> PC aa C32:1                          | .23 (-.004, .583)         | 0.05           | .100 (-.117, .360)                | 0.31           | -                                |
| <b>Log<sub>10</sub> PC aa C34:3</b>                    | <b>.30 (.132, 1.070)</b>  | <b>0.013</b>   | <b>.204 (.030, .778)</b>          | <b>0.03</b>    | <b>0.08</b>                      |
| Log <sub>10</sub> PC aa C34:4                          | .32 (.151, .906)          | 0.007          | .181 (-.022, .606)                | 0.06           | -                                |
| Log <sub>10</sub> PC aa C36:3                          | .24 (-.003, 2.383)        | 0.05           | .143 (-.246, 1.662)               | 0.14           | -                                |
| Log <sub>10</sub> PC aa C36:4                          | .22 (-.069, 1.939)        | 0.06           | .122 (-.292, 1.311)               | 0.20           | -                                |
| Log <sub>10</sub> PC aa C36:5                          | .33 (.198, 1.103)         | 0.006          | .123 (-.185, .663)                | 0.26           | -                                |
| Log <sub>10</sub> PC aa C36:6                          | .30 (.147, 1.089)         | 0.01           | .104 (-.212, .628)                | 0.32           | -                                |
| Log <sub>10</sub> PC aa C38:3                          | .28 (.093, 1.142)         | 0.02           | .110 (-.206, .691)                | 0.28           | -                                |
| Log <sub>10</sub> PC aa C38:4                          | .25 (.027, 1.014)         | 0.03           | .077 (-.252, .571)                | 0.44           | -                                |
| Log <sub>10</sub> PC aa C38:5                          | .30 (.115, .855)          | 0.01           | .103 (-.154, .479)                | 0.30           | -                                |
| Log <sub>10</sub> PC aa C38:6                          | .26 (.031, .646)          | 0.03           | .080 (-.156, .362)                | 0.42           | -                                |
| Log <sub>10</sub> PC aa C40:4                          | .32 (.223, 1.369)         | 0.007          | .139 (-.153, .832)                | 0.17           | -                                |
| Log <sub>10</sub> PC aa C40:5                          | .28 (.071, .816)          | 0.02           | .108 (-.142, .479)                | 0.28           | -                                |
| Log <sub>10</sub> PC ae C34:1                          | .21 (-.072, 1.465)        | 0.07           | .016 (-.602, .703)                | 0.87           | -                                |
| Log <sub>10</sub> PC ae C36:1                          | .18 (-.116, .894)         | 0.12           | .116 (-.155, .635)                | 0.22           | -                                |
| Log <sub>10</sub> PC ae C38:0                          | .26 (.043, 1.015)         | 0.03           | .079 (-.254, .574)                | 0.44           | -                                |
| Log <sub>10</sub> PC ae C38:3                          | .28 (.111, 1.187)         | 0.01           | .136 (-.151, .770)                | 0.18           | -                                |
| Log <sub>10</sub> PC ae C40:1                          | .22 (-.026, .921)         | 0.06           | .110 (-.177, .609)                | 0.27           | -                                |
| Log <sub>10</sub> PC ae C40:4                          | .24 (.013, 1.243)         | 0.04           | .057 (-.320, .662)                | 0.57           | -                                |
| Log <sub>10</sub> PC ae C40:5                          | .25 (.025, 1.051)         | 0.04           | .055 (-.370, .554)                | 0.59           | -                                |
| Log <sub>10</sub> PC ae C42:4                          | .23 (-.028, 1.297)        | 0.06           | .089 (-.292, .782)                | 0.36           | -                                |
| Log <sub>10</sub> SM (OH) C14:1                        | .24 (.010, 1.163)         | 0.04           | .099 (-.231, .707)                | 0.31           | -                                |
| <b>Log<sub>10</sub> Kynurenine / Trp</b>               | <b>-.38 (-1.60, -.40)</b> | <b>0.001</b>   | <b>-.29 (-1.24, -.30)</b>         | <b>0.002</b>   | <b>0.01</b>                      |
| <b><u>Log<sub>10</sub> HCB</u></b>                     |                           |                |                                   |                |                                  |
| Log <sub>10</sub> Kynurenine                           | -.33 (-1.228, -.230)      | 0.006          | -.176 (-.854, .056)               | 0.08           | -                                |
| Log <sub>10</sub> lysoPC a C16:0                       | .17 (-.133, .856)         | 0.014          | .091 (-.225, .595)                | 0.37           | -                                |
| Log <sub>10</sub> PC aa C32:1                          | .23 (-.007, .561)         | 0.056          | .091 (-.131, .346)                | 0.37           | -                                |
| Log <sub>10</sub> PC aa C32:3                          | .22 (-.029, 1.054)        | 0.063          | .135 (-.147, .753)                | 0.18           | -                                |
| <b>Log<sub>10</sub> PC aa C34:3</b>                    | <b>.29 (.11, 1.02)</b>    | <b>0.014</b>   | <b>.21 (.04, .78)</b>             | <b>0.02</b>    | <b>0.06</b>                      |
| Log <sub>10</sub> PC aa C34:4                          | .20 (-.061, .697)         | 0.098          | .067 (-.212, .422)                | 0.50           | -                                |
| Log <sub>10</sub> PC aa C36:5                          | .28 (.089, .980)          | 0.019          | -.017 (-.459, .394)               | 0.88           | -                                |
| Log <sub>10</sub> PC aa C36:6                          | .51 (.229, 1.128)         | 0.004          | .108 (-.211, .628)                | 0.32           | -                                |
| Log <sub>10</sub> PC aa C38:3                          | .19 (-.109, .929)         | 0.119          | -.03 (-.51, .38)                  | 0.77           | -                                |

|                                             |                                   |             |                                   |             |             |
|---------------------------------------------|-----------------------------------|-------------|-----------------------------------|-------------|-------------|
| Log <sub>10</sub> PC aa C40:4               | .21 (-.061, 1.084)                | 0.07        | -.001 (-.50, .49)                 | 0.99        | -           |
| Log <sub>10</sub> PC aa C42:2               | .22 (-.03, .86)                   | 0.07        | .10 (-.16, .57)                   | 0.28        | -           |
| Log <sub>10</sub> PC aa C42:5               | .21 (-.060, 1.126)                | 0.07        | .078 (-.305, .688)                | 0.44        | -           |
| Log <sub>10</sub> PC ae C34:1               | .16 (-.230, 1.272)                | 0.17        | .04 (-.502, .799)                 | 0.64        | -           |
| Log <sub>10</sub> PC ae C36:0               | .26 (.073, 1.321)                 | 0.02        | .11 (-.232, .831)                 | 0.26        | -           |
| <b>Log<sub>10</sub>PC ae C36:1</b>          | <b>.25 (.035, .996)</b>           | <b>0.03</b> | <b>.20 (.022, .793)</b>           | <b>0.03</b> | <b>0.07</b> |
| Log <sub>10</sub> PC ae C38:0               | .29 (.113, 1.043)                 | 0.01        | .08 (-.244, .582)                 | 0.41        | -           |
| Log <sub>10</sub> PC ae C38:3               | .28 (.095, 1.138)                 | 0.02        | .07 (-.301, .628)                 | 0.48        | -           |
| Log <sub>10</sub> PC ae C40:1               | .28 (.082, .985)                  | 0.02        | .11 (-.182, .603)                 | 0.28        | -           |
| Log <sub>10</sub> PC ae C40:4               | .19 (-.109, 1.094)                | 0.10        | .001 (-.515, .518)                | 0.99        | -           |
| Log <sub>10</sub> PC ae C40:5               | .24 (.002, .998)                  | 0.04        | .02 (-.387, .487)                 | 0.81        | -           |
| <b>Log<sub>10</sub>Kynurenine /<br/>Trp</b> | <b>-.30 (-1.363, -.<br/>.170)</b> | <b>0.01</b> | <b>-.19 (-.989, -.<br/>.004)</b>  | <b>0.04</b> | <b>0.06</b> |
| <b><u>Log<sub>10</sub> DDE</u></b>          |                                   |             |                                   |             |             |
| <b>Log<sub>10</sub>PC ae C34:3</b>          | <b>-.29 (-1.923, -.200)</b>       | <b>0.01</b> | <b>-.25 (-1.714, -.<br/>.110)</b> | <b>0.02</b> | <b>0.05</b> |
| Log <sub>10</sub> PC ae C36:0               | .20 (-.162, 1.998)                | 0.09        | .09 (-.633, 1.440)                | 0.44        | -           |
| <b>Log<sub>10</sub>Kynurenine /<br/>Trp</b> | <b>-.27 (-2.209, -.<br/>.158)</b> | <b>0.02</b> | <b>-.21 (-1.878, -.<br/>.036)</b> | <b>0.05</b> | <b>0.08</b> |

**Table 3.5 Unadjusted and adjusted multiple linear regression models for 1st-trimester maternal serum POPs ( $\Sigma$ PCBs, HCB and DDE) and a metabolite of interest in 1st-trimester maternal serum in N=67 women with healthy pregnancy from the RHEA pregnancy cohort.**

<sup>b</sup>Model adjusted for mothers' age, maternal education, and pre-pregnancy BMI. <sup>c</sup>Bootstrapping=Robustness of the models was tested by bootstrap using 1000 resamplings. POPs exposures and metabolites were log-transformed for normalisation.

Beta coefficients from significant ( $P \leq 0.05$ ) adjusted multiple regression models for 1<sup>st</sup>-trimester maternal serum metabolites associated with POPs exposures are shown in Figure 3.10. Adjusted multiple regression models for  $\Sigma$ PCBs showed significant negative associative with 1 amino acid: Citrulline ( $\beta = -.29$ ; 95%CI -1.13 to -.26); 1 biogenic amine: Kynurenine ( $\beta = -.22$ ; 95%CI .96 to -.02) and 1 metabolite ratio: Kynurenine / Tryptophan ( $\beta = -.29$ ; 95%CI -1.24 to -.30), while a significant positive correlation with 1 diacyl glycerophosphocholines PC aa C34:3 ( $\beta = .30$ ; 95%CI .13 to 1.07). Hexachlorobenzene (HCB) was significantly positively correlated with 1 diacyl glycerophosphocholines PC aa C34:3 ( $\beta = .21$ ; 95%CI .04 to .78), 1 plasmogen glycerophosphocholines PC ae C36:1 ( $\beta = .20$  95% CI .02 to .79) and negatively correlated with Kynurenine / Tryptophan ratio ( $\beta = -.19$  95% CI -.98 to -.004). Dichlorodiphenyldichloroethylene (DDE) was also negatively correlated with kynurenine / tryptophan ratio ( $\beta = -.21$  95% CI -1.87 to -.03) and 1 plasmogen glycerophosphocholines PC ae C34:3 ( $\beta = -.25$  95% CI -1.71 to -.11).

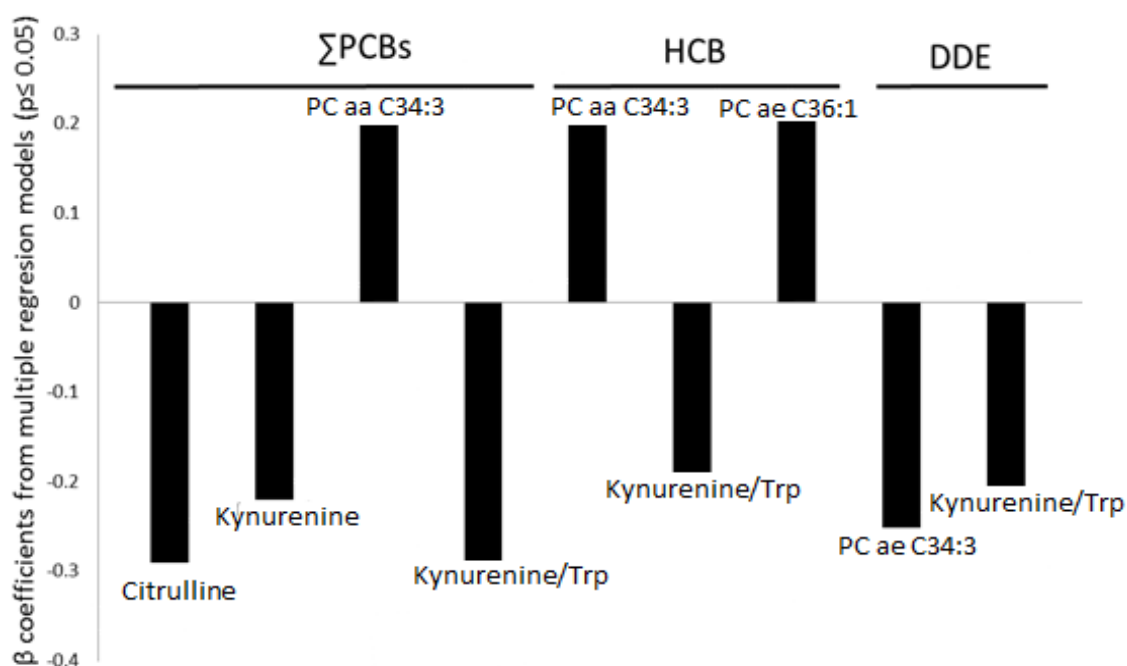


Figure 3.6  $\beta$  coefficients of 1st-trimester maternal serum metabolites associated with prenatal POPs exposures ( $\Sigma$ PCBs, HCB and DDE). These metabolites remained significant ( $P \leq 0.05$ ) after adjusting for covariates (mother's age, pre-pregnancy BMI and maternal education) in multiple regression models.

### 3.3.4 Associations between 1st-trimester organochlorine POPs exposure and the metabolite profile of cord serum

The aim of this analysis was to investigate potential links between current low levels of different POPs ( $\Sigma$ PCBs, HCB and DDE) in 1<sup>st</sup>-trimester maternal serum and metabolite profiles of umbilical cord serum in N=86 neonates with healthy birth outcomes from the RHEA pregnancy cohort. Spearman correlation coefficients were calculated to short list 'metabolites of interest' with significant associations ( $P \leq 0.05$ ); in total 14 metabolites and 3 metabolite ratios were identified to be linked with POPs exposure (Appendix Table 5). Of these, 3 metabolites and 1 metabolite ratio were associated with the sum of all PCB congeners ( $\Sigma$ PCBs), 7 metabolites and 1 metabolite ratio were associated with HCB, and 1 metabolite and 2 metabolite ratios were associated with DDE (Appendix Table 5).

To further assess the association of these metabolites of interest with POPs exposures, multiple regression models were performed with and without adjusting for POPs covariates found within this dataset. Table 3.6 represents Beta coefficients, 95% confidence intervals and their p values of unadjusted and adjusted multiple regression models for 1<sup>st</sup>-trimester maternal serum POPs exposures and metabolites of interest in cord serum.

|                                                        | <u>Unadjusted model</u>    |              | <u>Adjusted model<sup>b</sup></u> |              | <u>Bootstrapping<sup>c</sup></u> |
|--------------------------------------------------------|----------------------------|--------------|-----------------------------------|--------------|----------------------------------|
|                                                        | $\beta$ (95% CI)           | P value      | $\beta$ (95% CI)                  | P value      | p value (1000 models)            |
| <b><u>Log<sub>10</sub> <math>\Sigma</math>PCBs</u></b> |                            |              |                                   |              |                                  |
| Log <sub>10</sub> C3-DC (C4-OH)                        | .23 (.02, .56)             | 0.03         | .55 (.01, .93)                    | 0.08         | -                                |
| <b>Log<sub>10</sub> PC aa C34:2</b>                    | <b>.25 (.10, .99)</b>      | <b>0.01</b>  | <b>.26 (.17, .93)</b>             | <b>0.005</b> | <b>0</b>                         |
| <b>Log<sub>10</sub> PC aa C36:2</b>                    | <b>.29 (.18, 1.01)</b>     | <b>0.006</b> | <b>.29 (.22, .94)</b>             | <b>0.002</b> | <b>0.06</b>                      |
| <b>Log<sub>10</sub> Total DMA / Arg</b>                | <b>-.19 (-.38, .02)</b>    | <b>0.07</b>  | <b>-.19 (-.35, -.01)</b>          | <b>0.03</b>  | <b>0.03</b>                      |
| <b><u>Log<sub>10</sub> HCB</u></b>                     |                            |              |                                   |              |                                  |
| Log <sub>10</sub> PC aa C40:6                          | -.16 (-1.02, .14)          | 0.13         | -.12 (-.028, .008)                | 0.11         | -                                |
| Log <sub>10</sub> PC ae C42:3                          | -.11 (-1.05, .31)          | 0.28         | -.12 (-1.09, .29)                 | 0.25         | -                                |
| Log <sub>10</sub> C16-OH                               | -.16 (-1.21, .16)          | 0.13         | -.17 (-1.26, .16)                 | 0.13         | -                                |
| <b>Log<sub>10</sub> Citrulline</b>                     | <b>-.22 (-1.44, -.03)</b>  | <b>0.04</b>  | <b>-.23 (-1.50, -.07)</b>         | <b>0.03</b>  | <b>0.04</b>                      |
| <b>Log<sub>10</sub> Lysine</b>                         | <b>-.21 (-2.46, -.004)</b> | <b>0.05</b>  | <b>-.24 (-2.69, -.16)</b>         | <b>0.02</b>  | <b>0.06</b>                      |
| Log <sub>10</sub> Serine                               | -.14 (-2.15, .38)          | 0.17         | -.17 (-2.31, .26)                 | 0.11         | -                                |
| Log <sub>10</sub> Met-SO                               | .19(-.03, .65)             | 0.08         | .19 (-.03, .67)                   | 0.07         | -                                |
| Log <sub>10</sub> Met-So/Met                           | .14 (-.09, .53)            | 0.17         | .15 (-.08, .55)                   | 0.15         | -                                |
| <b><u>Log<sub>10</sub> DDE</u></b>                     |                            |              |                                   |              |                                  |
| Log <sub>10</sub> C16:1                                | .10 (-.20, .59)            | 0.33         | .02 (-.35, .45)                   | 0.81         | -                                |
| <b>Log<sub>10</sub> Tyr/Phe</b>                        | <b>.27 (.16, 1.22)</b>     | <b>0.01</b>  | <b>.24 (.11, 1.15)</b>            | <b>0.01</b>  | <b>0.08</b>                      |
| Log <sub>10</sub> Total DMA / Arg                      | .21 (.008, .53)            | 0.04         | .18 (-.03, .48)                   | 0.08         | -                                |

**Table 3.6 Unadjusted and adjusted multiple linear regression models for cord serum metabolites in relation to maternal serum concentrations of  $\Sigma$ PCBs, HCB and DDE.**

<sup>b</sup>Models adjusted for mothers' age and pre-pregnancy BMI. <sup>b</sup>bootstrapping=Robustness of the models was tested by bootstrap using 1000 resamplings. POPs exposures and metabolites were log-transformed for normalisation.

Beta coefficients from significant ( $P \leq 0.05$ ) adjusted multiple regression models for cord serum metabolites associated with POPs exposures are shown in Figure 3.11. Adjusted multiple regression models for  $\Sigma$ PCBs showed significant ( $p \leq 0.05$ ) positive association with cord serum metabolites PC aa C34:2 (adjusted  $\beta = 0.26$ ; 95%CI .17, to .93); PC aa C36:2 (adjusted  $\beta = 0.290$ ; 95%CI .229 to .942) and significant negative association with metabolite ratio Total DMA / Arg (adjusted  $\beta = -.19$ ; 95%CI-.354 to -.012). A negative association was observed between maternal HCB and cord serum metabolites citrulline (adjusted  $\beta = -.23$ ; 95%CI -1.50 to -.07) and lysine (adjusted  $\beta = -.24$ ; 95%CI -2.696 to -.164). For maternal DDE exposure, a significant positive correlation was found in cord serum metabolite ratio Tyrosine/Phenylalanine (adjusted  $\beta = .24$ ; 95%CI .11 to 1.15).

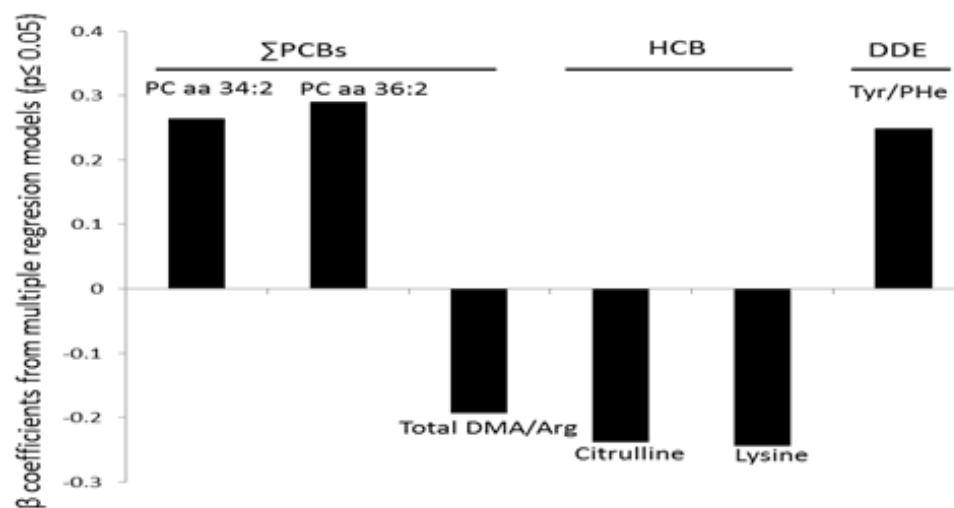


Figure 3.7  $\beta$  coefficients for the metabolites significant ( $p \leq 0.05$ ) in adjusted regression models for cord serum in relation to the maternal serum concentration of  $\Sigma$ PCBs, HCB, DDE and overall score. These metabolites remained significant ( $P \leq 0.05$ ) after adjusting for covariates (mother's age, pre-pregnancy BMI and maternal education) in multiple regression models.

### 3.4 Discussion

Targeted mass spectrometry-based metabolic profiles of healthy maternal and cord serum were distinguished by PCA analysis. This work describes for the first time, to best of my knowledge, a targeted LC/FIA-MS comparative study of maternal serum collected during the 1<sup>st</sup> trimester of pregnancy and umbilical cord serum collected at the time of the birth, with the aim to describe and quantify metabolic variations reflected in the serum of normal healthy pregnancies. Although in this study maternal and cord samples were not paired (due to logistics of sample availability and timing of the analysis), it still provides an average description of maternal-foetal (cord) metabolite profiles. This approach may not provide full details of the metabolic trajectory, but it demonstrates the more marked effects of pregnancy that were dominant over individual characteristics. One of the limitations of this work was that it did not cover the time-dependent course of metabolite variations throughout the pregnancy and only provides the basis for 1<sup>st</sup>-trimester maternal and cord metabolite profile. However, the results broadly agreed with those found in previous studies and confirmed the expected increase of essential amino acids in cord serum compared to maternal serum [51, 52, 53]. The levels of phosphatidylcholines, lysophosphatidylcholines except for (Lyso PC 16:1, Lyso PC 20:3 and Lyso PC 20:4), sphingolipids, and metabolites involved in fatty acid oxidation (free carnitine and long-chain acylcarnitines) were found to be higher in maternal serum whereas levels of short-chain carnitines (C2, C3, C4 and C5) were higher in cord serum. Considering *In utero*, the main sources for foetal metabolism are amino acids, glucose and free fatty acids [54], higher levels of amino acids and Lyso PC 20:4 observed in cord serum confirm this.

This work partly complements a recently published study using untargeted LC-MS and shotgun lipidomics to describe pregnancy induced metabolic phenotypes variations in maternal plasma [55]. This study followed metabolite trajectory of the metabolites classes including phospholipid metabolism (11 lipid classes), 6 amino acids, free carnitine and 10 acylcarnitines throughout the pregnancy. However, this study did not look at the metabolic variations in cord serum.

In the present study, the overall sum of all amino acid levels in maternal serum was less than those found in cord serum. These results are consistent with the literature where reduced levels of circulating amino acids are extensively studied in pregnancy and explained in terms of the increased placental transfer of amino acids to the foetus, favouring nitrogen metabolism which is considered extremely important for foetal development [51, 52]. The reduced levels of essential amino acids (lysine, histidine, methionine, tryptophan and threonine) noted here in 1<sup>st</sup>-trimester maternal serum compared to cord serum are shown to be particularly important in serving the foetus, early on in pregnancy [53]. The levels of tryptophan catabolite kynurenine (putatively

neuroprotective and neurotoxic) was also observed to be higher in cord serum compared to maternal serum, again these results are consistent with previously published study [56] which showed levels of kynurenine was significantly higher in umbilical artery and venous blood compared to maternal blood at the time of the birth.

Considering the importance of healthy dietary intake in pregnant women and energy demands for both mother and foetus, it is vital to focus on metabolite profile variations in lipids and carnitine species. Carnitine is biosynthesised mainly in liver and kidney from amino acids lysine (via trimethyllysine), and methionine and plasma concentrations of carnitines are known to be positively correlated with dietary consumption of carnitine [57]. In mammals, it is considered as a conditionally essential nutrient and plays a vital role in  $\beta$ -oxidation, a process in which fatty acid molecules are broken down and transported into the mitochondria of muscle cells including those in the heart for oxidation and to generate acetyl-CoA (Acetyl coenzyme A). Acylcarnitines are generated from carnitine by binding to fatty acids of varying carbon chain lengths, short-chain acylcarnitines being ( $\leq 10$  carbons) and long-chain acylcarnitines are those with ( $>10$  carbons) [58]. Carnitine palmitoyltransferase (CPT) plays a key role in the rate-limiting step in transporting long-chain fatty acyl-CoA across the inner mitochondrial membrane. It is suggested that CPT activity directly, or indirectly, may influence increased maternal blood consumption of long-chain acylcarnitines and may be associated with high-lipid diet-induced insulin resistance and free fatty acid levels developed in late pregnancy [59]. In the present study, maternal serum levels of free carnitine (C0) and long-chain acylcarnitines were higher, whereas short-chain acylcarnitines were higher in cord serum. In an older study, higher levels of long-chain acylcarnitines in maternal plasma were shown to reflect elevated fatty acid oxidation in maternal tissue while higher levels of short-chain acylcarnitines in cord plasma indicated the foetus is producing more or utilising less short-chain acylcarnitines [60]. In a pregnancy induced metabolic variation study [55] levels of free carnitine have been shown to decrease during pregnancy, while levels of short-chain acylcarnitines were found to increase in the second trimester but gradually decrease in late pregnancy. In contrast levels of long-chain acylcarnitines decreased in the second trimester and enhanced in late pregnancy. Decreased levels of plasma carnitine during pregnancy are also explained due to increased glomerular filtration rate [55]. Lower levels of free carnitine, short and long chain acylcarnitines are shown to be significantly associated with birth weight; levels were found to be considerably lower in SGA neonates compared to normal weight neonates [61].

Some of the lipids measured in this study were phosphatidylcholines (PC). These are glycerophospholipids which form about half of all cellular membrane phospholipids; they are the key building block of membrane bilayers and the major phospholipids found in plasma. A

schematic diagram of phosphatidylcholines is shown in Figure 3.12. Their structure consists of a glycerol backbone, and the hydroxyls of glycerol at C1 & C2 are esterified to fatty acids while the C3 hydroxyl is esterified to phosphate, the phosphate group is further esterified to choline as a polar head group. Phosphatidylcholines consists of a polar head group (hydrophilic: glycerol, phosphate and the polar head group-choline) and a non-polar tail (hydrophobic) 2 fatty acid.

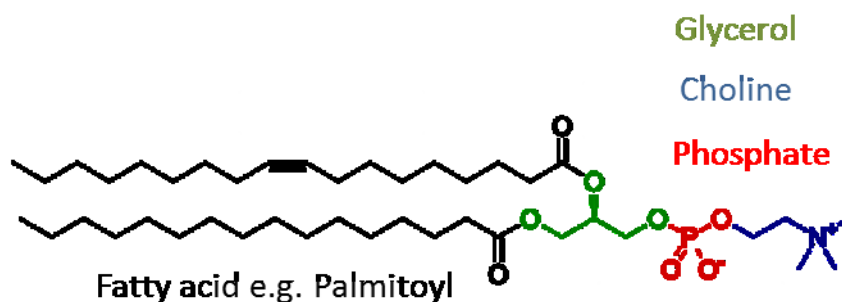


**Figure 3.8 A schematic diagram of Phosphatidylcholine structure.**

These fatty acids chains can be identical in structure but mostly tend to be non-identical with difference in carbon chain length and usually one is a saturated fatty acid (no double bond between carbon atoms of fatty acid chain) and/or unsaturated fatty acid (presence of one double bond = mono-unsaturated and presence more than one double bond= polyunsaturated). Phosphatidylcholines are further diversified into subclasses of diacyl (PC aa = diacyl) where both carbon atoms of the glycerol at C1 and C2 position are bound to ester linkage. Plasmalogens or ether Phosphatidylcholines (PC ae= acyl-alkyl) where one of the carbon atom on glycerol at C1 is bonded to an alkyl chain via an ether linkage and the other one at the C2 position bound to ester linkage. Lysophosphatidylcholine acyls (Lyso PC a) are also derived from phosphatidylcholines where one of the fatty acid chains is removed by partial hydrolysis [62].

Plasmalogens constitute around 20% of the total phospholipids in the cell membrane and makeup to 90% of the glycerophospholipids fraction in the brain while more than 50% in heart, neutrophils and eosinophils they are also abundant in kidney, lungs and skeletal muscle [63]. They play a vital role in protecting tissue-specific membrane lipids from oxidation and facilitating signalling processes. Plasmalogens are enriched with arachidonic fatty (AA, C20:4) acids, an omega 6 polyunsaturated fatty acid, and docosahexaenoic acid (DHA, C22:6) an omega 3 fatty acid. Their deficiency is linked to a peroxisome disorder (RCDP) [63].

Another class of phospholipid analysed in this study were sphingomyelins (SM); in mammals, these usually constitute 2-15% of the total organ phospholipid but could be more abundant in certain tissues like brain, peripheral nervous tissue and ocular lenses [64]. SM belongs to a class of lipid called sphingolipids which are characterised by the presence of a sphingosine base within

the hydrophobic part of the molecule and SM are the only phosphorus containing sphingolipids. The structure of SM consists of phosphocholine group from phosphatidylcholine as a polar head group, attached to hydrophobic (non-polar) sphingosine and a fatty acid (collectively categorised as ceramides), simultaneously yielding diacylglycerol. The SM which have a high degree of saturation, strong hydrogen bonding features and asymmetric molecular structure are strongly linked to their properties in the biological membrane. They undergo significant favourable interactions with cholesterol and other sterols; a growing body of evidence suggests that the functional role of SM is extensively linked to being a regulator of cholesterol distribution within biological membranes and maintaining cholesterol homeostasis in cells [64].

In the present study levels of diacyl (PC aa), plasmalogens (PC ae), and sphingomyelins (SM) in 1<sup>st</sup> trimester maternal were markedly higher compared to cord serum. This was also true for lyso-Phosphatidylcholine (lyso PC) except for LysoPC (C16:1), Lyso PC (C20:3) and Lyso PC (C20:4) which were significantly higher in cord serum. Studies in both mice and humans have demonstrated that the maternal sphingolipids metabolic pathway plays an important role in maintaining innate maternal-foetal immunity and fetomaternal tolerance by regulating proper uterine decidualization and blood vessel stability [65, 66]. In a pregnancy induced metabolic variation study [55] levels of PC were shown to increase throughout the pregnancy whereas levels of Lyso PC were found to be highest at 1<sup>st</sup> trimester and decreased significantly as the pregnancy progressed, especially for Lyso PC (C20:3) and Lyso PC (C20:4). The composition of essential fatty acids and long-chain polyunsaturated fatty acids (LCPUFA) determines its value in promoting health or disease in maternal, foetal, and neonatal medicine. The long chain polyunsaturated fatty acids (PUFA), C20 derivatives of essential fatty acids  $\alpha$ -linolenic acid, and linoleic acid (LA) such as dihomo gamma linolenic acid (DGLA 20:3, omega 6), arachidonic acid (AA; 20:4 omega-6), eicosapentaenoic acid (EPA; 20:5 omega-3) and docosahexaenoic acid (DHA; 22:6 omega -3) are biologically the most important fatty acids for proper foetal growth and development [67, 68]. Studies on maternal-foetal fatty acid composition have shown even though essential fatty acids represent about 10% of the total fatty acid in cord plasma and 30% in maternal plasma at birth the concentration levels of AA and DHA in foetal circulation are about twice as high as observed in the maternal circulation [68]. The maternal foetal transport of these PUFA is rapid during growth, and it is suggested that their deficiency can affect brain growth and has been linked to Intrauterine growth restriction (IUGR) [68, 69].

Phosphatidylcholines (PC) were positively associated with POPs after adjusting for maternal age, pre-pregnancy BMI and maternal education; for PCBs (PC aa C34:3) and HCB (PC aa C34:3, PC ae C36:1). These associations may indicate POPs affinity towards lipids and their positive association with long chain unsaturated fatty acids [70]. Although the exact position of the double

bonds and carbon atoms distribution in different fatty acid chains cannot be determined in this method, unsaturated fatty acids have been shown to be more vulnerable to lipid peroxidation which may promote the risk of disease [71, 72]. No study to date has associated POPs exposure with diacyl or plasmalogen phosphatidylcholines lipid metabolism and the role of maternal plasma phosphatidylcholines (apart from their omega 3 and omega 6 fatty acid content) in pregnancy is not very well understood. Interestingly, a study has linked variation in maternal plasma phosphatidylcholine fatty acids with child atopy and wheeze [73] at the age of 6. Another study has shown an increase in placental phosphatidylcholines (which are reflective of maternal phosphatidylcholines) in pre-eclamptic pregnancies versus normal pregnancies [74].

One other aim of this study was to assess the influence of low-level POPs exposures ( $\Sigma$ PCBs, HCB, and DDE) measured in 1<sup>st</sup>-trimester pregnancy on the metabolite profile of 1<sup>st</sup>-trimester pregnancy serum and cord serum and what biochemical pathways could be involved if any metabolic associations were observed. The results showed significant but not very strong associations between low doses prenatal exposure to  $\Sigma$ PCBs, HCB and DDE and metabolites of 1<sup>st</sup>-trimester maternal serum (Figure 3.10). Multiple regression models after adjusting for maternal age, pre-pregnancy BMI and maternal education showed a negative association between each of  $\Sigma$ PCBs, HCB and DDE exposure and the kynurenine to tryptophan ratio (Kyr/Trp).  $\Sigma$ PCBs was also negatively associated with kynurenine and citrulline while positively associated with PC aa C34:3. HCB was positively associated with PC aa C34:3 and PC ae C36:1. DDE was negatively associated with PC ae C34:3.

The kynurenine pathway represents a major route for tryptophan catabolism. 95% of the least abundant essential amino acid tryptophan can be metabolised through the kynurenine pathway. This pathway is mediated mainly in most mammalian tissues by the activity of the first and rate-limiting enzyme called Indoleamine 2, 3-dioxygenase (IDO). IDO is a powerful immunosuppressive enzyme its main function is to inhibit human immune system, the regulation of IDO biosynthesis and its activity in cells of the immune system can critically alter their immunosuppression or immune deficiency response to various immunological insults [75, 76]. The ratio of kyr/trp reflects the activity of IDO enzyme; a higher ratio is associated with accelerated depletion of tryptophan, increased kynurenine production and elevated IDO activity [75, 76].

The pioneering work done by Mellor *et al.* [77] first discovered the role of IDO in mammals which demonstrated placental cells express IDO-mediated immunosuppressive process that prevents foetal rejection *In utero* and protect the foetus from maternal T-cell attack. To date several studies have measured the levels of tryptophan and kynurenine and evaluated Kyr/Trp ratio under

various pathological conditions e.g. human immunodeficiency virus (HIV) [78], chronic kidney infections [79], schizophrenia [80], depression [81], cardiovascular disorders [82] and cancer [83]. Their objective was to estimate the degree of immune activation and to understand the relationship between tryptophan depletion, generation of kynurenine and kynurenine pathway in the diseased state. These studies demonstrated that the kyr/trp ratio increases are positively associated with progression of the disease, leading to the hypothesis that IDO-mediated immunosuppression is up-regulated as the inflammatory response associated with disease state progresses, compromising the body's immunity.

Another demonstrated mechanism of the kynurenine pathway and IDO activity is through interaction with aryl hydrocarbon receptors (AhR) [84]. Kynurenine is an endogenous ligand for AhR, activating AhR which in turn induces the expression of xenobiotic metabolising enzymes and genes such as cytochrome P450 needed for the detoxification of toxic AhR ligands. Some endogenous AhR ligands such as kynurenine may also bind and mediate AhR effects [85]. A recent study has shown that uremic toxins of tryptophan catabolism (kynurenine) share the ability to activate the AhR pathway similarly to dioxin environmental pollutants. This study also reviewed similarities between the effects of AhR-activating environmental pollutants and the effects of uremic kynurenine metabolites by inducing endothelial dysfunction, increasing oxidative stress, favour thrombosis and inflammation [86]. Another recent review has summarised several studies suggesting a possible mechanism between IDO and AhR in host-microbe interference within the gut [87] as AhR is known to regulate chronic gut inflammation. These studies have suggested a positive association between AhR and IDO expression.

In the present study, an opposite trend that to previously published data discussed above was observed. The ratio of Kyr/Trp was negatively associated with maternal POPs exposure with, for every 1-unit increase in POPs, an approximately 0.2 unit decrease in the Kyr/Trp ratio was observed. None of the above-discussed hypotheses explains this negative association. However, in a pregnancy study comparing normal and preeclampsia pregnancies, preeclamptic pregnancies showed a decrease in the maternal plasma Kyr/Trp ratio and reduced placental IDO activity compared to normal pregnancy [88]. In a normal pregnancy, the systemic inflammatory response is quite high which would explain increases in the Kyr/Trp ratio as maternal immunotolerance develops to protect and to avoid immunological rejection of the foetus. Following this, the ratio should be even greater in preeclampsia as a higher systemic inflammatory response instead of the converse was observed. This study hypothesised and demonstrated that Kyr/Trp ratio in pregnant women is regulated by placental IDO expression (IDO expression is highest in placenta) and if IDO expression is low in placenta due to dysfunctions observed during preeclampsia this will regulate the maternal Kyr/Trp ratio.

During normal healthy pregnancy, the maternal immunotolerance capacity increases so that antigens produced by the foetus don't get rejected by the mother. This immunotolerance is implicated as a driver of placental IDO activity. Following this hypothesis, it could be suggested that increase in POPs maternal exposure lowers maternal immunotolerance (expressed as Kyr/Trp ratio) towards the growing foetus which can cause implications on foetal growth and normal pregnancy. Further studies are required to understand the pregnancy-specific immune tolerance effect of POPs exposure on IDO expression and maternal Kyr/Trp ratio.

Another metabolite identified as negatively associated with PCBs was citrulline. Citrulline is a ubiquitous amino acid in mammals; it is synthesised in the liver from ornithine as part of the central reaction of the urea cycle by the key enzyme ornithine carbamoyltransferase (OCT). It is also produced from arginine as a by-product of the reaction catalysed by enzyme nitric oxide synthase (NOS) [89]. Studies have shown that restriction or manipulation in the supply of maternal citrulline induces nitric oxide (NO) deficiency which leads to oxidative stress in kidneys and the development of hypertension in newborn animals [89, 90, 91].

Overall, findings from this analysis provided metabolic insight of the potential metabolites and their ratios that could alter due to low-dose exposure of POPs resulting in adverse pregnancy outcomes, however, further research is required to fully understand the mechanism of POPs toxicity in pregnancy.

Analysis from this study also showed significant but not very strong associations between low prenatal exposure to  $\Sigma$ PCBs, HCB and DDE and metabolites of cord serum (Figure 3.11). Multiple regression models after adjusting for maternal age, pre-pregnancy BMI and maternal education showed a positive association between  $\Sigma$ PCBs and diacyl phosphatidylcholines (PC aa C34:2 and PC aa C36:2) while the negative association with a ratio of total dimethylated arginine to total unmodified arginine (Total DMA/Arg). As with the results found in maternal serum for PCBs exposure, diacyl-phosphatidylcholines carrying long-chain unsaturated fatty acids were also positively associated with prenatal PCBs exposure in cord serum. Previous studies have described polyunsaturated long-chain fatty acid alternations in cord blood related to intrauterine growth restriction and allergy development [92, 93]. A positive association was identified between prenatal DDE exposure and the ratio of tyrosine/phenylalanine (Tyr/Phe) in cord serum metabolites. This ratio is used to indicate the activity of phenylalanine hydroxylase (PAH) enzyme, and deficiency of this enzyme lead to phenylketonuria, an inborn error of metabolism.

### 3.5 Strengths and limitations

This study is the first to report targeted MS metabolic profiling approach with 145 metabolites and 21 metabolite sums and ratios in a mother-child prospective cohort, analysing biochemical effects of prenatal POPS exposure in 1<sup>st</sup> trimester maternal and umbilical cord serum. This study is also a first to compare and characterise metabolic phenotype variations observed between 1<sup>st</sup>-trimester maternal-foetal (cord) serums in healthy normal pregnancies. The clustering of two separate QCs among other samples and good inter-batch precision of measured metabolites demonstrates the quality of the data was the good. Although the sample size was small, the potential use of these data could be to provide a baseline or reference data set to be compared against specific conditions related to pregnancy (e.g. Preeclampsia, diabetes, preterm birth).

However, this study also has some limitations: (1) Small sample size and mother-child pairs were not matched: The number of samples used to analyse the association between prenatal POPs and metabolite profile of 1<sup>st</sup>-trimester maternal serum and cord serum was based on a small population set. The associations reported in this study were significant but not very strong, and none of the associations passed multiple testing corrections (FDR), and unwarranted biological significance may be attributed to false positive results. Also, we cannot rule out the modest effect of POPs on metabolites given the wide confidence intervals. The mother-cord serum samples were not matched in this study (mainly due to logistics of the samples). The present study still provided an average description of maternal-foetal (cord) metabolite profile. This approach may not provide full details of the metabolic trajectory, but it demonstrates the more marked effects of pregnancy dominant over individual characteristics.

(2) POPs exposure data in cord serum were not available: The association analysis between prenatal POPs exposure and metabolite profile of cord serum doesn't represent the exposure profile of POPs in cord blood itself. It has been suggested that levels of POPs in cord blood are marginal compared to those found in maternal blood because cord blood is not rich in lipids. The concentration of POPs levels in the RHEA cohort are amongst the lowest reported in 12 European cohorts [25] and does not generalise the exposure levels of the other European countries or rest of the world population. Samples from separate independent cohorts need to be analysed to validate any findings from this study. Clearly, in future work, the findings of this study need to be validated in comparable studies in other larger birth cohorts are required. This future work will be carried out in Human Early-Life Exposome (HELIX) project [94]. HELIX is explained in more detail in Chapter 4.

(3). Models adjusted for limited covariates: Some of the other limitations of this study that need to be addressed in any future work are: Only a limited number of covariates for POPs exposure

were considered (pre-pregnancy BMI, maternal age and maternal education, these covariates did attenuate the initial trends observed). Although fish consumption was not found to be associated with POPs within this data set, no other dietary source was adjusted in multiple regression models. A detailed analysis of diet and other lifestyle confounding factors which are main sources of POPs exposure and their effects on metabolites identified in this study could be a future study.

### 3.6 Conclusion

This study is the first to employ a targeted mass spectrometry-based approach to study the variation in metabolite profile of N=67 1<sup>st</sup>-trimester pregnancy serum from those of N=86 cord serum of healthy pregnancies. Results showed profound variation in maternal-foetal metabolites levels; including altered metabolic pathways of amino acids, biogenic amines, carnitines, lyso-phosphatidylcholines, phosphatidylcholines and sphingomyelins were associated with normal pregnancy. Although this comparison had limitations like maternal- cord samples were not paired, and analysis didn't provide a time-dependent metabolic trajectory throughout the pregnancy, it still demonstrated the more marked effects of pregnancy were dominant over individual characteristics. The potential use of these data could be to provide a baseline or reference dataset to be compared against specific conditions related to pregnancy (e.g. pre-eclampsia, diabetes, preterm birth).

This study is also the first exploratory work to demonstrate the potential of targeted mass spectrometry-based metabolic profiling approach to characterise the metabolic variations associated with low-dose prenatal POPs exposure in maternal and cord serum in a mother-child prospective study. From the discriminating metabolites identified in maternal serum, it could be hypothesised that low-dose maternal POPs exposure could alter the immunosuppressive activity of the indoleamine 2, 3-dioxygenase (IDO) enzyme indicated by kynurenine/tryptophan ratio. This study also demonstrated that low-dose prenatal POPs exposure is associated with concentration variation in citrulline and diacyl phosphatidylcholines in both maternal and cord serum which could have an impact on child health and pregnancy outcomes. However, the validations of these observations in larger population cohort and comparative studies with other cohorts are required. In future work, these observations will be further examined in Human Early-Life Exposome (HELIX) project.

Targeted mass spectrometry-based metabolic profiling has the potential to identify novel biomarkers related to the effects of persistent environmental toxicants exposures. In future work, the method used in this study could be employed on carefully selected populations set with high-dose prenatal POPs exposure to observe more marked effects of the mechanism of their toxicity and thus improve risk assessment models.

### 3.7 References

1. Mrema, E. J. *et al.* Persistent organochlorinated pesticides and mechanisms of their toxicity. *Toxicology* 307, 74–88 (2013).
2. Androutsopoulos, V.P. *et al.* A mechanistic overview of health associated effects of low levels of organochlorine and organophosphorus pesticides. *Toxicology* 307, 89-94 (2013).
3. Gourounti. K. Mechanism of actions and health effects of organochlorine substances. *Health Science Journal*, 2, 89-98 (2008).
4. Lee D.H. *et al.* Low dose organochlorine pesticides and polychlorinated biphenyls predict obesity, dyslipidemia, and insulin resistance among people free of diabetes. *PLoS One* 6 (1), (2011).
5. Vested . A. *et al.* Persistent organic pollutants and male reproductive health. *Asian Journal of Andrology*. 16, 71–80 (2014).
6. Wigle, D. T. *et al.* Epidemiologic evidence of relationships between reproductive and child health outcomes and environmental chemical contaminants. *J. Toxicol. Environ. Health. B. Crit. Rev.* 11, 373–517 (2008).
7. Faroon, O. *et al.* Effects of polychlorinated biphenyls on the nervous system. *Toxicol. Ind. Health* 16, 305–333 (2000).
8. Jacobson, J. L. & Jacobson, S. W. Intellectual impairment in children exposed to polychlorinated biphenyls in utero. *N. Engl. J. Med.* 335, 783–789 (1996).
9. Kamel, F. & Hoppin, J. A. Association of pesticide exposure with neurologic dysfunction and disease. *Environ. Health Perspect.* 112, 950–958 (2004).
10. Aronson, K. J. *et al.* Breast adipose tissue concentrations of polychlorinated biphenyls and other organochlorines and breast cancer risk. *Cancer Epidemiol. Biomarkers Prev.* 9, 55–63 (2000).
11. Cohn, B. A. *et al.* Prenatal DDT exposure and testicular cancer: a nested case-control study. *Arch. Environ. Occup. Health* 65, 127–34 (2010).
12. ATSDR. Toxicological Profile for PCB. U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES. Public Health Service. Agency for Toxic Substances and Disease Registry (2000). [ONLINE] Available at: <http://www.atsdr.cdc.gov/ToxProfiles/TP.asp?id=142&tid=26>
13. Grimm. F.A. *et al.* Metabolism and metabolites of polychlorinated biphenyls (PCBs). *Critical Reviews in Toxicology*. 45 (2015).
14. ATSDR. Toxicological Profile for Hexachlorobenzene. U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES. Public Health Service. Agency for Toxic Substances and Disease Registry (2013). [ONLINE] Available at: <http://www.atsdr.cdc.gov/toxprofiles/tp.asp?id=627&tid=115>
15. Gocemen A. *et al.* Hexachlorobenzene episode in Turkey. *Biomed Environ Sci.* 2, 36-43 (1989).
16. Bailey, R. E. Global hexachlorobenzene emissions. *Chemosphere* 43, 167–182 (2001).
17. Cripps D.J. *et al.* Porphyria turcica due to hexachlorobenzene: a 20 to 30 year follow-up study on 204 patients. *British Journal of dermatology* 111(4), 413-22 (1984).
18. Randi A.S. *et al.* Hexachlorobenzene is a tumour co-carcinogen and induces alterations in insulin-growth factors signalling pathway in the rat mammary gland. *Toxicology* 89, 83-92 (2005).
19. ATSDR. Toxicological Profile for DDT, DDE and DDD. U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES. Public Health Service. Agency for Toxic Substances and Disease Registry (2013). [ONLINE] Available at: <http://www.atsdr.cdc.gov/toxprofiles/tp.asp?id=81&tid=20>

20. Rogan, W. J. *et al.* Neonatal effects of transplacental exposure to PCBs and DDE. *J. Pediatr.* 109, 335–341 (1986).
21. Vizcaino, E. *et al.* Transport of persistent organic pollutants across the human placenta. *Environ. Int.* 65, 107–15 (2014).
22. Rice, D. & Barone, S. Critical periods of vulnerability for the developing nervous system: Evidence from humans and animal models. *Environ. Health Perspect.* 108, 511–533 (2000).
23. Grandjean P.B.D *et al.* The Faroes Statement: Human Health Effects of Developmental Exposure to Chemicals in Our Environment. *Basic & Clinical Pharmacology & Toxicology*, 102, 73–75 (2007)
24. Nickerson, K. Environmental contaminants in breast milk. *J. Midwifery Womens. Health* 51, 26–34 (2006).
25. Govarts, E. *et al.* Birth weight and prenatal exposure to polychlorinated biphenyls (PCBs) and dichlorodiphenyldichloroethylene (DDE): A meta-analysis within 12 European birth cohorts. *Environ. Health Perspect.* 120, 162–170 (2012).
26. Longnecker, M. P. *et al.* Association between maternal serum concentration of the DDT metabolite DDE and preterm and small-for-gestational-age babies at birth. *Lancet* 358, 110–4 (2001).
27. Wolff, M. S. *et al.* Prenatal pesticide and PCB exposures and birth outcomes. *Pediatr Res* 61, 243–250 (2007).
28. Stillerman, K. P. *et al.* Environmental exposures and adverse pregnancy outcomes: a review of the science. *Reprod. Sci.* 15, 631–650 (2008).
29. La Merrill, M. & Birnbaum, L. S. Childhood obesity and environmental chemicals. *Mt. Sinai J. Med.* 78, 22–48 (2011).
30. Valvi, D. *et al.* Prenatal exposure to persistent organic pollutants and rapid weight gain and overweight in infancy. *Obesity* 22, 488–496 (2014).
31. Valvi, D. *et al.* Prenatal concentrations of polychlorinated biphenyls, DDE, and DDT and overweight in children: A prospective birth cohort study. *Environ. Health Perspect.* 120, 451–457 (2012).
32. Winneke, G. Developmental aspects of environmental neurotoxicology: Lessons from lead and polychlorinated biphenyls. *J. Neurol. Sci.* 308, 9–15 (2011).
33. Torres-Sánchez, L. *et al.* Prenatal p,p'-DDE exposure and neurodevelopment among children 3.5-5 years of age. *Environ. Health Perspect.* 121, 263–8 (2013).
34. Park, H.-Y. *et al.* Neurodevelopmental toxicity of prenatal polychlorinated biphenyls (PCBs) by chemical structure and activity: a birth cohort study. *Environ. Health* 9, 51 (2010).
35. Sharpe, R. & Irvine, D. How strong is the evidence of a link between environmental chemicals and adverse effects on human reproductive health? *Br. Med. J.* 328, 447–451 (2004).
36. Papadopoulou, E. *et al.* Maternal diet, prenatal exposure to dioxins and other persistent organic pollutants and anogenital distance in children. *Sci. Total Environ.* 461-462, 222–9 (2013).
37. Ellis, J. K. *et al.* Metabolic profiling detects early effects of environmental and lifestyle exposure to cadmium in a human population. *BMC Med.* 10, 61 (2012).
38. Kuo, C. H. *et al.* Metabolomic characterization of labourers exposed to welding fumes. *Chem. Res. Toxicol.* 25, 676–686 (2012).
39. Bonvallot, N. *et al.* Metabolomics tools for describing complex pesticide exposure in pregnant women in Brittany (France). *PLoS One* 8, e64433 (2013).
40. Siskos, A. P. *et al.* Inter-laboratory reproducibility of a targeted metabolomics platform for analysis of human serum and plasma. *Anal. Chem.* (2016).

41. Ramirez, T. *et al.* Metabolomics in toxicology and preclinical research. *ALTEX* 30, 209–25 (2013).
42. Chatzi, L. *et al.* Metabolic syndrome in early pregnancy and risk of preterm birth. *Am. J. Epidemiol.* 170, 829–836 (2009).
43. Vafeiadi, M. *et al.* Persistent organic pollutants exposure during pregnancy, maternal gestational weight gain, and birth outcomes in the mother-child cohort in Crete, Greece (RHEA study). *Environ. Int.* 64, 116–23 (2014).
44. Koponen, J. *et al.* Determination of selected perfluorinated alkyl acids and persistent organic pollutants from a small volume human serum sample relevant for epidemiological studies. *J. Chromatogr. A* 1309, 48–55 (2013).
45. Schisterman, E. F. *et al.* Lipid adjustment in the analysis of environmental contaminants and human health risks. *Environ. Health Perspect.* 113, 853–857 (2005).
46. Xu, T. *et al.* Effects of smoking and smoking cessation on human serum metabolite profile: results from the KORA cohort study. *BMC Med.* 11, 60 (2013).
47. Floegel, A. *et al.* Reliability of serum metabolite concentrations over a 4-month period using a targeted metabolomics approach. *PloS One* 6, (2011).
48. Carpenter, J. *et al.* Bootstrap confidence intervals: When, which, what? A practical guide for medical statisticians. *Statistics in Medicine*, 19(9), 1141–1164 (2000).
49. Xia, J. *et al.* Translational biomarker discovery in clinical metabolomics: an introductory tutorial. *Metabolomics* 2, 280–299 (2013).
50. Storey, J. D. The positive false discovery rate: A Bayesian interpretation and the q-value. *Ann. Stat.* 31, 2013–2035 (2003).
51. Tea, I. *et al.* 1H-NMR-based metabolic profiling of maternal and umbilical cord blood indicates altered materno-foetal nutrient exchange in preterm infants. *PLoS One* 7, e29947 (2012).
52. Regnault, T. R. *et al.* Fetoplacental transport and utilization of amino acids in IUGR--a review. *Placenta* 26 Suppl A, S52–62 (2005).
53. Velazquez, A. *et al.* Amino Acid Pools in the Feto-Maternal System. *Neonatology*. 29, 1–2 (1976).
54. Houwelingen, A. C. & Hornstra, G. Long-chain polyunsaturated fatty acids, pregnancy, and pregnancy outcome. *Am. J. Clin. Nutr.* 71, 285S–91S (2000).
55. Luan, H. *et al.* Pregnancy-induced metabolic phenotype variations in maternal plasma. *J. Proteome Res.* 13, 1527–1536 (2015).
56. Kazda, H. *et al.* Maternal, umbilical, and amniotic fluid concentrations of tryptophan and kynurenine after labour or caesarean section. *Pediatr Res* 44, 368–373 (1998).
57. Steiber, A. *et al.* Carnitine: A nutritional, biosynthetic, and functional perspective. *Mol. Aspects Med.* 25, 455–473 (2004).
58. Flanagan, J. L. *et al.* Role of carnitine in disease. *Nutr. Metab. (Lond)*. 7, 30 (2010).
59. Bruce, C. R. *et al.* Overexpression of carnitine palmitoyltransferase-1 in skeletal muscle is sufficient to enhance fatty acid oxidation and improve high-fat diet-induced insulin resistance. *Diabetes* 58, 550–558 (2009).
60. Novak, M. *et al.* Carnitine in the perinatal metabolism of lipids. I. Relationship between maternal and foetal plasma levels of carnitine and acylcarnitines. *Pediatrics* 67, 95–100 (1981).
61. Meyburg, J. *et al.* Postnatal changes in neonatal acylcarnitine profile. *Pediatr. Res.* 49, 125–129 (2001).
62. Phosphatidylcholine and Related Lipids. [ONLINE] Available at: (<http://www.lipidhome.co.uk/lipids/complex/pc/index.htm>)

63. Braverman, N. E. & Moser, A. B. Functions of plasmalogen lipids in health and disease. *Biochim. Biophys. Acta - Mol. Basis Dis.* 1822, 1442–1452 (2012).
64. Slotte, J. P. & Ramstedt, B. The functional role of sphingomyelin in cell membranes. *Eur. J. Lipid Sci. Technol.* 109, 977–981 (2007).
65. Mizugishi, K. *et al.* Sphingolipid Pathway Regulates Innate Immune Responses at the Fetomaternal Interface during Pregnancy. *J. Biol. Chem.* 290(4), 2053–68 (2014).
66. Mizugishi, K. *et al.* Maternal disturbance in activated sphingolipid metabolism causes pregnancy loss in mice. *J. Clin. Invest.* 117, 2993–3006 (2007).
67. Duttaroy, A. K. Transport of fatty acids across the human placenta: A review. *Prog. Lipid Res.* 48, 52–61 (2009).
68. Jones, M. L. *et al.* Maternal dietary omega-3 fatty acids and placental function. *Reproduction* 1–26 (2014).
69. Guerranti, C. & Focardi, S. E. Differences in Pop Levels between Conventional and Omega-3 Fatty Acid-Enriched Milk and Dairy Products. *ISRN Toxicol.* 2011, 541694 (2011).
70. Gavino, V. C. *et al.* Effect of polyunsaturated fatty acids and antioxidants on lipid peroxidation in tissue cultures. *J. Lipid Res.* 22, 763–769 (1981).
71. Cai, F. *et al.* Role of polyunsaturated fatty acids and lipid peroxidation on colorectal cancer risk and treatments. *Curr. Opin. Clin. Nutr. Metab. Care* 15, 99–106 (2012).
72. Pike, K. C. *et al.* Maternal plasma phosphatidylcholine fatty acids and atopy and wheeze in the offspring at age of 6 years. *Clin. Dev. Immunol.* 2012, 474613 (2012).
73. Huang, X. *et al.* Increased placental phospholipid levels in pre-eclamptic pregnancies. *Int. J. Mol. Sci.* 14, 3487–3499 (2013).
74. Houwelingen, A. C. & Hornstra, G. Long-chain polyunsaturated fatty acids, pregnancy, and pregnancy outcome. *Am. J. Clin. Nutr.* 71, 285S–91S (2000).
75. Chen, Y. & Guillemin, G. J. Kynurenine pathway metabolites in humans: Disease and healthy states. *Int. J. Tryptophan Res.* 2, 1–19 (2009).
76. Mbongue J.C. The role of indoleamine 2, 3-dioxygenase in immune suppression and autoimmunity. *Vaccines.* 10, 703-29 (2015).
77. Mellor A. L. *et al.* Indoleamine 2,3-dioxygenase, immunosuppression and pregnancy. *J. Reprod. Immunol.* 57, 143–150. (2002).
78. Boasso, A. & Shearer, G. M. How does indoleamine 2,3-dioxygenase contribute to HIV-mediated immune dysregulation. *Curr. Drug Metab.* 8, 217–223 (2007).
79. Schefold, J. C. *et al.* Increased indoleamine 2,3-dioxygenase (IDO) activity and elevated serum levels of tryptophan catabolites in patients with chronic kidney disease: a possible link between chronic inflammation and uraemic symptoms. *Nephrol. Dial. Transplant* 24, 1901–8 (2009).
80. Müller, N. *et al.* Kynurenine pathway in schizophrenia: pathophysiological and therapeutic aspects. *Curr. Pharm. Des.* 17, 130–136 (2011).
81. Christmas, D. M. *et al.* A biological pathway linking inflammation and depression: Activation of indoleamine 2,3-dioxygenase. *Neuropsychiatr. Dis. Treat.* 7, 431–439 (2011).
82. Polyzos, K. A. & Ketelhuth, D. F. J. The role of the kynurenine pathway of tryptophan metabolism in cardiovascular disease. An emerging field. *Hamostaseologie* 35, 128–136 (2015).
83. Zamanakou, M. *et al.* Tumor immune escape mediated by indoleamine 2,3-dioxygenase. *Immunol. Lett.* 111, 69–75 (2007).
84. Nguyen, N. T. *et al.* Aryl hydrocarbon receptor and kynurenine: recent advances in autoimmune disease research. *Front. Immunol.* 5, 551 (2014).

85. Heath-Pagliuso, S. *et al.* Activation of the Ah receptor by tryptophan and tryptophan metabolites. *Biochemistry* 37, 11508–11515 (1998).
86. Sallée, M. *et al.* The aryl hydrocarbon receptor-activating effect of uremic toxins from tryptophan metabolism: A new concept to understand cardiovascular complications of chronic kidney disease. *Toxins (Basel)*. 6, 934–949 (2014).
87. Zelante T. *et al.* Tryptophan feeding of the IDO1-AhR axis in host-microbial symbiosis. *Front Immunol.* 25 (2014).
88. Kudo, Y. *et al.* Decreased tryptophan catabolism by placental indoleamine 2,3-dioxygenase in preeclampsia. *Am. J. Obstet. Gynecol.* 188, 719–726 (2003).
89. Curis, E. *et al.* Almost all about citrulline in mammals. *Amino Acids* 29, 177–205 (2005).
90. Chien, S. *et al.* Two different approaches to restore renal nitric oxide and prevent hypertension in young spontaneously hypertensive rats: l-citrulline and nitrate. *Transl. Res.* 163, 43–52 (2014).
91. Tain, Y.-L. *et al.* Effects of maternal L-citrulline supplementation on renal function and blood pressure in offspring exposed to maternal caloric restriction: the impact of nitric oxide pathway. *Nitric Oxide* 23, 34–41 (2010).
92. Barman, M. *et al.* High levels of both n-3 and n-6 long-chain polyunsaturated fatty acids in cord serum phospholipids predict allergy development. *PLoS One* 8, e67920 (2013).
93. Cetin, I. *et al.* Intrauterine growth restriction is associated with changes in polyunsaturated fatty acid fetal-maternal relationships. *Pediatric. Res.* 52, 750–5 (2002).
94. Vrijheid, M. *et al.* The Human Early-Life Exposome (HELIX): Project Rationale and Design. *Environ. Health Perspect.* 122, 535–44 (2014).

## **Chapter 4. Untargeted LC-MS metabolic profiling to determine sources of analytical and short-term temporal variation in urinary metabolite measurements: application to the Human Early-Life Exposome (HELIX) project.**

### **4.1 Introduction**

#### **4.1.1 The Human Early-Life Exposome (HELIX) project**

One of the current initiatives in exposome science seeking successful integration of ‘omics’ technologies within epidemiological studies is the European Commission funded project HELIX (Human Early-Life Exposome), initiated in 2012 [1]. The general aim of the HELIX project is to characterise the ‘early-life exposome’ by scheduling the measurement of exposure from critical periods of early pregnancy to childbirth and following up through early childhood years (‘early life’) using mother-child cohorts. The conceptual framework of HELIX project is shown in Figure 4.1.

The HELIX project was divided into three key research areas. The first is to measure a wide range of environmental contaminants, e.g. water contaminants, air pollution, consumer products, noise pollution, UV radiation, persistent organic pollutants together with lifestyle factors, e.g. diet, mobility, by personal exposure monitoring, remote sensing, geographical information system-based spatial methods devices and questionnaires. The second is to assess the internal exposome by implementing ‘omics’ based approaches to measure molecular signatures in biofluids from the mothers and children in the cohorts. The measurement of the internal exposome using ‘omics’ tools is a relatively new and important approach compared to traditional exposure measurement approaches. It will enable holistic understanding of the relationship between combined early life environmental exposures and their impact on key child health outcomes (specifically in this project: obesity, neurodevelopment, respiratory and reproductive health). The third element is to integrate the external and internal exposome by developing novel bioinformatics/statistical approaches.

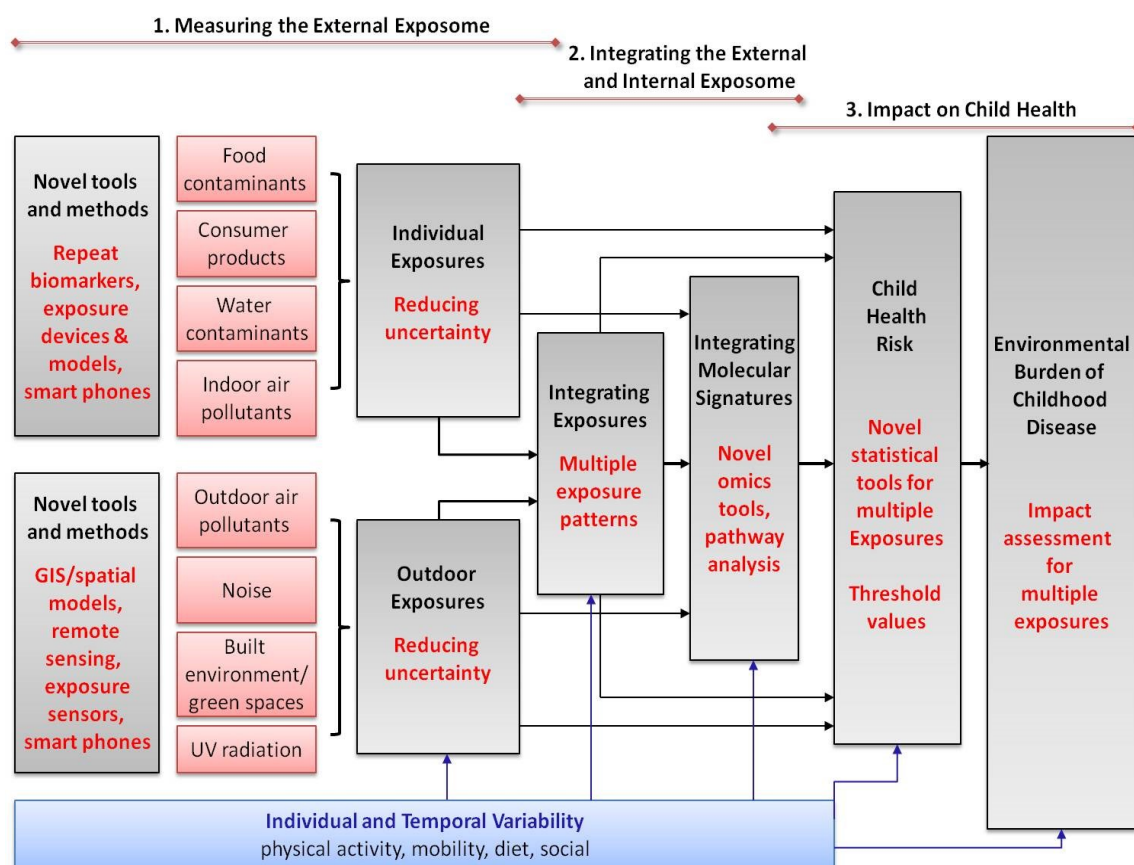


**Figure 4.1 Human Early-Life Exposome (HELIX) project conceptual framework [Figure is taken from Ref 1]**

HELIX consists of six prospective longitudinal European population-based birth cohorts (total 32,000 mother-child pairs) including RHEA (Greece) [2], BiB (UK) [3], EDEN (France) [4], INMA (Spain) [5], KANC (Lithuania) [6] and MoBa (Norway) [7]. Selected from these cohorts, 1200 mother-child pairs (~200 pairs from each cohort) will form the initial basis of the biomarker discovery work. The participating cohorts already have a large amount of existing exposure and 'omics' data following from early pregnancy through to childhood. Their inclusion in the HELIX project will allow harmonisation of the existing data with the new follow-up research on the children health outcomes at a similar age (6-9 years) across all six European cohorts. The follow-up examination allowed collection of new biofluids from children for exposure assessment and 'omics' profiles, integration of new questionnaires' and clinical examinations using harmonised standardised protocols. The details of the HELIX project including exposure assessments and study design is explained by Vrijheid *et al.* [1].

One of the initially stated objectives of the HELIX project was to define and identify sources of exposure variations (analytical, temporal, behavioural, diet, toxicokinetic) within and between individuals. Therefore, the first step in metabolome studies was to conduct a pilot study to assess

the analytical and short-term temporal variability of a subset of metabolites detected in urine and evaluate their implications for optimal study design for future HELIX studies.

#### **4.1.2 Untargeted metabolic profiling for characterising the human exposome**

Untargeted metabolic profiling methods are central to efforts to define the internal exposome as these provide – in principle - wide metabolite profile coverage of individuals. The implementation of this approach has been discussed with examples in recent publications [8, 9]. As described in Chapter 1 of this thesis, the inherent complexity of the human metabolome, due to immense chemical diversity and high dynamic range, make it a complex part of the exposome. Also, analytical platforms used to characterise the metabolome via untargeted profiling in large population studies need to be high throughput, robust, reproducible, provide good coverage of a range of metabolites all at a reasonable cost.

Both LC-MS and  $^1\text{H}$  NMR spectroscopic based metabolic profiling are extensively used and appraised in the context of population-based studies [10, 11]. These studies show that both the techniques are well suited to meet the requirements for some of the challenges in large population studies such as high throughput analysis, costs associated with it, time requirement and sample volume availability.

The main strength of  $^1\text{H}$  NMR is reproducibility; the analytical reproducibility for urinary metabolites has been demonstrated to be <3% in identical samples [12, 13]. Due to its robustness and non-destructive nature, NMR spectroscopy-based metabolic profiling has been favoured in population-based studies as a screening technique despite its limitation regarding method sensitivity. It has been used to present an overall view of similarities and outliers within the sample sets, screen out any sources of contamination and help to avoid any adverse analytical effects on the larger subsequent runs.

In contrast to  $^1\text{H}$  NMR, the main advantage of the LC-MS profiling is its sensitivity. However, compared to  $^1\text{H}$  NMR, LC-MS is more prone to analytical variability typically due to the preliminary chromatography stage and matrix effects leading to ion suppression. LC-MS profiling approaches are often used in a targeted manner on selected subgroups of metabolite classes for more sensitivity. One of the reasons for the growing utilisation of LC-MS in an untargeted manner is due to its widespread availability and more trained researchers are available who can operate and acquire LC-MS data compared to NMR [14].

### 4.1.3 Sources of metabolite level variability

The two most common biofluids collected and studied in metabolic profiling are urine and blood samples. Urine is a popular biofluid for investigation, especially in studies involving children, due to the simplicity of collecting repeated samples and it is completely non-invasive. Also, urine often provides a dense metabolic composition and higher metabolite concentration compared to blood plasma [15]. In large epidemiological studies, the general aim is to define 'usual or normal' metabolite level at homeostasis. A certain degree of homeostasis is required to occur in normal healthy urine metabolite levels, the elevation of these 'normal levels' is presumed to be an indicator of altered metabolism that could indicate disease risk or exposure.

While some studies have shown the potential of metabolic profiling to find intermediate markers of exposure and disease risk in both urine and blood plasma samples [16, 17, 18], the utility of such markers in clinical settings can be diminished by intra- (within) and inter- (between) individual variance. This has been shown to occur due to the circadian rhythm or lunar rhythm of the body [19], temporal variation [20], gender and age [21, 22], diet interventions [23], menstrual cycle [24], sample degradation over time [25] and experimental analytical variability [26]. In addition to the above factors, concerns have also been raised on the representativeness of a single spot urine sample, which often the only biofluid sample available in biomonitoring/exposure studies [27, 28]. Furthermore, in urine samples the concentration of endogenous metabolites can vary based on water consumption; variations in urine volume up to 15-fold have been reported in the literature [29]. In most cases, the urine samples are normalised for the volume variation by dividing the data to a common, relatively constant and measurable metabolite, creatinine [30]. Metabolites that vary largely due to day to day variability between individuals, or are measured with low experimental, analytical reliability can further reduce the power of associations between exposure and disease risk [30].

The two criteria for a single or suite of biomarkers to be useful in epidemiological studies are: that their intra-individual variation should not vary too much such that measurement of a single point in time would lose statistical power to identify an association. The second criterion is that its levels should not solely be influenced by genotype especially if disease risk is influenced by environmental exposures [31]. Therefore, for the successful transition of candidate metabolic biomarkers from epidemiological studies into clinical settings, it is vital to characterise sources of intra- (within) and inter- (between) individual variance within a normal healthy population.

Recently there has been increased interest in understanding how pre-analytical factors such as study design, sample collection [32, 33], sample storage conditions [33, 34] and sample preparation before analytical run [15, 34] can have an impact on the quality of the metabolite

profile of serum or plasma. However, there is a scarcity of similar information for urine samples. Therefore, it is vital to standardise pre-analytical factors as the mistakes made during sample collection stages cannot be rectified by even the most advanced analytical technique or statistical methods [26].

Irrespective of these issues, currently there is a lack of information detailing longitudinal normal urinary metabolite variation in healthy populations, especially in children. To the best of my knowledge, there are only three previously published LC-MS untargeted metabolic profiling studies which systematically investigated the sources of variation (analytical and temporal variation) in urine samples of a healthy population [20, 30, 35]. As finance and other resources are a major objective in running repeat samples from large epidemiological studies, no study has explored the alternative experimental design of pooling multiple time point samples from the same individual and analyse as a single measurement to investigate the reliability of the pooled samples.

#### **4.1.4 Aim**

Driven by the lack of information on the degree of LC-MS urinary metabolome variability over time the aim of the present study was to characterise sources of analytical and short-term temporal biological variation in urine samples from healthy children participating in the HELIX project.

This pilot study was the first to investigate such variability in healthy children's urine. Also, it was the first study to investigate the stability of the metabolite profile in pooled diurnal urine samples compared to individual time points (morning and night) from the same child, providing important information for the design of the full HELIX sub-cohort study.

#### **4.1.5 Objectives**

The main objectives were:

- To implement an untargeted LC-MS protocol for the analysis of urine
- To build a chemical reference standard library to identify metabolites in untargeted LC-MS profiling.
- To calculate the experimental, analytical variation in pooled QC samples.
- To investigate the short-term intra- and inter-individual biological variability (6 days) of a selected set of metabolites in a small panel study of children urine samples (N= 20, age 7-9 years) from the INMA cohort within the HELIX project using LC-MS analysis.

- To estimate the reliability and stability of pooled diurnal (morning void and before bed) time point samples compared to individual time point (morning and night) samples from the same child.

## 4.2 Materials and Methods

### 4.2.1 Materials

The solvents and reagents required were: Water, Acetonitrile, Formic acid, sodium formate, and isopropanol were all LC-MS grade and purchased from Sigma-Aldrich. The UPLC-ESI-MS/MS instrument consisted of a Waters Xevo® G2-S quadrupole-Time of flight (Q-TOF) mass spectrometer coupled with a Waters ACQUITY UPLC system. Both the UPLC and the MS were controlled by Waters software (MassLynx version 4.1). Chromatographic separation was carried out on a Waters UPLC C18 HSS T3 column. Other equipment required were: 350 µL 96-well plates (Waters), plate mats (VWR) and Eppendorf MixMate® shaker.

### 4.2.2 Study design

The children's urine samples (N=20, age 7-9 years) analysed in the present study were drawn from the Spanish cohort INMA within the HELIX project. These samples formed a part of a children 'panel' study (N=150) nested within the main sub-cohort (N=1200) which collected in-depth person exposure data and repeat urine samples for a one-week period. Two urine samples from each of the 20 children (first-morning void and before bed night-time void) were collected over one entire week (7 days and morning of the 8<sup>th</sup> day).

A detailed questionnaire on biological variables which could impact metabolite profiles such as the use of any medication, meal time, diet, gender and physical activity were recorded. The likely sources of technical variability which might also influence metabolite profiles such as timings between sample collection and storage and any deviation from the protocol were collected. Health examinations of the participating children in the study were carried out by trained nurses following a harmonised protocol recording factors such as height, weight, skinfolds, BMI, blood pressure and spirometry.

For the UPLC-MS metabolic profiling analysis, the incorporation of longitudinal morning and night sampling in the study design allowed investigating the short-term (6 days) temporal biological variability of the detected metabolites. The morning and night samples were pooled (50:50) to obtain daily pool (or diurnal pools) samples per child. The incorporation of the daily pool samples allowed an investigation into whether a pooled sample of multiple points (day and night) would generate more a reliable and reproducible metabolic profile of an individual, compared to analysing single spot samples from a specific time point. This information could be an alternative study design when resources are limited and do not allow analysis of repeat samples from large epidemiological studies. Figure 4.2 shows the study design and urine sample pooling strategy (simplified to show collection across 6 days) of the samples whose results are

reported and discussed in the present chapter. While urine samples were collected and analysed over 7 days and only the morning of day 8, the samples from the day 7 (morning and night) and day eight morning were excluded from the data analysis due to the difference in their sample handling and storage temperature.

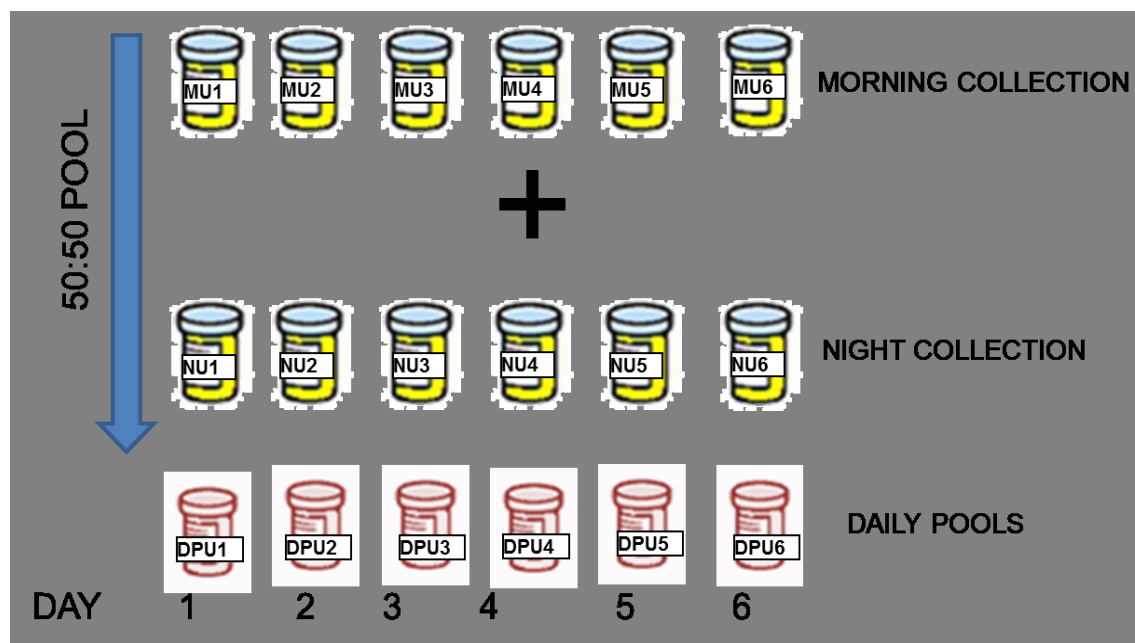


**Figure 4.2 Children urine (N=20) morning void and before bed night void sample collection and daily pool (50:50 morning and night urine) strategy across six days.** Day 1 to Day 6 urine sample collection from N=20 children, in total 18 urine samples per child were analysed = 6 mornings, 6 nights and 6 daily pools. From 20 children N=360 urine samples were used to estimate the short-term temporal biological variability in a set of metabolites detected via untargeted LC-MS metabolic profiling.

In addition to daily pools, weekly pools were also prepared, and their LC-MS metabolite profiles were captured. The total number of urine samples analysed per child were 25: 6 morning urine, 6 night urine, 6 daily pools (50:50 morning and night pool), 1 weekly morning pool (pooled urine from 6 mornings per child), 1 weekly night pool (pooled urine from 6 nights per child), 1 weekly pool of all daily pools, 1 weekly pool of mornings, nights and daily pool per child. Additional 3 more weekly pools were also prepared per child including morning and night urine of 7 and morning urine on day 8. Consequently, from N=20 children, in total 500 samples were analysed. These 500 samples were divided into two aliquots, one for the LC-MS profiling and another one for the  $^1\text{H}$  NMR profiling (results from which are presented elsewhere [36]). In this chapter, only data analysis of N=360 samples were carried out (as shown in study design Figure 4.2) to answer the main aim and objectives of the study. In the future work the remaining 140 samples including 1 weekly morning pool (pooled urine from 6 mornings per child), 1 weekly night pool (pooled

urine from 6 nights per child), 1 weekly pool of all daily pools, 1 weekly pool of mornings, nights and daily pool per child will be also analysed.

### **QC pooling**

Quality control (QC) samples were prepared by pooling 500  $\mu$ L from each of the 360 samples obtained from 20 children described above. The pooled QC samples were aliquoted in 1.3 mL cryovials and stored at -40°C freezer to be used for all HELIX study related future analysis. These QC samples were used to assess analytical precision and reliability of the annotated metabolites captured using the untargeted LC-MS method. The QC samples were incorporated after every 10 urine samples in the LC-MS experimental run per batch. This extensive replication of QC samples within the present study allowed estimation of *non-biological* variation or analytical variation introduced by experimental error and day-to-day instrument technical variability such as a shift in retention time or signal intensity over time [37, 38]. It also allowed assessing the reproducibility of the methodology in large-scale continuous extended analytical runs.

### **4.2.3 LC-MS sample preparation and untargeted metabolic profiling**

#### **Urine sample preparation**

The untargeted LC-MS urinary metabolic profiling analysis used in the present study was based on the method published by Want *et al.* in Nature Protocol [37]. Urine sample aliquots which were stored at -40°C were thawed on ice and homogenised on the vortex mixer. The samples were prepared fresh before the LC-MS analysis per 96-well plate/batch. 60  $\mu$ L of urine sample aliquots were centrifuged at 10,000g (4 °C) for 10 min to remove any particulates. Supernatants (50  $\mu$ L) of urine sample were removed and diluted in LC-MS grade water in 1:2 ratio and were mixed over the vortex for 10 seconds before transferring the samples to 96-well plates.

As a method of experimental control and to minimise the impact of any analytical bias/drift introduced via batch effects or run order, sample order in the 96-well plates were randomised. Sample order randomisation was carried out to minimise any bias or confounding factors observed due to the placement of homogenous samples within the same batch. To reduce the impact of batch effects, an ideal experimental design should ensure that the homogeneous test samples are evenly spread across different batches. This approach would allow observing the separation of the real biological significance from the effects observed due to batch effects or any other confounding factor more statistically powerful. Plate maps for 6 X 96-well plates were generated using Microsoft Excel function and samples were placed according to the plate maps. The 96-well plates were covered with cap mats for LC-MS analysis.

The urine samples in 96-well plates were transferred to LC autosampler maintained at 4°C and analysed within one day. The total number of samples analysed were 556 across 6 plates/batches including 6 blanks (1 blank per plate: to check for any contamination within the solvents or UPLC-MS system), 500 urine samples (86 urine samples in plate 1 to 5 and 70 urine samples in plate 6) and 51 QCs (9 QCs in plate 1 to 5 and 6 QCs in plate 5). The total LC-MS runtime for the entire analysis was two weeks in continuous operation (Week 1 positive ionisation mode and second-week negative ionisation mode). After running a plate in the positive mode, it was stored at -20°C till it was run again in negative mode. The LC-MS was in continuous operation between switching from positive to negative ionisation mode. As a measure of good UPLC-MS practise, five blanks and five QCs were injected at the beginning of the runs in each ionisation mode for the system equilibration and conditioning as mentioned in Want *et al.* Nature Protocol [37]. Using pooled QCs to condition the LC or UPLC-MS system at the beginning of the run allows to reduce the retention time variability as the column is more stabilised and provide a better representation of the actual test samples through the remainder of the run [37].

### LC-MS untargeted metabolic profiling

The UPLC-MS method was performed on a Waters ACQUITY UPLC® system using reversed phase high silica chromatography (ACQUITY UPLC HSS T3 C18 2.1 X 100 mm 1.7 µm column) coupled to a Waters Xevo G2-S Q-ToF mass spectrometer with an ESI source. The column temperature was maintained at 40°C and injection volume was five µL. The mobile phase A consisted of 0.1% formic acid in water and mobile phase B as 0.1% formic acid in acetonitrile. The gradient was run at 500 µL/min and consisted of an isocratic separation for 1 minute of 1% mobile phase B and then B was increased at linear rate to 95% at 9 minutes, held at 95% until 10 minutes, and then re-equilibrated for 2 minutes with 1 % mobile phase B (Table 4.1). Total UPLC-MS analysis time per sample was 12 minutes/ sample, per plate was approximately 19 hours. Samples were stored at 4°C in the auto-sampler throughout the analysis.

| Time (min) | Flow rate (µL/min) | A (%)<br>Water + 0.1% Formic acid | B (%)<br>Acetonitrile + 0.1% Formic acid |
|------------|--------------------|-----------------------------------|------------------------------------------|
| 0          | 500                | 99                                | 1                                        |
| 1          | 500                | 99                                | 1                                        |
| 3          | 500                | 85                                | 15                                       |
| 6          | 500                | 50                                | 50                                       |
| 9          | 500                | 5                                 | 95                                       |
| 10         | 500                | 5                                 | 95                                       |
| 10.1       | 500                | 99                                | 1                                        |

**Table 4.1 Gradient elution with Mobile phase A (Water + 0.1% formic acid) and Mobile phase B (Acetonitrile + 0.1% formic acid), total run time 12 minutes per sample (including 2 minutes re-equilibration time).**

The MS was calibrated with 0.1mg/mL sodium formate solution and checked for mass accuracy by infusing 200 pg/μl leucine enkephalin solution into the instrument. The MS was set to measure ions between 60-1000 m/z in both positive and negative ionisation mode to enhance metabolite coverage and detectability. Other parameters of the positive mode MS ionisation were: capillary voltage 3.20kV, cone voltage 27, ESI source temperature 150°C, desolvation gas temperature 600 °C and collision energy 2 eV. Parameters used in the negative ionisation mode were: capillary voltage 2.5kv, cone voltage 25v and collision energy 6v.

#### **4.2.4 LC-MS metabolite identification**

Metabolite identification was based on linking the retention time and masses of the previously generated chemical reference standard library to mass spectra peaks to identify individual metabolites and determine their peak areas (relative quantities).

The reason to use this approach was to give more confidence in the set of urinary metabolites (covering different metabolite classes) considered in LC-MS metabolic profiling and to facilitate future comparison with other metabolomics platforms, e.g. NMR spectroscopy. A small set of metabolites covering different metabolite classes could still provide a useful representation of the LC-MS metabolome and evaluate sources of their variation as a function of analytical and temporal parameters. Furthermore, this approach was convenient and practical as the Waters software MarkerLynx module TargetLynx used to build the chemical reference standard library is integrated with the MS instrument and automated the process of linking chemical standards with mass spectra peaks. This method is far quicker and more reliable than the application of different retention time alignment, peak detection and peak matching software.

It was not within the scope of the current work to compare and provide the details of the steps involved in the peak picking software such as XCMS, MZmine, Metalign etc. However, the choice and application of any of this softwares can have varied effects on the outcome of the results. Figure 4.3 shows the steps that were followed to generate chemical reference standard library and identify metabolites in mass spectra of samples analysed within the HELIX exposome project.

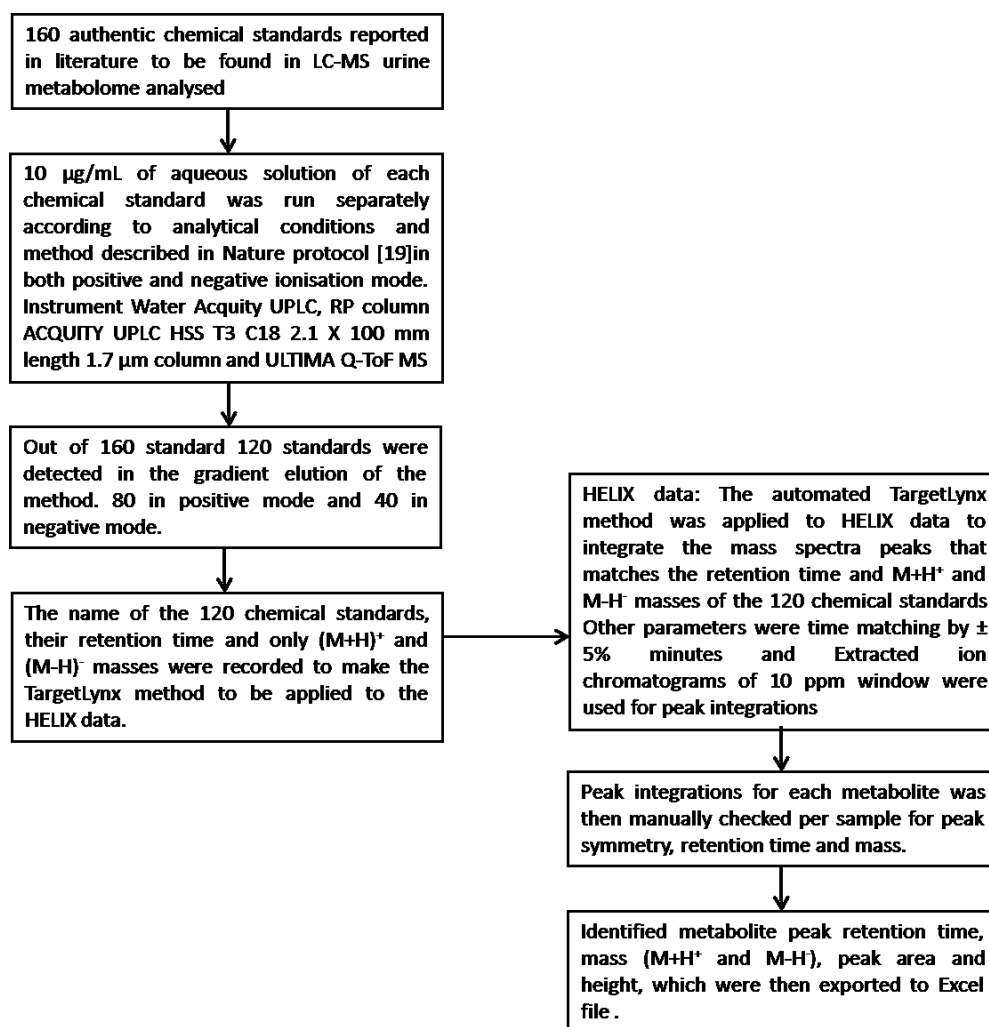


**Figure 4.3 Schematic diagram showing various steps followed to identify metabolites in untargeted LC-MS metabolic profiling**

In a separate set of experiments, 160 authentic chemical standards were analysed. These compounds were expected to be found in the metabolic composition of urine detected by various LC-MS methodologies according to the published literature and online human urine database [39, 40]. A 10 µg/mL aqueous solution of each of the 160 chemical standards were run separately according to the same LC-MS method and analytical conditions described above in both positive and negative ionisation mode. Out of 160 chemical standards analysed only 120 standards: 80 in positive mode and 40 in negative mode (Appendix Table 6) were detected in the gradient elution of the method. These standards were grouped into eight chemical classes: amino acids, acylcarnitines, sugars, lipids, nucleotide metabolites, organic acids, bile acids and peptides. Their retention time and mass of only (M+H)<sup>+</sup> and (M-H)<sup>-</sup> were searched and noted. Initially, (and for the work presented in this chapter) masses of only protonated and deprotonated compounds were searched.

This experiment was an initial attempt to generate a chemical reference standard library of metabolites which are detected in urine using the Want *et al.* LC-MS method. It facilitated the identification of metabolites and their peak areas in actual urine samples analysed in the present study. The identification of the metabolite in a sample used two parameters, retention time and mass ( $M+H$ )<sup>+</sup> and ( $M-H$ )<sup>-</sup> that matched to those of a chemical standard analysed under identical analytical conditions.

Waters MassLynx software module TargetLynx was used to create an automated method by incorporating 120 chemical standards names, retention time and mass ( $m/z$ ). Extracted ion chromatograms of 10 ppm window were used for peak integration. Only compounds eluting after 1 minute and before 9 minutes of the gradient elution were considered. The blanks, QCs and urine samples were then processed using this automated method per batch. The metabolite peak integrations were then manually checked per sample for peak symmetry, retention time and mass. The processed raw data provided integrated metabolite peak retention time, mass, peak area and height, which were then exported to Excel file for subsequent data processing and analysis.

Importantly, there was a difference in MS instruments used to run 160 authentic chemical standards (ULTIMA Q-ToF) described above, and the MS used to capture untargeted metabolic profiling of the actual test samples (Waters Xevo G2-S Q-ToF) analysed in the present study. However, the LC system and column used were same for both set experiments. The use of separate instruments was due to unavailability of the Waters Xevo G2-S Q-ToF at the time of analysing authentic chemical standards. Ideally, in good practice, both sets of experiments should have been conducted using same MS instrument. The use of different MS instruments can impact the findings as a parameter such as an adduct formation in the source of different instruments cannot be controlled. Therefore, initially only retention time and mass of only ( $M+H$ )<sup>+</sup> and ( $M-H$ )<sup>-</sup> were searched and noted.

#### **4.2.5 Data normalisation for analytical and biological variability**

The peak area data from the metabolites that were identified and present in 100% of QC and urine samples were pre-treated to account for biases introduced due to analytical and biological variations. The data normalisation/correction consisted of three steps as summarised in Figure 4.4: Step 1 was to correct for analytical variation caused due to batch effect using QC-RLSC: quality control based LOESS (locally estimated scatterplot smoothing) signal correction [38]. Step 2 was to correct for biological variation introduced due to variation in urine volume by expressing peak area as ratios relative to creatinine (Note: Creatinine was not among the annotated metabolites in the LC-MS data, its concentrations for each urine sample was taken from the <sup>1</sup>H NMR spectral

data acquired on the same QC and urine samples (Dr Lea Matire & Dr Chung Ho Lau, personal communication). In step 3 due to the non-normal distribution of these data, it was  $\log_e$  transformed to meet the assumption of normal distribution. The rationale behind normalising the data before calculating the analytical and biological variations was to reflect the actual urinary metabolite variance in normal healthy children and to exclude the possibility that the observed variance was due to systematic inconsistency from batch to batch, or due to urine volume effect or non-normal distribution of the data.

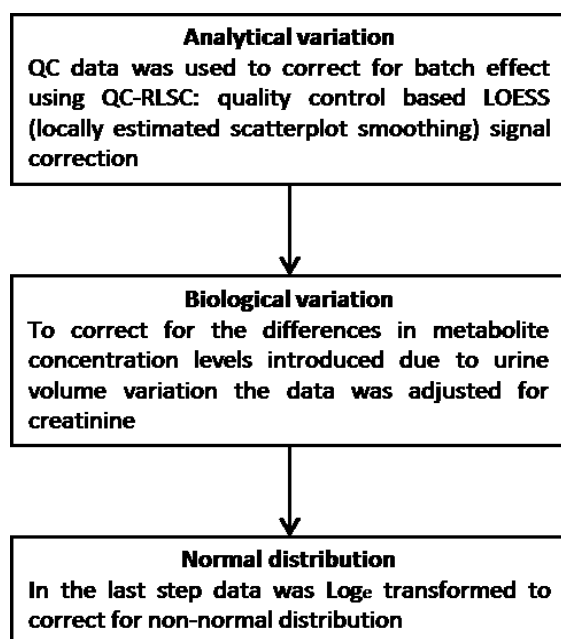


**Figure 4.4 Schematic diagram showing data pre-treatment steps followed to normalise analytical and biological variations**

Changes in the signal intensity over time are considered one of the main analytical factors that could introduce apparent variability in the finally reported metabolites levels [37, 38]. It is considered possible that the analysis of a large number of samples (N=556) and necessity to capture the metabolic profile of the samples in several batches (6 plates/batches) would introduce signal intensity drift which consequently introduces batch effects. To evaluate this 51 pooled QCs spread across six plates/batches were used to assess the batch effect on the metabolites detected via positive mode LC-MS metabolic profiling. The peak areas of 17 metabolites assigned to the positive ionisation mode data were used for this purpose. Inter-batch CV% was calculated on the raw (peak areas without any normalisation) and normalised data where peak area of each metabolite was corrected to the QC-RLSC. The rationale behind calculating precision (CV%) on non-normalised data was to assess the systematic error introduced by time and batch-to-batch variations and to compare the performance of QC-RLSC before assessing any specific metabolite level variability.

### **QC-RLSC: quality control based LOESS (locally estimated scatterplot smoothing) signal correction**

To eliminate the biases in signal intensity drift influenced by the run order and batching of the analysis a univariate approach called QC-RLSC, quality control based- LOESS signal correction, was applied. Application of the method is shown to minimise any attenuation in the peak response due to the order of the injections [41]. QC-RSLC, as explained by Dunn et al. [38], is a form of LOESS (Locally Weighted Scatterplot Smoother) [42] algorithm which generates a trend line for the QC samples. In the QC-RSLC method, the changes observed in the signal (in the present study it was peak area) of specific metabolites in bracketing QC samples (QC samples bracketing every ten urine subject samples) was used to correct temporal signal drift in the same metabolite in the subject urine samples.

The details of the QC-RSLC are explained by Dunn *et al.* [38]. In brief LOESS correction curve uses a flexible combination of weighted quadratic least squares regression with the nonlinear regression modelled by a set of local polynomials. Weighted least squares used to fit polynomials is based on the concept that the data points closer to QC data (estimation point) best describe the local model hence influence it, and data points far from the QC data least influence the overall local model parameters. The outcome of the model describes point by point the deterministic part of the data variation and avoids the fact that data follows any one global model hence improving its sensitivity. A parameter termed as 'smoothing parameter' ( $\alpha$ ) is used to calculate the regression curves and to determine how much data is needed to be fit in each polynomial. The correction curve for the whole analytical run is normalised using cubic spline normalisation after LOESS curve fits the QC data.

A MATLAB script written by Dr Gregory Tredwell was used to apply the QC-RLSC normalisation on the QC and urine samples. SIMCA-P+ version 12 was used to visualise the degree of reproducibility and reliability of the QC data from six plates/batches and performance of the QC-RSLC using PCA.

#### **4.2.6 Temporal biological variability**

The data from only those batches which passed the QC acceptance criteria were used to calculate the temporal biological variations. QC-RLSC normalised, creatinine-adjusted and  $\log_e$  transformed metabolites peak area data were used to calculate the intraclass correlation coefficient (ICC) on data for 17 metabolites in urine acquired in positive ionisation mode. The calculate ICC R script (R Project, version 3.1.3) used was written by Dr Pooja Jain. The Intraclass correlation coefficient (ICC) [43] is a quantitative measure of the reproducibility or reliability of the measurements. In this study, estimates of ICC were calculated to quantify how much

consistency or reproducibility there are the measurements of the metabolite profiles within and between individuals measured over time. However, ICC is a relative measure of biological variability- it depends on the variability in the samples as it quantifies the total variance that is the result of differences between subjects.

The ICC score is defined by score ranges of 0 to 1. The closer the score is to 1, better the consistency or reproducibility of the measurements. Estimates of ICC were computed as the ratio of the between-subject variance over the total variability (attributed to the sum of the square of between and square of within variance).

$$ICC = \frac{(between\ individual\ variance)^2}{(between\ individual\ variance)^2 + (within\ individual\ variance)^2}$$

The individual variances (between and within) were calculated by one-way ANOVA.

## 4.3 Results

### 4.3.1 Metabolites identified

Out of 120 chemical reference standards that were used to link the retention time and  $(M+H)^+$  and  $(M-H)^-$  masses of the chemical standards with the mass spectra peaks, a total of 33 metabolites were identified in 100% of the urine and the QC samples. Out of which ten metabolites were identified only in positive ionisation mode data, 16 were identified only in negative ionisation mode data, and seven were found in both positive and negative mode data (Figure 4.5). These identified metabolites were from 5 chemical classes: amino acids, acylcarnitines, carbohydrates, organic acids and nucleotide metabolites. Peak integration of these metabolites were manually checked per sample for retention time, peak shape and mass.

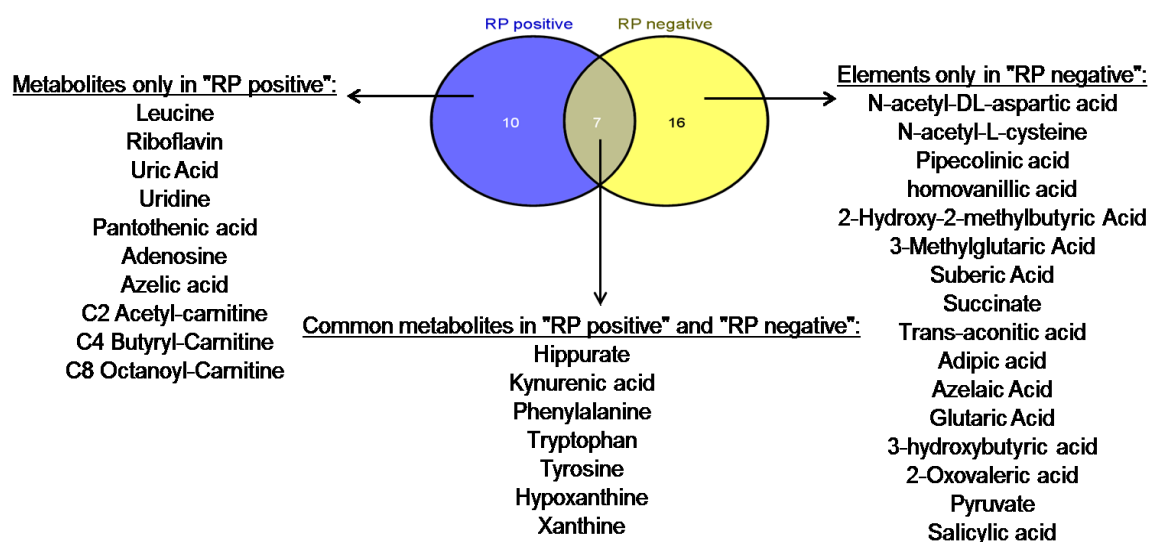


**Figure 4.5** List of metabolites assigned in positive and negative ionisation mode urine data of N=20 healthy children and QC samples. RP positive and RP negative refer to reverse phase positive ionisation mode and reversed phase negative ionisation mode data respectively. The experimental and temporal biological variability was explored using urinary metabolites assigned only to positive ionisation mode data.

The metabolite identification work is on-going with the aim of expanding the number of metabolites that are detected based on the retention time and  $m/z$  match to authentic chemical standard so that the data analysis can be semi-targeted in future analyses. However, in untargeted metabolic profiling, the possibility of an ambiguous identification can be likely especially for stereoisomers. Stereoisomers are likely to have very similar retention times and have the same mass. Therefore, optimised chromatographic and targeted MS methods are required to differentiate and identify these isomers.

### 4.3.2 LC-MS analytical variation

The inter-batch precision (across six plates) of the raw data (peak areas) for the majority of the metabolites was  $>20\%$  including four metabolites with  $CV\% >30\%$  (Figure 4.6). Although according to the US FDA guidelines for the biomarker discovery the acceptance criteria for  $CV\%$  are set up to  $30\%$ , it is suggested that for untargeted metabolic profiling the acceptance tolerance should be aimed at  $20\%$  [33]. After the data normalisation (Figure 4.7), the inter-batch precision (across six plates)  $CV\%$  for most of the metabolites was reduced to  $<20\%$ , but for the three metabolites the  $CV\%$  remain unchanged or increased to be more than  $30\%$ .

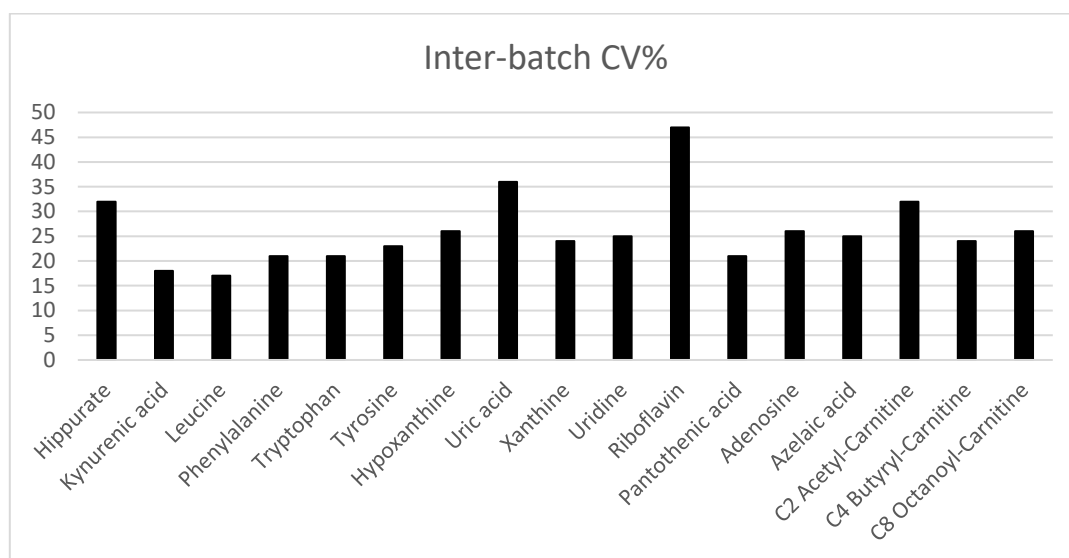


**Figure 4.6 Inter-batch precision (CV %) of the peak areas of the 17 metabolites assigned in positive ionisation mode data from 51 pooled QCs across six plates.**

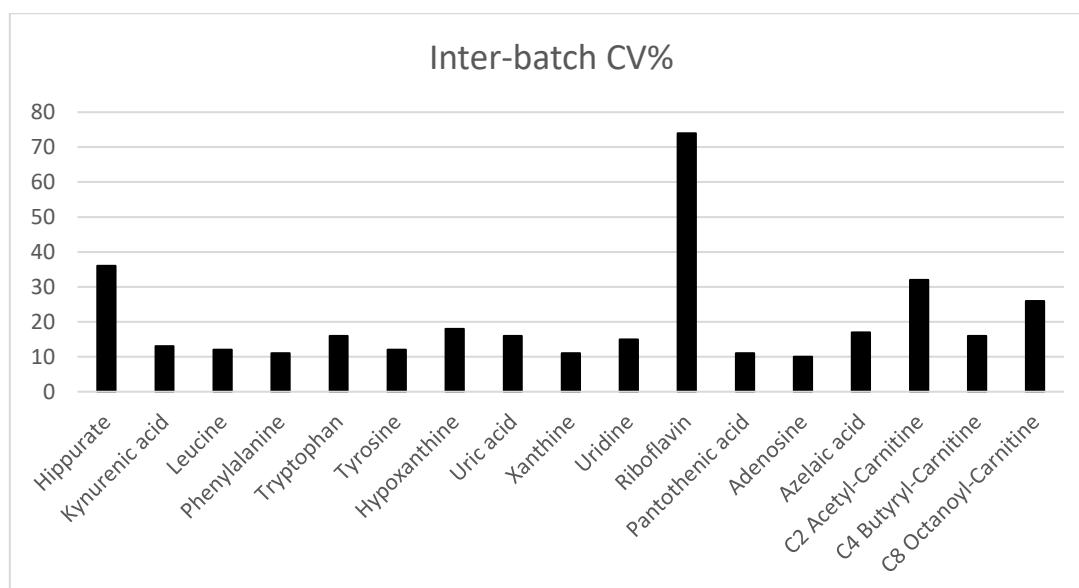


**Figure 4.7 Inter-batch precision (CV%) of the normalised peak area data (QC-RLSC, creatinine-adjusted and log transformed) of the 17 metabolites assigned in positive ionisation mode data from 51 pooled QCs across six plates.**

This suggested that although normalisation reduced the CV% for the majority of the metabolites, it failed to correct for high variation observed in metabolites where the CV% was more than 30% in raw data. To further examine the QC data quality, PCA scores plots were constructed using raw and normalised data (360 urine samples and 51 QC samples) from six batches in the positive ionisation mode as shown in Figures 4.8-4.9. The PCA scores plot in Figure 4.8 shows a visualisation of the degree of distribution of QC samples using peak areas of 17 metabolites. The PC 1 ( $t[1]$ ) accounts for most of the biological variability, and PC 2 ( $t[2]$ ) accounts for a batch to batch variation. The red circles represent 51 pooled QC samples and grey circles represent 360 urine samples. As it can be seen, the QC samples do not show tight clustering and spread across in the PC1 and PC2, data indicating the influence of both batch to batch variation and biological variation.

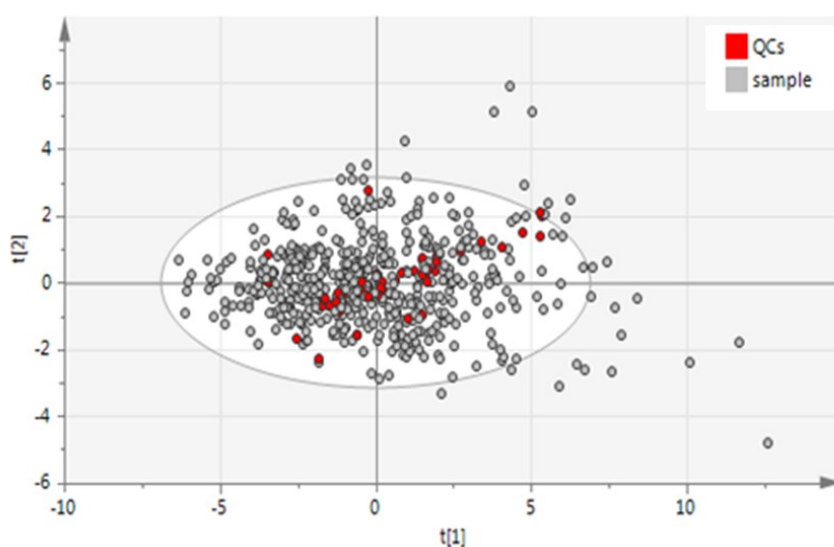


**Figure 4.8** PCA scores plot is showing the distribution of 51 pooled QCs (red circles) among 360 children urine samples (grey circles) from six batches. Raw data, i.e. peak areas of 17 metabolites were used in this analysis.

Considering that the 51 QC samples were identical in composition, their tight clustering in the PCA scores plot would indicate good reproducibility and reliability of the metabolites. To further monitor the QCs, PCA scores plot on a reduced dataset comprising only QCs from six batches were examined. This allowed evaluating the relation or closeness of the QCs from one batch to those obtained from other five batches. As seen in Figure 4.9 the QCs from batch 1, 2, 3, 4, and 6 show tight clustering indicating good intra and inter-batch reproducibility. However, 6 QCs from the batch 5 were not tightly clustered in PC 1 and PC 2 indicating increases variability in both experimental and biological source. Furthermore, it also indicates that the QC-RSLC normalisation was not a sufficient or appropriate method to correct for this increased experimental variability.

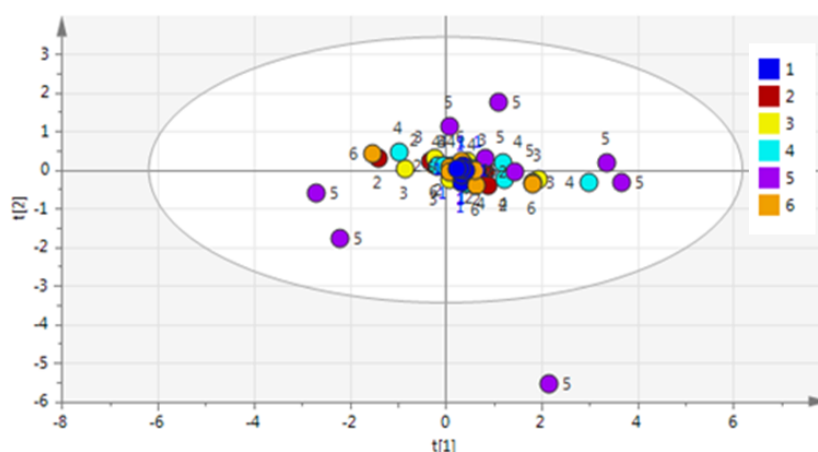


**Figure 4.9 PCA scores plot showing clustering of 51 pooled QCs from six separate batches after normalisation. Colour code represents batches 1 to 6.**

The finding shows that the measurements of the QCs from batch 1,2,3,4 and 6 were more reproducible as they overlap. While the QC data obtained from batch 5 shows unexplained experimental and biological variability. In Figure 4.9 two QCs from batch 5 cluster and overlap with QCs from other batches. However, the remaining 7 QCs from batch 5 are much further apart. This trend cannot be explained solely by the systematic error introduced by batch to batch variation since QCs from batch six does not show any increased experimental variability and overlap with QCs from batch 1, 2,3 and 4. Furthermore, the spread observed in the QCs of batch five did not indicate run order effect within the batch, the QC observed most extreme was third in the QC order list. QC variance in batch 5 raised concern about the data quality of the children urine samples analysed in batch 5.

The most likely explanation of the variability observed in QCs from batch five was due to the manual error made by placing the samples in the 96-well plate. Mismatch of the samples placed in the 96-well plate in batch 5 and run order list registered in UPLC-MS system would not only introduce error in the QC samples but the entire run list for the test samples as well. The variations observed in QCs of batch 5 raised an implication of unreliable data quality for the urine samples within this batch and was excluded from any subsequent analysis.

These results also implied that variation observed in the QC data from batch 5, would influence the inter-day precision shown earlier in Figure 4.6 and Figure 4.7. For this reason, the inter-day precision (CV%) was re-calculated using QC data from batches 1, 2,3, 4 and six excluding QCs from batch 5. The inter-batch precision (45 QCs across five plates) of the raw data (peak areas) for all of the 17 metabolites was <30% (Figure 4.10). After the data normalisation (Figure 4.11), the inter-batch precision (45 QCs across five plates) CV% for all the metabolites reduced to be <10%.

After excluding QCs from batch five the inter-batch precision (CV%) of both raw and normalised data (Figure 4.10 and Figure 4.11) shows good reproducibility and reliability of the metabolites detected using LC-MS metabolic profiling. These results suggest that despite the inevitable experimental variability introduced due to various reasons such as instrument and sample handling, the untargeted LC-MS metabolic profiling of the repeated samples is reliable and stable over continuously extended run. It also demonstrated the importance of QCs in untargeted LC-MS metabolic profiling for correcting systematic signal drift in intra-batch and inter-batch to batch variation using QC-RLSC normalisation.

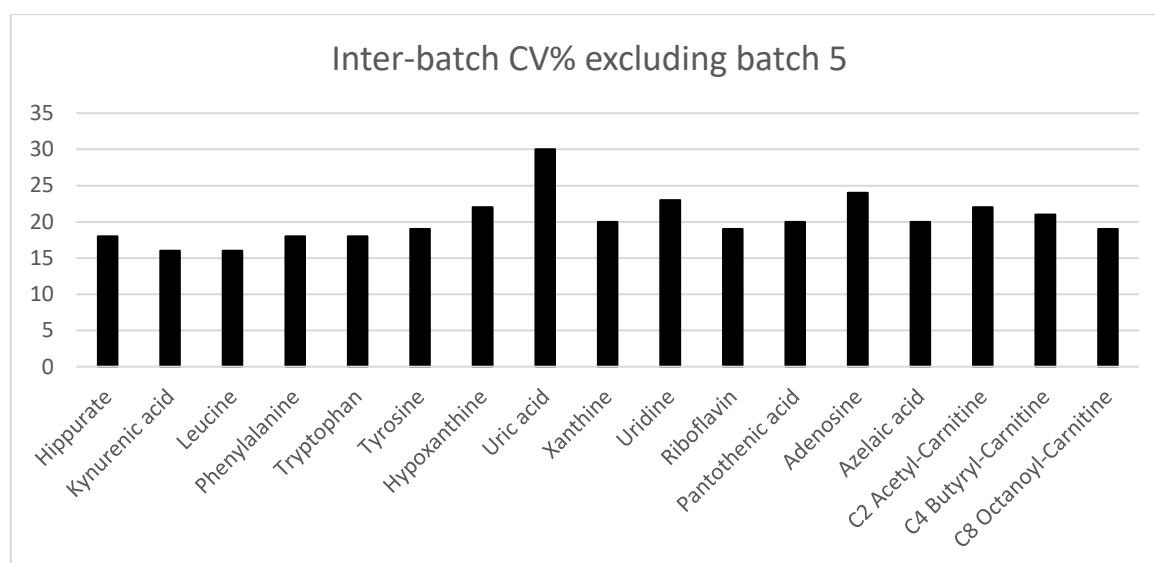


**Figure 4.10** Inter-batch precision (CV%) of the raw peak areas of the 17 metabolites identified in positive ionisation mode data from 45 pooled QCs across five plates (batch 1, 2,3, 4 and 6), excluding QCs from batch 5.

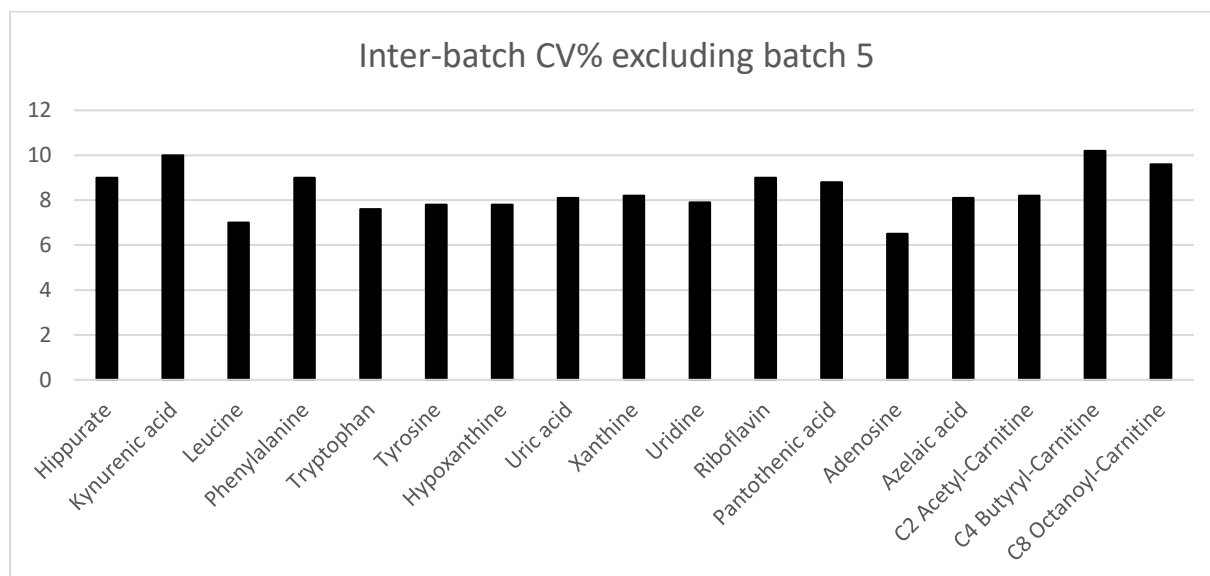


**Figure 4.11** Inter-batch precision (CV%) of the normalised peak areas of the 17 metabolites identified in positive ionisation mode 45 pooled QCs across five plates (batch 1, 2,3, 4 and 6), excluding QCs from batch 5.



0.78), riboflavin (ICC: 0.58, 95% CI 0.18 to 0.66) and adenosine (ICC: 0.56, 95% CI 0.06 to 0.67). The daily pools ICCs for leucine, uric acid, uridine and C4 Butyryl carnitine were close to 0.5. This suggests that daily pool urines were more reproducible and a better representative of the metabolite profile over one week.

Low ICCs (<0.5) indicated high variability in the metabolite levels between individuals over the short longitudinal time frame (6 days). The within and between variance components of the ICC are also reported for each metabolite. As seen in Table 4.2 the within-individual variance is higher compared to the between individual variance indicating that majority of the biological variance was explained by intra-individual variability. Given that in most large population studies a single spot sample is all that is available per individual the power of the study to detect any association of metabolites with exposure or disease increases when intra-individual variability is low and greater proportion of total variability is attributed to inter-individual differences.

|                       | <b>Morning ICC (95% CI)</b> | <b>Night ICC (95% CI)</b> | <b>Daily pool ICC (95% CI)</b> |
|-----------------------|-----------------------------|---------------------------|--------------------------------|
| Hippurate             | 0.28 (0.18 to 0.38)         | 0.18 (0.03 to 0.19)       | 0.37 (0.06 to 0.47)            |
| Within                | 0.19                        | 0.19                      | 0.09                           |
| Between               | 0.05                        | 0.09                      | 0.07                           |
| Kynurenic acid        | 0.25 (0.03 to 0.33)         | 0.25 (0.06 to 0.45)       | 0.24 (0.00 to 0.37)            |
| Within                | 0.05                        | 0.05                      | 0.05                           |
| Between               | 0.04                        | 0.03                      | 0.03                           |
| Leucine               | 0.28 (0.04 to 0.44)         | 0.24 (0.05 to 0.28)       | 0.47 (0.05 to 0.68)            |
| Within                | 0.02                        | 0.03                      | 0.02                           |
| Between               | 0.02                        | 0.00                      | 0.06                           |
| Phenylalanine         | 0.31 (0.01 to 0.39)         | 0.29 (0.03 to 0.41)       | 0.35 (0.14 to 0.56)            |
| Within                | 0.01                        | 0.02                      | 0.02                           |
| Between               | 0.01                        | 0.01                      | 0.04                           |
| Tryptophan            | 0.36 (0.15 to 0.58)         | 0.31 (0.11 to 0.52)       | <b>0.61 (0.33 to 0.66)</b>     |
| Within                | 0.03                        | 0.03                      | 0.01                           |
| Between               | 0.02                        | 0.02                      | 0.03                           |
| Tyrosine              | 0.34 (0.06 to 0.47)         | 0.27 (0.07 to 0.48)       | <b>0.69 (0.41 to 0.78)</b>     |
| Within                | 0.04                        | 0.06                      | 0.03                           |
| Between               | 0.02                        | 0.03                      | 0.03                           |
| Hypoxanthine          | 0.12 (0.05 to 0.29)         | 0.14 (0.02 to 0.33)       | 0.36 (0.08 to 0.49)            |
| Within                | 0.17                        | 0.24                      | 0.12                           |
| Between               | 0.02                        | 0.00                      | 0.05                           |
| Uric.acid             | 0.38 (0.01 to 0.48)         | 0.20 (0.01 to 0.36)       | 0.42 (0.04 to 0.53)            |
| Within                | 0.20                        | 0.18                      | 0.12                           |
| Between               | 0.04                        | 0.04                      | 0.04                           |
| Xanthine              | 0.20 (0.07 to 0.24)         | 0.15 (0.02 to 0.33)       | 0.30 (0.03 to 0.37)            |
| Within                | 0.09                        | 0.09                      | 0.06                           |
| Between               | 0.01                        | 0.02                      | 0.01                           |
| Uridine               | 0.20 (0.09 to 0.25)         | 0.13 (0.06 to 0.27)       | 0.44 (0.09 to 0.51)            |
| Within                | 0.06                        | 0.08                      | 0.03                           |
| Between               | 0.01                        | 0.01                      | 0.01                           |
| Riboflavin            | 0.51 (0.10 to 0.60)         | 0.28 (0.01 to 0.36)       | <b>0.58 (0.30 to 0.66)</b>     |
| Within                | 0.64                        | 0.69                      | 0.44                           |
| Between               | 0.28                        | 0.15                      | 0.31                           |
| Pantothenic.acid      | 0.17 (0.07 to 0.25)         | 0.13 (0.07 to 0.25)       | 0.24 (0.01 to 0.35)            |
| Within                | 0.04                        | 0.06                      | 0.04                           |
| Between               | 0.00                        | 0.01                      | 0.00                           |
| Adenosine             | 0.42 (0.14 to 0.56)         | 0.15 (0.09 to 0.29)       | <b>0.56 (0.06 to 0.67)</b>     |
| Within                | 0.02                        | 0.05                      | 0.04                           |
| Between               | 0.02                        | 0.00                      | 0.01                           |
| Azelic acid           | 0.11 (0.06 to 0.27)         | 0.13 (0.10 to 0.16)       | 0.18 (0.16 to 0.54)            |
| Within                | 0.31                        | 0.30                      | 0.17                           |
| Between               | 0.04                        | 0.01                      | 0.01                           |
| C2.Acetyl.carnitine.  | 0.29 (0.05 to 0.46)         | 0.24 (0.02 to 0.34)       | 0.38 (0.03 to 0.44)            |
| Within                | 0.32                        | 0.29                      | 0.22                           |
| Between               | 0.11                        | 0.06                      | 0.04                           |
| C4 Butyryl.carnitine  | 0.32 (0.03 to 0.42)         | 0.20 (0.06 to 0.26)       | 0.49 (0.09 to 0.56)            |
| Within                | 0.05                        | 0.05                      | 0.03                           |
| Between               | 0.01                        | 0.01                      | 0.01                           |
| C8 Octanoyl.carnitine | 0.35 (0.08 to 0.49)         | 0.23 (0.01 to 0.38)       | 0.42 (0.02 to 0.53)            |
| Within                | 0.12                        | 0.25                      | 0.20                           |
| Between               | 0.23                        | 0.19                      | 0.20                           |

**Table 4.2 ICCs (95% confidence interval) for 17 metabolites in the morning, night and daily pool samples of 9 children. The within and between values represent within the individual and between individual variances.**

To further examine the intra-individual variation of metabolites as a function of the time, a line graph showing temporal variation of tyrosine and tryptophan levels in all of 20 children across six days was shown in the morning, night and daily pool samples (Figure 4.13). Tyrosine and tryptophan were chosen as an example to show temporal metabolite level trends as these had better ICCs compared to all other metabolites. Considering daily pool samples is 50:50 mixture of morning and night urine. Levels and longitudinal temporal trend of daily pool samples should lie between morning and night urine. However, the results in Figure 4.13, does not support this hypothesis. This data shows high fluctuation and variability in metabolite levels within all 20 children across six days for the morning, night and daily pool samples.

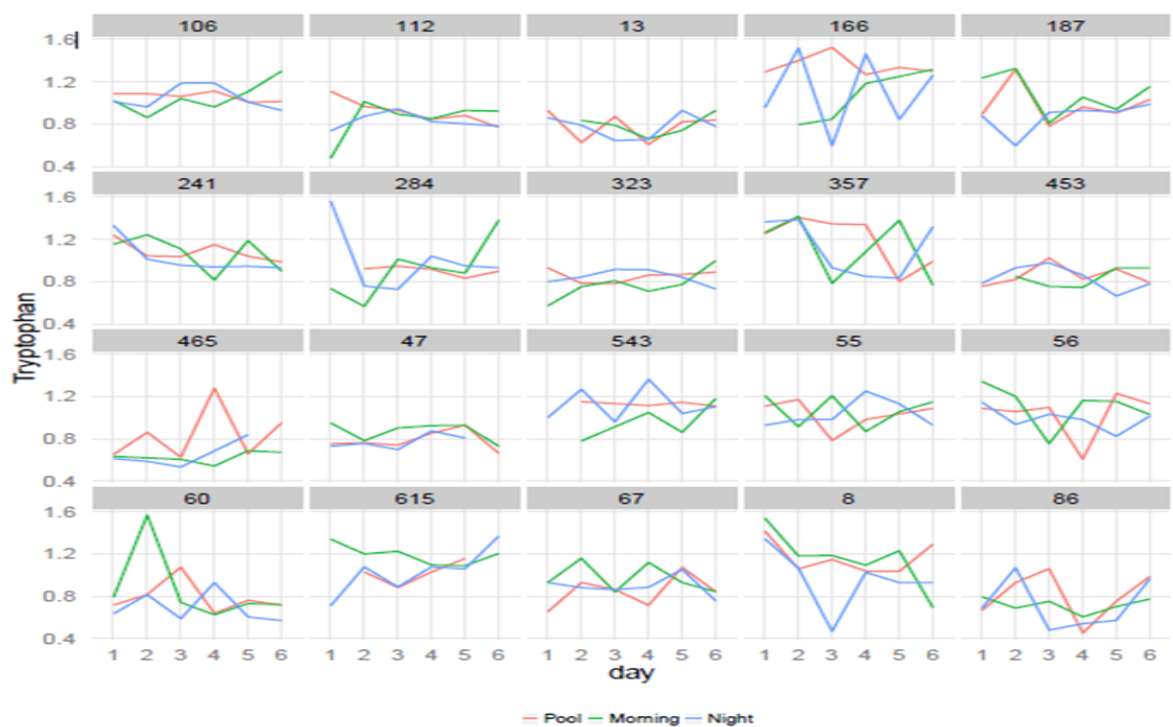
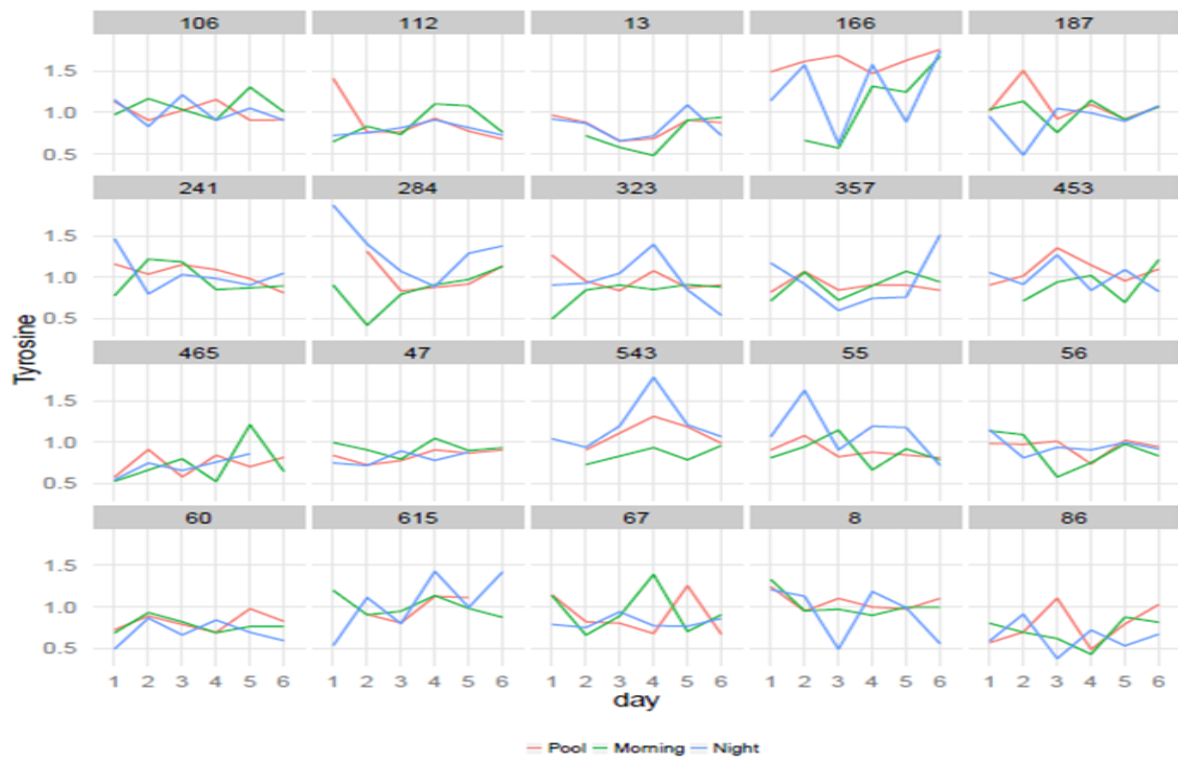


**Figure 4.13** Line graphs are showing intra-individual variation in levels of tyrosine and tryptophan across six days in N=20 children urine samples. Green, blue and red lines represent morning, night and daily pool samples respectively. A number on the top of each graph is arbitrary for 20 children.

## 4.4 Discussion

The present work was a pilot study to assess the LC-MS metabolome short-term temporal variability in children (age 7 to 9 years) urine samples with specific focus on implications of the results for designing future molecular epidemiology and HELIX exposome research. The current study has also extended the limited pre-existing knowledge of the sources of variation in the urine of healthy children. Furthermore, it is also the first study to explore the reliability and reproducibility of pooled diurnal samples compared to single spot urine from one single time point per individual.

The observed variability in the data was explained by analytical and short-term (6 days) intra and inter-individual temporal biological variations. The 17 metabolites examined in this study were all major metabolites ranging over four metabolite classes (amino acids, acylcarnitines, organic acids and nucleotide metabolites. Their identification was based on linking the mass spectral peaks with retention time and  $(M+H)^+$  masses of the chemical reference standards and MS/MS. Future studies using Want *et al.* [37] method and same LC-MS system/conditions (as used in the present study) to acquire urinary metabolite profiles can benefit from the reference chemical standard library. For future studies, it will facilitate quick and efficient identification of main metabolite peaks, assess the data quality of urine samples and estimate the degree of analytical and temporal variations.

### Analytical variation

The aim of evaluating the analytical variability of the 17 urine metabolites detected by LC-MS was to determine the reproducibility and reliability of the analytical method (LC-MS system) in continuous use over the period of one full week. The length of the run is considered both feasible and substantial in large population-metabolic profiling studies. The objective was to evaluate the degree of systematic error introduced as a result of analytical system deterioration, which could be misinterpreted as of biological significance. According to the US FDA and EMEA guidelines large-scale LC-MS metabolomics studies with analytical reproducibility of CV% <10% is deemed very good and provides confidence in the analytical method and the quality of the data [44, 45]. At first glance, the raw data (non-normalised) results from QCs across six batches (54 QCs) showed that the systematic inconsistency from batch-to-batch was responsible for around CV% =25% to 40% in the metabolites measurement. The QC data was then used to minimise this systematic drift using QC-RLSC normalisation, and the results showed that even after QC-RLSC normalisation QCs from batch five were driving most of the variation and were excluded from the data analysis due to unacceptable data quality. The post-data correction was applied, and the

analytical reproducibility of 17 urinary metabolites assessed in 100% of QC samples (45 QCs) was high for the five batches 1, 2, 3, 4, and 6 with CV%  $\leq 10$

The variability observed in the QC data from batch five cannot be explicitly pinpointed or solely attributed to systematic signal drift, since QC data from batch six did not show increased variation. Further, there was no drift in retention time or symmetry of peaks that were observed in the QCs from batch 5. This illustrates as an example of the error in pre-analytical steps that could be introduced either via manual mishandling, misplacement of the QC samples on the plate or mismatch of placements of test samples and QCs on plate 5 with the run order with the sample list in the MS. Despite the use of the state-of-the-art UPLC-MS system and bare minimum sample preparation as used in this study, relatively large proportion of variability was observed in QCs of batch 5 (in relation to batch 1, 2, 3, 4 and 6). This emphasises the need and importance to optimise pre-analytical steps required to minimise this source of variation. It also highlights that errors introduced during experimental stages (mismatch of the order of the samples on the plate and run order list) cannot be corrected by even the most advanced analytical technique or statistical methods.

The results also clearly illustrated that the importance of QCs in untargeted metabolic profiling. The present study showed that in a high throughput experiments where continuous consecutive batches are analysed, the performance of the evenly distributed QC samples per batch could be efficiently monitored and analysed to assess the robustness of the analytical method. Any increased variability in the QC samples noticed just from visual examination via PCA, can be used to investigate the source of the variation. These sources of analytical variations could range from ionisation source cleaning/maintenance, change in room temperature, to any error introduced via manual handling of the samples on the plates.

Other studies which investigated the reproducibility and robustness of the untargeted LC-MS metabolic profiling methods for human urine analysis have reported inconsistent results. One of these studies investigating the experimental reproducibility for the 10 QC replicates reported ICCs for 72% (N=207) of the detected metabolites to be more than 0.8 [20]. While another study, exploring the experimental variability of the urine LC-MS metabolome, reported the experimental/technical source as the biggest source of variation in metabolite measurements [35] with >50% of the accounted variability. It is worth pointing that in both of these studies the total number of the QC samples analysed were smaller (10 in former and 13 in a later study) than the sample size used in the present study. Another study which investigated the QC approach to determine the robustness of the urine untargeted LC-MS metabolic profiling method in five batches (each of 80 injections, 60 samples). The reported results showed good day-to-day QC reproducibility (Relative standard deviation <5%) for first three batches, while break of two days

between the analysis showed a negative effect on the reproducibility [26]. The same study also reported the instability of the LC-MS system after analysing five batches and need for maintenance and cleaning to restore the level of performance.

The use of a test mix solution (containing compounds of known composition and concentration) at the beginning and the end of the run along with the QC samples is also being recommended in untargeted LC-MS metabolic profiling methods [37]. This test mix solution is shown to be used to monitor LC-MS system stability and the quality of the separation of the peaks of known compounds. The use of test mix solution could be incorporated in the future analysis. While the LC-MS experimental/analytical variability may never be eliminated, paying attention to following parameters will definitely reduce experimental inconsistencies: careful sample handling, placement of the samples on the plate, monitoring sample storage condition and temperature, bench time of the samples, consistent methods following standard optimised procedures for sample collection and data pre-processing steps.

### **Temporal biological variation**

The two major limitations of calculating temporal biological variation within this study were: the temporal variability was assessed using data from only nine subjects over a short period (6 days). This may underpower this study to detect any inter-individual differences among metabolites. Furthermore, the temporal variability was calculated for an only small number of metabolites.

Studies have demonstrated that levels of certain metabolites such as melatonin, cortisol, vitamin D, hormones differ due to age, time of sample collection and temporal variability [21, 24, 46, 47]. A short-term variability in the levels of amino acids, lipids and carbohydrates were observed to increase in urine samples compared to blood and saliva after meal time and dietary intervention in an NMR analysis based study [23]. It is critical to characterise the normal level of variations in concentrations of metabolites influenced by different sources so that they are not misinterpreted as of biological relevance. For example, if metabolite concentration is influenced by sample collection time then special procedures will be needed to be followed to avoid any biased variability's other than that of biological relevance (e.g. due to disease state).

To the best of my knowledge, there are only three urinary LC-MS metabolic profiling studies which examined the longitudinal intra-individual and inter-individual variations each conducted by Xiao *et al.* [20], Sampson *et al.* [30] and Kim *et al.* [35]. All three studies reported relatively inconsistent results. Kim *et al.* [35] reported that most of the variability in their data >50% was driven by analytical/experimental sources. Sampson *et al.* [30] reported high analytical reliability (median ICC 0.8) but intra-individual variability over time was also high. This study accounted the majority of the variability (64%) in the metabolites levels due to intra-individual and

analytical functions. Xiao *et al.* [20] reported high analytical reproducibility (median ICC =0.92) and the majority of variation was attributed to inter-individual variation in 72% of the measured metabolites. Overall all three studies suggested that larger sample sizes are required to detect any urinary metabolite associations with moderate effect sizes to exposure or disease. For example, Xiao *et al.* [20] calculated the effect of sample size and concluded that a larger population study with more than 500 samples is required to detect 1% of metabolite association for a relative risk.

Within this study, though the ICCs for morning and night urine samples were low, the ICCs for the pooled diurnal samples were >0.5 for four metabolites, close to 0.5 in six metabolites and relatively better than morning and night samples. These results were the main key findings in the present study, clearly indicating the pooled diurnal samples represent more stable and reproducible part of the urinary LC-MS metabolome. These results have important implication for future large molecular epidemiology and exposome research studies specifically HELIX. As a result of the findings presented here and the related NMR-based study [36] analysis on the full HELIX, sub-cohort was applied to daily pooled urine samples.

The high intra-individual fluctuation of metabolite levels indicates high day to day variability. These results do corroborate with some previously reported work from Sampson *et al.* [30] and Kim *et al.* [35], where high intra-individual variations and low inter-individual variations were observed. However, direct quantitative comparisons between the present study and the results showed by Sampson *et al.* and Kim *et al.* are difficult to make due to differences in measured metabolites, study design and data analysis.

The results indicate that an LC-MS urinary metabolite profile from a single child is capable of varying hugely over a short period. Unlike blood, urine is not in constant circulation throughout the body. Its composition rather consists of a large number of waste biological by-products and metabolite excretion is a reflection of the vast number of metabolic processes that occur over time. Hence, the high intra-individual variability could perhaps be a representation of the normal biological function of urine. Kim *et al.* [35] observed higher instability in the urine samples compared to blood samples, and the effects were more pronounced by meal and sample collection time. Similarly, Nicholson *et al.* [31] have shown more variability in urine samples compared to blood, reporting 60% stability of blood <sup>1</sup>H NMR metabolome whereas only 47% stability in urine <sup>1</sup>H NMR metabolome. Since children in the present study were a homogenous, i.e. same gender, age and ethnic background most likely followed a similar diet; this could probably mean that larger sample size was required to see the more marked effect of larger inter-individual metabolite profile differences.

Given the high amount of variability observed in intra-individual urine samples analysed via untargeted LC-MS metabolic profiling, it is essential that due diligence is applied in context to some of the limitation of the technique about its robustness, reproducibility and reliability of the information. Some of the limiting factors of this analysis that potentially contributed to low ICCs are explored further below.

(1) Non-linear response of the assay: despite the precision of the analytical method for the identical QC samples  $CV \leq 10\%$ , the non-linear response of the detector can result in a substantial difference between the measured intensities and make an appropriate adjustment to correct for this error could be difficult to compute. Options of using pooled QC dilution series to check for non-linear matrix effects in randomised order in each batch have been suggested [48]. This method has been applied in untargeted LC-MS metabolic profiling studies post peak annotation; the relative concentration (peak height or area) was validated for linearity by establishing pooled QC dilution series or dilution series of the reference standards response curve for the corresponding peak [46, 47].

(2) Short time frame: in the majority of the previous studies which showed stable part of intra-individual urine metabolome acquired using LC-MS or NMR, the urine samples were collected over 30 days to few months to a couple of years [20, 55, 56]. In contrast to the present study, the temporal longitudinal sample collection spanned over six days. There is a possibility that for some metabolites the variability in their levels could be large when the time interval between sample collections is small. To confirm the stability of the urine LC-MS metabolome the measurement over few days may not represent the long-term normal level stability. The variability observed over short timescale cannot be implied as variability over longer time scale: in other words, variations observed six days might not persist for several months or years.

(3) Sample size: ICCs calculated on the data from only nine children are likely to be approximate. Studies which have shown the consistency instability of urine metabolome either using LC-MS or NMR platform have used a minimum of 60 subjects to hundreds of subjects. An NMR study by Saude *et al.* [15] observed the existence of the substantive amount of fluctuation and higher intra-individual variability relative to inter-individual population variation in some metabolites (Citrate and tyrosine) used six subjects in their analysis. Investigators of HUSERMET study [49] phenotyped 1200 individuals using untargeted LC-MS metabolic profiling to define normal baseline concentration in human serum suggested that minimum sample size of 600 is necessary to detect subtle inter-individual metabolite levels changes. Nonetheless, the data shown in this chapter inject a healthy dose of caution into the design of the future metabolic profiling studies and translation of any markers of interest into clinical settings.

The work done by Sampson *et al.* [30] which observed high intra-individual variability (40% of biological variability) in LC-MS serum metabolite profiles of 60 healthy women over 1 year discussed in his paper that '*metabolic profiling studies should not be discouraged by high intra-individual variability and should plan for such results to overcome imprecision of such magnitude*'. In the past studies, the top vs bottom quartile comparison in epidemiological studies has overcome the issue of low inter-individual variability and nonetheless went on to establish associations between metabolites and exposure levels and disease risk. E.g. insulin and estradiol with moderate inter-individual ICCs 0.58 and 0.68 respectively over the course of 1-3 years were associated with breast cancer in post-menopausal women [50, 51]. Top vs low cotinine level comparison was associated with lung cancers with an odds ratio of up to 20 [52]. Similarly, top and bottom quartile comparison between population studies established an association between branched-chain amino acids and risk of diabetes and heart disease with an odds ratio of 2 to 4 [53, 54]. Further ICCs of some of the well-established clinical markers such as serum cholesterol (ICC=0.65), blood glucose (ICC=0.52), and blood pressure (ICC= 0.60-0.64) are reported to be within moderate range for several years of reproducibility.

From the points discussed above, it can be concluded that low ICCs should not be considered as a limiting factor in establishing the association between metabolites levels, exposure levels and disease risk. However, to enhance the study power and to detect metabolite association with moderate size effects in case/control studies would require large sample sizes (upwards of 1000 subjects) as calculated in studies by Nicholson *et al.* [31], Sampson *et al.* [30] and Xiao *et al.* [20].

To summarise, some of the key points that can be implemented in future untargeted LC-MS metabolic profiling studies to enhance its study power, data quality and robustness are: (1) An attempt to reduce or minimise the experimental variability should be made by optimising sample preparation, sample handling, their placements of the plates, double checking run list order, analysing QC and test mix data batch by batch as the samples are running. (2) To use pooled QC dilution series to check for non-linear matrix effects of the assay. (3) Sample size: Increase the sample size to detected larger-inter-individual differences and subsequently ICCs. (4). Diurnal daily pool samples per individual should be analysed to increase the statistical power of detection of a biologically significant variation.

## **4.5 Strengths and limitations**

The main strength of the present work was its study design. It is the first study which used children urine samples which are central to HELIX exposome project to assess short-term intra- and inter-individual temporal variations over the course of six days using LC-MS platform. It is also the first study which evaluated the stability and reliability of daily pooled diurnal samples in

comparison to the stability of specific time points (morning or night) per subject. It also allowed evaluating the analytical variability in metabolites levels that need to be taken into account for any variation that is not of biological significance. Take together all the above points will help in effective study design of large molecular epidemiology and exposome studies.

In addition to points already discussed (limited sample size and study power, short duration of the study) the work had further limitations. The temporal variability was calculated for an only small number of metabolites. However, these metabolites were spread across four different metabolite classes. The analytical and temporal variability was calculated using two different statistical methods and rather than model both sources together. To get the estimate of the total variances (analytical + temporal) and the impact of the individual or combination of variances a different approach of data analysis could have been employed, such as decomposition of variances [20].

## **4.6 Conclusion**

The present work was a pilot study designed and conducted specifically in the context of HELIX exposome project to assess short-term temporal variability in the LC-MS metabolome of urines from healthy children. Two different sources of metabolite level variations were assessed one was analytical variability and second was short-term temporal biological variability. To realise the full potential of LC-MS metabolic profiling and exposome research, it is essential to define variability caused by such sources and identify normal baseline, so they cannot be misinterpreted as a biological significance.

Overall analytical reproducibility in identical QC samples (45 QCs across five batches) was high  $CV\% \leq 10\%$  for all the 17 metabolites examined in this study. The results demonstrated the importance of QC samples in untargeted LC-MS metabolic profiling to investigate the experimental robustness of the analytical method, to use QC data to exclude samples that were not deemed acceptable by data quality and to correct for systematic error introduced during analysis. Despite high analytical precision, the data demonstrated large inter- and intra-individual variation in metabolite levels (ICCs  $<0.5$ ) in both morning and night samples. ICCs for daily pooled diurnal samples were relatively better than morning and night urine samples, indicating that daily pooled diurnal samples are a better stable representation of an individual's metabolite profile and should be used to enhance the study power. The results from the present study inject a cautionary note about the reproducibility and reliability of the untargeted LC-MS metabolic profiling as it may not provide sufficient precision required to identify subtle changes in the urine metabolome.

## 4.7 References

1. Vrijheid, M. *et al.* The Human Early-Life Exposome (HELIX): Project Rationale and Design. *Environ. Health Perspect.* 122, 535–44 (2014).
2. Chatzi, L. *et al.* Metabolic syndrome in early pregnancy and risk of preterm birth. *Am. J. Epidemiol.* 170, 829–836 (2009).
3. Wright, J. *et al.* Cohort profile: The born in bradford multi-ethnic family cohort study. *Int. J. Epidemiol.* 42, 978–991 (2013).
4. Drouillet, P. *et al.* Association between maternal seafood consumption before pregnancy and fetal growth: Evidence for an association in overweight women. The EDEN mother-child cohort. *Paediatr. Perinat. Epidemiol.* 23, 76–86 (2009).
5. Guxens M. *et al.* Cohort Profile: The INMA-INfancia y Medio Ambiente-(Environment and Childhood) Project. *Int J Epidemiol.* 41, 930-40 (2012).
6. Grazuleviciene, R. *et al.* Maternal smoking, GSTM1 and GSTT1 polymorphism and susceptibility to adverse pregnancy outcomes. *Int. J. Environ. Res. Public Health* 6, 1282–1297 (2009).
7. Magnus P. *et al.* Cohort profile: the Norwegian Mother and Child Cohort Study (MoBa). *Int J Epidemiol* 35(5), 1146- 1150 (2006).
8. Athersuch, T.J. & Keun, H.C. Metabolic profiling in human exposome studies. *Mutagenesis* 30 (6): 755-762 (2015).
9. Athersuch T. Metabolome Analyses in Exposome Studies: Profiling Methods for a Vast Chemical Space. *Arch Biochem Biophys.* 9861(15), 30082-5 (2015).
10. Holmes, E. *et al.* Metabolic Phenotyping in Health and Disease. *Cell* 134, 714–717 (2008).
11. Holmes, E. *et al.* Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature* 453, 396–400 (2008).
12. Keun, H. C. *et al.* Analytical reproducibility in 1H NMR-based metabonomic urinalysis. *Chem. Res. Toxicol.* 15, 1380–1386 (2002).
13. Dumas, M. E. *et al.* Assessment of analytical reproducibility of 1H NMR spectroscopy based metabonomics for large-scale epidemiological research: The INTERMAP study. *Anal. Chem.* 78, 2199–2208 (2006).
14. Theodoridis, G. A. *et al* , I. D. Liquid chromatography-mass spectrometry based global metabolite profiling: A review. *Anal. Chim. Acta* 711, 7–16 (2012).
15. Saude, E. J. *et al* . Variation of metabolites in normal human urine. *Metabolomics* 3, 439–451 (2007).
16. Ellis, J. K. *et al.* Metabolic profiling detects early effects of environmental and lifestyle exposure to cadmium in a human population. *BMC Med.* 10, 61 (2012).
17. Kuo, C. H. *et al.* Metabolomic characterization of labourers exposed to welding fumes. *Chem. Res. Toxicol.* 25, 676–686 (2012).
18. Bonvallet, N. *et al.* Metabolomics tools for describing complex pesticide exposure in pregnant women in Brittany (France). *PLoS One* 8, e64433 (2013).
19. Dallmann, R. *et al* . The human circadian metabolome. *Proc. Natl. Acad. Sci. U. S. A.* 109, 2625–2629 (2012).
20. Xiao, Q. *et al.* Sources of Variability in Metabolite Measurements from Urinary Samples. *PLoS One* 9, 1–8 (2014).
21. Slupsky, C. M. *et al.* Investigations of the effects of gender, diurnal variation, and age in human urinary metabolomic profiles. *Anal. Chem.* 79, 6995–7004 (2007).

22. Psihogios, N. G. *et al.* Gender-related and age-related urinalysis of healthy subjects by NMR-based metabonomics. *NMR Biomed.* 21, 195–207 (2008).
23. Walsh, M. C. *et al.* Effect of acute dietary standardization on the urinary, plasma, and salivary metabolomic profiles of healthy humans. *Am. J. Clin. Nutr.* 84, 531–539 (2006).
24. Wallace, M. *et al.* Effects of menstrual cycle phase on metabolomic profiles in premenopausal women. *Hum. Reprod.* 25, 949–956 (2010).
25. Lenz, E. M. & Wilson, I. D. Analytical strategies in metabonomics. *J. Proteome Res.* 6, 443–458 (2007).
26. Gika, H. G. *et al.* A QC approach to the determination of day-to-day reproducibility and robustness of LC–MS methods for global metabolite profiling in metabonomics/metabolomics. *Bioanalysis* 4, 2239–2247 (2012).
27. Bradman, A. *et al.* Variability of organophosphorous pesticide metabolite levels in spot and 24-hr urine samples collected from young children during 1 week. *Environ. Health Perspect.* 121, 118–124 (2013).
28. Frederiksen, H. *et al.* Temporal variability in urinary phthalate metabolite excretion based on spot, morning, and 24-h urine samples: Considerations for epidemiological studies. *Environ. Sci. Technol.* 47, 958–967 (2013).
29. Warrack, B. M. *et al.* Normalization strategies for metabonomic analysis of urine samples. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 877, 547–552 (2009).
30. Sampson, J. N. *et al.* Metabolomics in epidemiology: Sources of variability in metabolite measurements and implications. *Cancer Epidemiol. Biomarkers Prev.* 22, 631–640 (2013).
31. Nicholson, G. *et al.* Human metabolic profiles are stably controlled by genetic and environmental variation. *Mol. Syst. Biol.* 7, 525 (2011).
32. Teahan, O. *et al.* Impact of analytical bias in metabonomic studies of human blood serum and plasma. *Anal. Chem.* 78, 4307–4318 (2006).
33. Yin, P. *et al.* Effects of pre-analytical processes on blood samples used in metabolomics studies. *Anal. Bioanal. Chem.*, 407 (17), 4879–4892 (2015).
34. Pinto, J. *et al.* Human plasma stability during handling and storage: impact on NMR metabolomics. *Analyst* 139, 1168–77 (2014).
35. Kim, K. *et al.* Mealtime, temporal, and daily variability of the human urinary and plasma metabolomes in a tightly controlled environment. *PLoS One* 9, (2014).
36. Matrie, L. *et al.* Assessment of metabolic phenotypic variability in children's urine using <sup>1</sup>HNMR spectroscopy. *Scientific Reports* 7, (2017).
37. Want, E. J. *et al.* Global metabolic profiling procedures for urine using UPLC-MS. *Nat. Protocol.* 5, 1005–1018 (2010).
38. Dunn, W. B. *et al.* Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat. Protocol.* 6, 1060–1083 (2011).
39. The Human metabolome database version 3.6. <http://www.hmdb.ca/>
40. Bouatra, S. *et al.* The human urine metabolome. *PLoS One* 8, e73076 (2013).
41. Zelena, E. *et al.* Development of a robust and repeatable UPLC - MS method for the long-term metabolomic study of human serum. *Anal. Chem.* 81, 1357–1364 (2009).
42. Cleveland, W. S. Robust Locally Weighted Regression and Smoothing Scatterplots. *J. Am. Stat. Assoc.* 74, 829–836 (1979).
43. Shrout, P. E. & Fleiss, J. L. Intraclass correlations: uses in assessing rater reliability. *Psychol. Bull.* 86, 420–428 (1979).

44. U.S. Department of Health and Human Services, Food and Drug Administration. Guidance for Industry, Bioanalytical Method Validation (2001). at <http://www.fda.gov/downloads/Drugs/Guidances/ucm070107.pdf>
45. European medicine agency. Guideline to Bioanalytical method validation (2011). At [http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Scientific\\_guideline/2011/08/W\\_C500109686.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2011/08/W_C500109686.pdf)
46. Bottcher, C. *et al.* Metabolome Analysis of Biosynthetic Mutants Reveals a Diversity of Metabolic Changes and Allows Identification of a Large Number of New Compounds in Arabidopsis. *Plant Physiol.* 147(4): 2107–2120 (2008).
47. Berg M. *et al.* LC-MS METABOLOMICS FROM STUDY DESIGN TO DATA-ANALYSIS – USING A VERSATILE PATHOGEN AS A TEST CASE. *Computational and Structural Biotechnology Journal* 4 (5), 1-8 (2013).
48. Gika, H. & Theodoridis, G. Within-day reproducibility of an HPLC-MS-based method for metabonomic analysis: application to human urine. *J. Proteome Res.* 6, 3291–3303 (2007).
49. Dunn, W. B. *et al.* Molecular phenotyping of a UK population: defining the human serum metabolome. *Metabolomics* 11, 9–26 (2014).
50. Kabat, G. C. *et al.* Repeated measures of serum glucose and insulin in relation to postmenopausal breast cancer. *Int. J. Cancer* 125, 2704–2710 (2009).
51. Gunter, M. J. *et al.* Insulin, insulin-like growth factor-I, and risk of breast cancer in postmenopausal women. *J. Natl. Cancer Inst.* 101, 48–60 (2009).
52. Boffetta, P. *et al.* Serum cotinine level as predictor of lung cancer risk. *Cancer Epidemiol. Biomarkers Prev.* 15, 1184–1188 (2006).
53. Wang, Z. *et al.* Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* 472, 57–63 (2011).
54. Wang, T. J. *et al.* Metabolite profiles and the risk of developing diabetes. *Nat. Med.* 17, 448–53 (2011).
55. Assfalg, M. *et al.* Evidence of different metabolic phenotypes in humans. *Proc. Natl. Acad. Sci. U. S. A.* 105, 1420–1424 (2008).
56. Bernini, P. *et al.* Individual human phenotypes in metabolic space and time. *J. Proteome Res.* 8, 4264–4271 (2009).

## Chapter 5. Summary and Discussion

This PhD project examined three different LC-MS metabolic profiling platforms to assess their potential and limitations in applications into exposome and molecular epidemiological studies. This was achieved by incorporating LC-MS metabolome analysis into three different kinds of studies in the context of functional genomics, biomarker discovery and defining the exposome, each answering novel research questions. In addition to generating novel data the work in this thesis provides perspective on the importance of precise and standardised LC-MS metabolic profiling methods that can overcome **four key challenges** faced by the field which hamper the discovery of clinically useful and acceptable molecular biomarkers for multifactorial diseases.

The first challenge is achieving the right balance between sensitivity and quantification vs coverage (detecting a maximum number of metabolites) in the experimental strategies employed to analyse metabolome in population studies. This approximates to the choice between targeted and untargeted LC-MS metabolic profiling. Conventional wisdom and established practice in the field of metabolic profiling in large epidemiological studies is that since the aim is to generate hypotheses, it is better not be selective about the molecules to be measured as to use an untargeted approach [1]. The trade-off for using targeted LC-MS metabolic profiling over untargeted metabolic profiling is losing an approach which holds the potential to discover completely novel molecules and unanticipated associations. However, despite some advancements made in analytical techniques [2, 5] data processing techniques, metabolite databases and spectral libraries [3, 4] identifying and annotating unknown peaks in the untargeted LC-MS metabolic profiling is the biggest bottleneck within metabonomics [2, 3, 5]. Introducing another challenge which is that for scientists from other fields such as geneticists, epidemiologist and statisticians who are used to working with well-defined annotated datasets (e.g. genetic data) a panel of unannotated peaks are not readily interpretable.

This raises a question that whether the maximum biochemical information is necessarily gained by acquiring a 'global' profile without prior annotation. Christians *et al.* [6] and Ramautar *et al.* [7] have suggested that using the validated platform to quantify metabolites of selected metabolic pathways could more robust and reliable approach than untargeted metabolic profiling in population studies. Christian *et al.* [6] further discussed the untargeted metabolic profiling is not an unbiased approach to defining the human metabolome. The data acquired from any current experimental strategy or analytical technologies are inherently biased because they capture only part of the metabolome. While subject to the bias of preselection of metabolites, a targeted profile method can still be applied in a hypothesis-free manner. The study undertaken in Chapter 3 was an example of this, seeking to discover the metabolic response to environmental chemical exposure in a mother-child cohort. This method proved efficient and reliable in measuring a wide

panel of metabolites. The study identified novel associations, importantly including to metabolite ratios, which can reflect proxies for different enzymatic reactions, and the data, since quantitative can be readily compared to future quantitative studies independently of the platform used. The use of Biocrates AbsoluteIDQ p180 assay provides a useful compromise to bridge the gap between sensitivity and coverage at the expense of excluding unexpected associations.

Targeted panels, of course, have a place in hypothesis-driven studies. In Chapter 2 a targeted method was applied for the reliable and reproducible absolute quantification of 14 endogenous metabolites of the oestrogen metabolic pathway. This method proved to be a very useful tool to identify that the genetic polymorphism in CYP3A gene is associated with reduced levels of 2-hydroxylation pathway oestrogen metabolites in premenopausal women. The novelty of this study came from the fact that it provided a detailed snapshot of intermediate metabolic phenotype correlating genotype-phenotype data by establishing a link between genetic variant, oestrogen pathway and reduced levels of serum metabolites in carefully genotyped population. However, this method was very specific to the biological question to be answered and provided in-depth but limited coverage of metabolites. Overall the method was fit for purpose in a small, hypothesis-driven study regarding sensitivity, accuracy and precision. However, improvements will be made in future work to reduce the run time of the assay if it were to be applied to large cohorts.

By contrast in Chapter 4 an untargeted LC-MS metabolic profiling method was employed in an exposome study to analyse short-term temporal variation in annotated metabolites in healthy children. This method, despite being analytically robust and reproducible in the large-scale continuous analysis (5 batches of 96-well plates) showed high intra-individual variations compared to inter-individual variations. The data from 'metabo-ring initiative' showed that 11 different LC-MS instruments using untargeted LC-MS metabolic profiling were unable to detect inter-individual variability in low biological contrast test groups compared to high-biological test groups [8]. Therefore, it is crucial when the study aims to detect subtle changes in metabolic effect in a healthy population with a similar biological background that absolute precision of the analytical technique is very high and the sample size is big enough to observe any moderate differences.

In this work, the challenge of peak annotations and identification was addressed by building a chemical standard reference library of individual standards and matching mass spectra peaks with reference standard's retention time and monoisotopic mass. Early efforts of this approach proved to be a time-saving method to assist in the identification of unknown peaks unlike using peak picking software (e.g. XCMS) when annotation occurs after an unannotated feature is identified therefore delaying the process of metabolite identification and interpretation of

biological significance. This approach has the potential to significantly improve the real-time annotation of the peaks by incorporating reference standard mixtures which are relevant to the study design, resemble biological matrix or deemed important from any prior analysis (e.g. NMR). However, this approach requires additional work and more computational time to build chemical reference mixtures that provide a breadth of biochemical coverage. In future work, standard reference mixtures will be constructed possibly at different concentrations (to observe the linearity response of the detector) to be run within any large population study to improve the speed and reliability of metabolite identification.

The second major challenge is a lack of standardisation of the assays and data comparability. The comparisons of experimental metabolic datasets from one population are often not verified in another comparative independent study due to experimental differences. However, adoption of standard protocols validated according to the FDA & EMA guidelines for Bioanalytical method validation can undoubtedly facilitate direct comparability of the data. In targeted metabolite profiling, sample preparation steps plus a choice of instrument can be easily harmonised in inter-laboratory experiments as it focused towards metabolites of interest. Also, the use of isotope labelled standards and internal standards provides additional confidence in absolute metabolite quantification. In Chapter 2 the oestrogen metabolites LC-MS/MS quantification method allowed direct comparison of the data between the concentrations of metabolites in work done in this chapter with that of Nurses' health population II study carried out in the US [9]. This not only facilitated a broader understanding of oestrogen metabolism in healthy pre-menopausal women but also assured the normal level of oestrogen metabolite concentrations in two different population cohorts. Defining endogenous oestrogen metabolites concentrations in premenopausal women has been seen as a cumbersome analysis in the literature due to variability introduced via menstrual cycle. In this work the use of standardised method allowed to provide absolute concentrations of oestrogen metabolites in follicular and compare with previously published luteal phase data.

Similarly, Biocrates AbsoluteIDQ method has been demonstrated fit for the purpose of many studies. Its participation in 'international ring trial' [10] is a major step in standardisation and quality assessment of the assay. In the international ring trial study, five different laboratories across Europe participated in assessing the reliability and comparability of Biocrates assay and its metabolite datasets. The results showed  $CV < 15\%$  for both LC- and FIA-MS/MS metabolites in all five groups.

In untargeted LC-MS metabolic profiling even after following similar sample preparation protocol inter-laboratory data comparability is often difficult and unreliable for three main reasons: 1. variations introduced within and between instruments; 2. Ion suppression and adduct formation

cannot be harmonised 3. Any subtle change in the column and chromatographic method will change the retention time of any identified metabolites making automation of data processing between laboratories very difficult. Several groups around the world using untargeted LC-MS metabolic profiling are building their internal databases of detected and identified metabolites using reference chemical standards or through MS/MS experiments. However, the question remains as to how reliable these datasets will be when compared to experiments carried out in other laboratories. The HUSERMET study implemented standardised experimental design and incorporated pooled QC samples to define normal relative baseline metabolite concentrations in serum samples of 1200 healthy individuals using non-targeted LC- and GC-MS platforms [11]. To provide comparable metabolite profiles, this study implemented the use of standard pool QC samples to track instrument performance drift, feature-wise bias correction and to normalise analytical batch variation. A similar QC strategy to correct for analytical variation and metabolite signal correction using QC based robust LOESS signal correction was also incorporated in Chapter 4 of this thesis. Standardised experimental design, dividing study into multiple analytical experiments and incorporating pooled QC strategy is critical for quality assurance of untargeted-chromatographic-MS analysis of routine, large-scale size continuous studies.

The third challenge is to minimise the effect of the diurnal or circadian rhythm of the body on metabolite level variations to maximise biologically relevant extractable information. It also ties in with the question of the reliability of single spot urine metabolite profiles. To address this attention was turned to work done in Chapter 4 where data demonstrated that urine metabolite levels are more reliably determined in pooled diurnal (pool of morning void and before bed) samples compared to single spot urine from any single time point within a day. These results were supported by the NMR analysis carried out on same sample sets. This highlighted important factor that metabolite levels differ according to the urine sampling time which can be considered as a pre-analytical variability and can bias the overall outcome of the studies. These results also have huge implications for experiment designs for future exposome and molecular epidemiological studies. Pooling daily diurnal urine sample before experimental and analytical analysis is also cost-effective especially for population studies where multiple time point samples are available by reducing the number of samples that need to be analysed without losing biological information.

Finally, irrespective of the choice of experimental strategy or analytical method used to characterise the metabolome there is a lack of longitudinal LC-MS data that explains the normal level of variation in identified metabolites across the healthy population. To add to this limited knowledge work done in Chapter 4 showed high short-term temporal intra- and inter-individual variation in identified metabolites.

Though this work had some limitations and still brings to light that short-term temporal variations are drivers of variances in urine metabolite profiles in healthy individuals. This may affect the interpretation of molecular epidemiological and exposome studies. This work highlighted the importance and need of large sample size longitudinal population studies to enhance the study power of metabonomics studies. This information was vital to design optimal sampling strategies for the future work on HELIX project. It allowed making decisions on which analytical platform to use for untargeted metabolic profiling and allowed better estimation of the power to detect associations to exposure profiles. It is also noteworthy to mention that while untargeted LC-MS metabolic profiling is considered as an economical and cost-effective way to acquire data on large sample sets, additional requirements of the approach in terms of time spent on data processing, purchase of authentic chemical standards, running additional experiments including sample preparation to identify metabolites become increasingly expensive, including human capital cost and time required per project. Conversely, in targeted LC-MS metabolic profiling where metabolites of interest are predefined, the cost of isotope labelled internal standards, consumables required before sample preparation, and the time required to prepare calibration curves are no less expensive but can be estimate upfront and operational costs can be reduced. Most often no additional experiments are required, and data analysis is straightforward as observed in the case of the oestrogen method in Chapter 2. Target profiling kits provide a middle ground to reduce the operational and post-operational expense when large-scale studies require analysis of a large panel of metabolites. The Biocrates method is highly automated and standardised from sample preparation to metabolite quantification which will not only allow better prediction of cost, increase throughput but also decreases sample-handling errors and is likely to improve comparability across time in large-scale studies.

An underlying aim of all the efforts in this PhD project was to try to gain understanding as to the factors determining the appropriate experimental in characterising metabolic phenotypes in large population studies, particularly the trade-offs between sensitivity, pre-annotation, quantification and metabolome coverage. It is hoped that findings from this PhD work are a foundation for the designing future metabolome-exposome studies and will facilitate the successful transition of any molecular biomarker into clinical settings.

## 5.1 References

1. Lindon, J. C. & Nicholson, J. K. Spectroscopic and statistical techniques for information recovery in metabonomics and metabolomics. *Annu. Rev. Anal. Chem. (Palo Alto. Calif)*. 1, 45–69 (2008).
2. Kuehnbaum, N. L. & Britz-McKibbin, P. New advances in separation science for metabolomics: resolving chemical diversity in a post-genomic era. *Chem. Rev.* 113, 2437–68 (2013).
3. Neumann, S. & Böcker, S. Computational mass spectrometry for metabolomics: Identification of metabolites and small molecules. *Anal. Bioanal. Chem.* 398, 2779–2788 (2010).
4. Menikarachchi, L. C. *et al.* MolFind: A software package enabling HPLC/MS-based identification of unknown chemical structures. *Anal. Chem.* 84, 9388–9394 (2012).
5. Scalbert, A. *et al.* Mass-spectrometry-based metabolomics: Limitations and recommendations for future progress with particular focus on nutrition research. *Metabolomics* 5, 435–458 (2009).
6. Christians, U. *et al.* How unbiased is non-targeted metabolomics and is targeted pathway screening the solution? *Curr. Pharm. Biotechnol.* 12, 1053–66 (2011).
7. Ramautar, R. *et al.* Human metabolomics: strategies to understand biology. *Curr. Opin. Chem. Biol.* 1–6 (2013).
8. Martin, J. C. *et al.* Can we trust untargeted metabolomics? Results of the metabo-ring initiative, a large-scale, multi-instrument inter-laboratory study. *Metabolomics* 11, 807–821 (2015).
9. Eliassen, A. H. *et al.* Reproducibility of fifteen urinary estrogens and estrogen metabolites over a 2- to 3-year period in premenopausal women. *Cancer Epidemiol. Biomarkers Prev.* 18, 2860–2868 (2009).
10. Siskos, A. P. *et al.* Inter-laboratory reproducibility of a targeted metabolomics platform for analysis of human serum and plasma. *Anal. Chem.* (2016).
11. Dunn, W. B. *et al.* Molecular phenotyping of a UK population: defining the human serum metabolome. *Metabolomics* 11, 9–26 (2014).

## Appendix

**Appendix Table 1. List of metabolites measured using Biocrates p180 LC-MS/MS assay**

| <b>Amino acids</b>  | <b>Biogenic amines</b>             |
|---------------------|------------------------------------|
| Alanine (Ala)       | Acetylmethionine (Ac-Met)          |
| Arginine (Arg)      | Asymmetric dimethylarginine (ADMA) |
| Asparagine( Asn)    | Symmetric dimethylarginine (SDMA)  |
| Aspartate (Asp)     | Total dimethylarginine (total DMA) |
| Citrulline (Cit)    | Alpha-Aminoadipic acid (alpha-AAA) |
| Glutamine (Gln)     | Carnosine                          |
| Glutamate (Glu)     | Creatinine                         |
| Glycine (Gly)       | Histamine                          |
| Histidine (His)     | Kynurenine                         |
| Isoleucine (Ile)    | Methioninesulfoxide (Met-SO)       |
| Leucine (Leu)       | Nitrotyrosine (nitro-Tyr)          |
| Lysine (Lys)        | Hydroxyproline (OH-Pro)            |
| Methionine (Met)    | Phenylethylamine (PEA)             |
| Ornithine (Orn)     | Putrescine                         |
| Phenylalanine (Phe) | Sarcosine                          |
| Proline (Pro)       | Serotonin                          |
| Serine (Ser)        | Spermidine                         |
| Threonine (Thr)     | Spermine                           |
| Tryptophan (Trp)    | Taurine                            |
| Tyrosine (Tyr)      |                                    |
| Valine (Val)        |                                    |

**Appendix Table 2. List of metabolites measured using Biocrates p180 FIA-MS/MS assay**

| <b>Acylcarnitine</b>                      | <b>Sphingolipids</b>                       |
|-------------------------------------------|--------------------------------------------|
| Carnitine (C0)                            | Hydroxysphingomyeline C14:1 (SM (OH)C14:1) |
| Acetylcarnitine (C2)                      | Sphingomyeline C16:0 (SM C16:0)            |
| Propionylcarnitine (C3)                   | Sphingomyeline C16:1 (SM C16:1)            |
| Hydroxypropionylcarnitine (C3-OH)         | Hydroxysphingomyeline C16:1 (SM (OH)C16:1) |
| Butyrylcarnitine (C4)                     | Sphingomyeline C18:0 (SM C18:0)            |
| Butenylcarnitine (C4:1)                   | Sphingomyeline C18:1 (SM C18:1)            |
| Hydroxybutyrylcarnitine (C4-OH (C3-DC))   | Sphingomyeline C20:2 (SM C20:2)            |
| Valeryl carnitine (C5)                    | Sphingomyeline C22:3 (SM C22:3)            |
| Tiglylcarnitine (C5:1)                    | Hydroxysphingomyeline C22:1 (SM (OH)C22:1) |
| Glutaconylcarnitine (C5:1-DC)             | Hydroxysphingomyeline C22:2 (SM (OH)C22:2) |
| Hydroxyhexanoylcarnitine ((C5-DC (C6-OH)) | Sphingomyeline C24:0 (SM C24:0)            |
| Methylglutaryl carnitine (C5-M-DC)        | Sphingomyeline C24:1 (SM C24:1)            |
| Methylmalonylcarnitine ((C5-OH (C3-DC-M)) | Hydroxysphingomyeline C24:1 (SM (OH)C24:1) |
| Fumaryl carnitine (C6(C4:1-DC))           | Sphingomyeline C26:0 (SM C26:0)            |
| Hexenoylcarnitine (C6:1)                  | Sphingomyeline C26:1 (SM C26:1)            |
| Pimelylcarnitine (C7-DC)                  |                                            |
| Octanoylcarnitine (C8)                    |                                            |
| Nonaylcarnitine (C9)                      | <b>Sugars</b>                              |
| Decanoylcarnitine (C10)                   | Hexose (H1)                                |
| Decenoylcarnitine (C10:1)                 |                                            |
| Decadienylcarnitine (C10:2)               |                                            |

|                                                    |                                                    |
|----------------------------------------------------|----------------------------------------------------|
| Dodecanoylcarnitine (C12)                          |                                                    |
| Dodecanoylcarnitine (C12:1)                        |                                                    |
| Dodecanedioylcarnitine (C12-DC)                    |                                                    |
| Tetradecanoylcarnitine (C14)                       |                                                    |
| Tetradecenoylcarnitine (C14:1)                     |                                                    |
| Hydroxytetradecenoylcarnitine (C14:1-OH)           |                                                    |
| Tetradecadienylcarnitine (C14:2)                   |                                                    |
| Hydroxytetradecenoylcarnitine (C14:2-OH)           |                                                    |
| Hexadecanoylcarnitine (C16)                        |                                                    |
| Hexadecenoylcarnitine (C16:1)                      |                                                    |
| Hydroxyhexadecenoylcarnitine (C16:1-OH)            |                                                    |
| Hexadecadienylcarnitine (C16:2)                    |                                                    |
| Hydroxyhexadecadienylcarnitine (C16:2-OH)          |                                                    |
| Hydroxyhexadecanoylcarnitine (C16-OH)              |                                                    |
| Octadecanoylcarnitine (C18)                        |                                                    |
| Octadeceanoylcarnitine (C18:1)                     |                                                    |
| Hydroxyoctadecenoylcarnitine (C18:1-OH)            |                                                    |
| Octadecadienylcarnitine (C18:2)                    |                                                    |
| Glycerophospholipids                               |                                                    |
|                                                    | Phosphatidylcholine diacyl C42:0 (PC aa C42:0)     |
| lysoPhospatidylcholine acyl C14:0 (lysoPC a C14:0) | Phosphatidylcholine diacyl C42:1 (PC aa C42:1)     |
| lysoPhospatidylcholine acyl C16:0 (lysoPC a C16:0) | Phosphatidylcholine diacyl C42:2 (PC aa C42:2)     |
| lysoPhospatidylcholine acyl C16:1 (lysoPC a C16:1) | Phosphatidylcholine diacyl C42:4 (PC aa C42:4)     |
| lysoPhospatidylcholine acyl C17:0 (lysoPC a C17:0) | Phosphatidylcholine diacyl C42:5 (PC aa C42:5)     |
| lysoPhospatidylcholine acyl C18:0 (lysoPC a C18:0) | Phosphatidylcholine diacyl C42:6 (PC aa C42:6)     |
| lysoPhospatidylcholine acyl C18:1 (lysoPC a C18:1) | Phosphatidylcholine acyl-alkyl C30:0 (PC ae C30:0) |
| lysoPhospatidylcholine acyl C20:3 (lysoPC a C20:3) | Phosphatidylcholine acyl-alkyl C30:1 (PC ae C30:1) |
| lysoPhospatidylcholine acyl C20:4 (lysoPC a C20:4) | Phosphatidylcholine acyl-alkyl C30:2 (PC ae C30:2) |
| lysoPhospatidylcholine acyl C24:0 (lysoPC a C24:0) | Phosphatidylcholine acyl-alkyl C32:1 (PC ae C32:1) |
| lysoPhospatidylcholine acyl C26:0 (lysoPC a C26:0) | Phosphatidylcholine acyl-alkyl C32:2 (PC ae C32:2) |
| lysoPhospatidylcholine acyl C26:1 (lysoPC a C26:1) | Phosphatidylcholine acyl-alkyl C34:0 (PC ae C34:0) |
| lysoPhospatidylcholine acyl C28:0 (lysoPC a C28:0) | Phosphatidylcholine acyl-alkyl C34:1 (PC ae C34:1) |
| lysoPhospatidylcholine acyl C28:1 (lysoPC a C28:1) | Phosphatidylcholine acyl-alkyl C34:2 (PC ae C34:2) |
| Phosphatidylcholine diacyl C24:0 (PC aa C24:0)     | Phosphatidylcholine acyl-alkyl C34:3 (PC ae C34:3) |
| Phosphatidylcholine diacyl C26:0 (PC aa C26:0)     | Phosphatidylcholine acyl-alkyl C36:0 (PC ae C36:0) |
| Phosphatidylcholine diacyl C28:1 (PC aa C28:1)     | Phosphatidylcholine acyl-alkyl C36:1 (PC ae C36:1) |
| Phosphatidylcholine diacyl C30:0 (PC aa C30:0)     | Phosphatidylcholine acyl-alkyl C36:2 (PC ae C36:2) |
| Phosphatidylcholine diacyl C30:2 (PC aa C30:2)     | Phosphatidylcholine acyl-alkyl C36:3 (PC ae C36:3) |
| Phosphatidylcholine diacyl C32:0 (PC aa C32:0)     | Phosphatidylcholine acyl-alkyl C36:4 (PC ae C36:4) |
| Phosphatidylcholine diacyl C32:1 (PC aa C32:1)     | Phosphatidylcholine acyl-alkyl C36:5 (PC ae C36:5) |
| Phosphatidylcholine diacyl C32:2 (PC aa C32:2)     | Phosphatidylcholine acyl-alkyl C38:0 (PC ae C38:0) |
| Phosphatidylcholine diacyl C32:3 (PC aa C32:3)     | Phosphatidylcholine acyl-alkyl C38:1 (PC ae C38:1) |
| Phosphatidylcholine diacyl C34:1 (PC aa C34:1)     | Phosphatidylcholine acyl-alkyl C38:2 (PC ae C38:2) |
| Phosphatidylcholine diacyl C34:2 (PC aa C34:2)     | Phosphatidylcholine acyl-alkyl C38:3 (PC ae C38:3) |
| Phosphatidylcholine diacyl C34:3 (PC aa C34:3)     | Phosphatidylcholine acyl-alkyl C38:4 (PC ae C38:4) |
| Phosphatidylcholine diacyl C34:4 (PC aa C34:4)     | Phosphatidylcholine acyl-alkyl C38:5 (PC ae C38:5) |
| Phosphatidylcholine diacyl C36:0 (PC aa C36:0)     | Phosphatidylcholine acyl-alkyl C38:6 (PC ae C38:6) |
| Phosphatidylcholine diacyl C36:1 (PC aa C36:1)     | Phosphatidylcholine acyl-alkyl C40:1 (PC ae C40:1) |
| Phosphatidylcholine diacyl C36:2 (PC aa C36:2)     | Phosphatidylcholine acyl-alkyl C40:2 (PC ae C40:2) |
| Phosphatidylcholine diacyl C36:3 (PC aa C36:3)     | Phosphatidylcholine acyl-alkyl C40:3 (PC ae C40:3) |

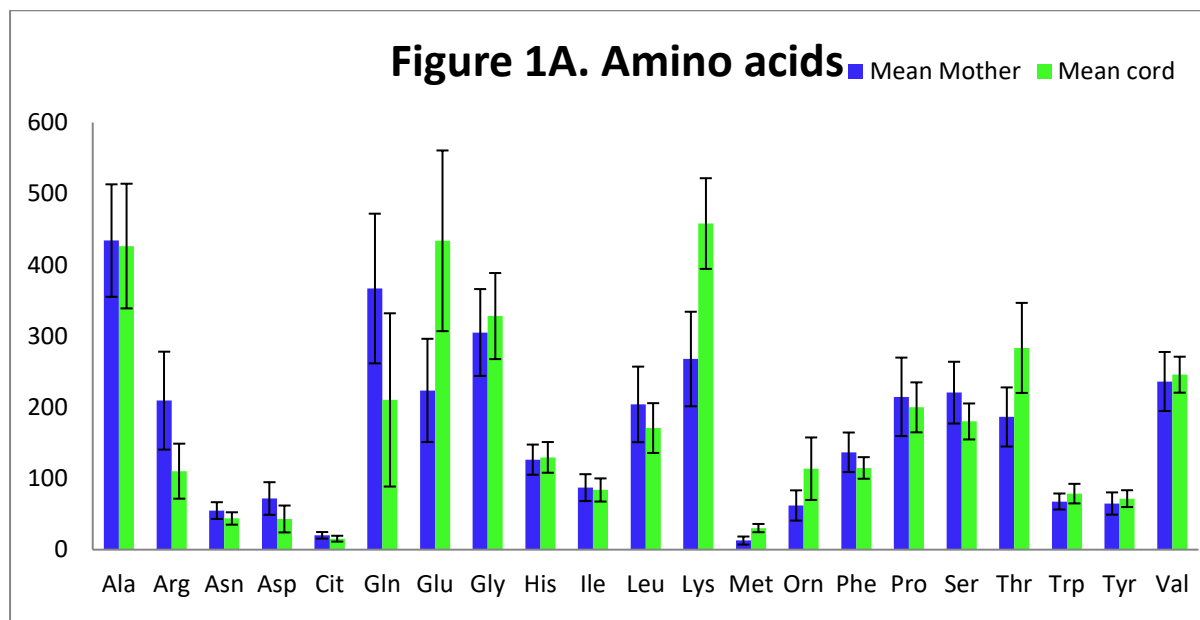
|                                                 |                                                    |
|-------------------------------------------------|----------------------------------------------------|
| Phosphatidylcholine diacyl C36:4 (PC aa C36:4)  | Phosphatidylcholine acyl-alkyl C40:4 (PC ae C40:4) |
| Phosphatidylcholine diacyl C36:5 (PC aa C36:5)  | Phosphatidylcholine acyl-alkyl C40:5 (PC ae C40:5) |
| Phosphatidylcholine diacyl C36:6 (PC aa C36:6)  | Phosphatidylcholine acyl-alkyl C40:6 (PC ae C40:6) |
| Phosphatidylcholine diacyl C38:0 (PC aa C38:0)  | Phosphatidylcholine acyl-alkyl C42:0 (PC ae C42:0) |
| Phosphatidylcholine diacyl C38:1 (PC aa C38:0)  | Phosphatidylcholine acyl-alkyl C42:1 (PC ae C42:1) |
| Phosphatidylcholine diacyl C38:3 (PC aa C38:03) | Phosphatidylcholine acyl-alkyl C42:2 (PC ae C42:2) |
| Phosphatidylcholine diacyl C38:4 (PC aa C38:04) | Phosphatidylcholine acyl-alkyl C42:3 (PC ae C42:3) |
| Phosphatidylcholine diacyl C38:5 (PC aa C38:5)  | Phosphatidylcholine acyl-alkyl C42:4 (PC ae C42:4) |
| Phosphatidylcholine diacyl C38:6 (PC aa C38:6)  | Phosphatidylcholine acyl-alkyl C42:5 (PC ae C42:5) |
| Phosphatidylcholine diacyl C40:1 (PC aa C40:1)  | Phosphatidylcholine acyl-alkyl C44:3 (PC ae C44:3) |
| Phosphatidylcholine diacyl C40:2 (PC aa C40:2)  | Phosphatidylcholine acyl-alkyl C44:4 (PC ae C44:4) |
| Phosphatidylcholine diacyl C40:3 (PC aa C40:3)  | Phosphatidylcholine acyl-alkyl C44:5 (PC ae C44:5) |
| Phosphatidylcholine diacyl C40:4 (PC aa C40:4)  | Phosphatidylcholine acyl-alkyl C44:6 (PC ae C44:6) |
| Phosphatidylcholine diacyl C40:5 (PC aa C40:5)  |                                                    |
| Phosphatidylcholine diacyl C40:6 (PC aa C40:6)  |                                                    |

**Appendix Table 3. List of 28 ratios or sums of selected metabolites with significance or proxy to provide detailed insight into a wide range of disorders of the metabolic diseases.**

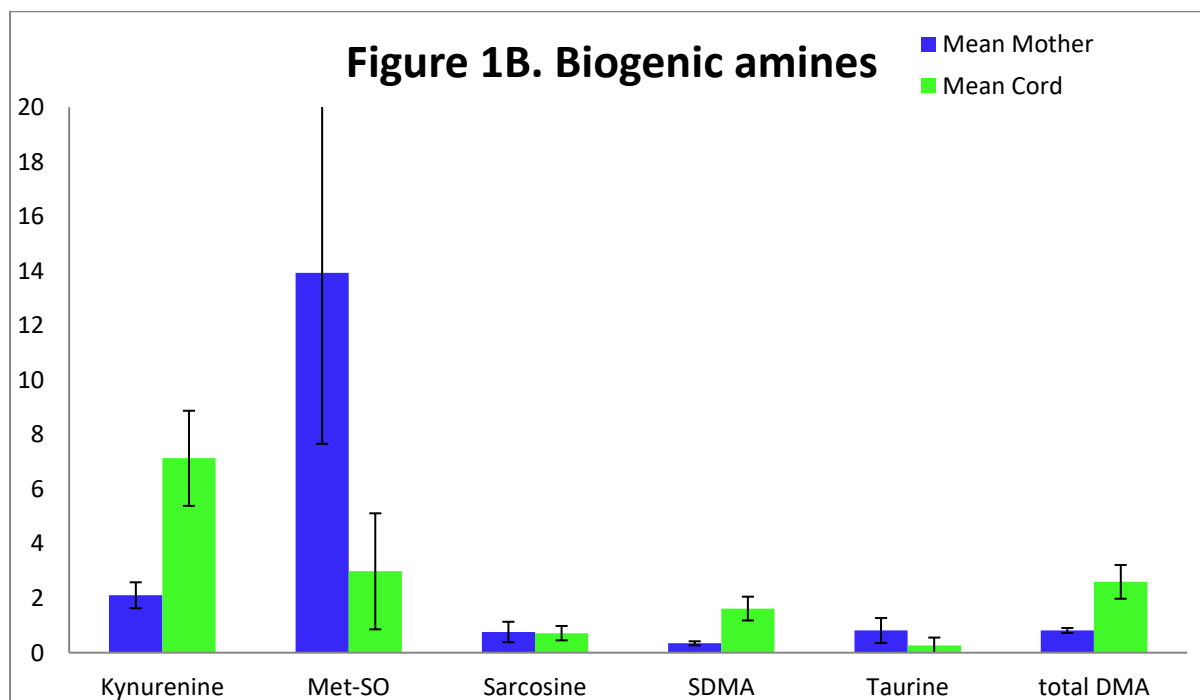
| Ratio or Sum     | Description                                                                | Metabolic significance                                                                                |
|------------------|----------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------|
| Essential AA     | Sum of essential amino acids                                               | Indicator of nutritional significance                                                                 |
| Non-essential AA | Sum of non-essential amino acids                                           | Indicator metabolic catabolic status                                                                  |
| BCAA             | Sum of branched chained amino acids                                        | Indicator of short-term metabolic control ( analogy with muscular insulin resistance)                 |
| Fischer ratio    | Ratio of BCAA to aromatic AA                                               | Indicator of liver damage                                                                             |
| Glucogenic AA    | Sum of selected glucogenic amino acids (alanine, glycine, serine)          | Indicator of balance between glycolysis and gluconeogenesis                                           |
| ADMA/Arg         | Ratio of asymmetrically dimethylated arginine to total unmodified arginine | Inhibition of NO synthase, associated with endothelial dysfunction and cardiovascular risk in general |
| SDMA/Arg         | Ratio of symmetrically dimethylated arginine to total unmodified arginine  | Indicator of impaired kidney function/renal failure                                                   |
| Total DMA/Arg    | Ratio of total dimethylated arginine to total unmodified arginine          | Activity of protein arginine methyl transferases                                                      |
| Cit/Arg          | Ratio of citrulline to arginine                                            | Activity of nitric oxide synthase (NOS)                                                               |
| Cit/Orn          | Ratio of citrulline to ornithine                                           | Activity of ornithine carbamoylphosphate transferase                                                  |
| Orn/Arg          | Ratio of ornithine to arginine                                             | Activity of arginase                                                                                  |
| Tyr/Phe          | Ratio of tyrosine to phenylalanine                                         | Activity of phenylalanine hydroxylase                                                                 |
| Met-SO/Met       | Ratio of sulfoxidized methionine to total unmodified methionine            | Measure of systemic oxidative stress                                                                  |

|                       |                                                                  |                                                                                                                 |
|-----------------------|------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------|
| Kynurenine/Trp        | Ratio of kynurenine to tryptophan                                | Rate of Trp degradation to niacin , indicator of indole dioxygenase acitivity (immunosuppression/tolerance)     |
| Serotonin/Trp         | Ratio of serotonin to tryptophan                                 | Rate of Trp degradation to serotnin                                                                             |
| Putrescine/Orn        | Ratio of putrescine to tryptophan                                | Activity of ornithine decarboxylase                                                                             |
| Spermidine/Putrescine | Ratio of spermidine to putrescine                                | Activity of spermidine syntase                                                                                  |
| Spermine/Spermidine   | Ratio of spermine to spermidine                                  | Activity of spermine synthase                                                                                   |
| C2/C0                 | Rate of acetylcarnitine to free carnitine                        | Measure of $\beta$ -oxidation of even numbered fatty acids                                                      |
| (C2+C3)/C0            | Ratio of short chain acylcarnitine to free carnitine             | Measure of overall $\beta$ -oxidation activity                                                                  |
| CPT-I([C16+C18]/C0)   | Ratio of long chain acylcarnitine to free carnitine              | Activity of carnitine palmitoyltransferase I, rate-limiting step in the uptake of fatty acids into mitochondria |
| Total AC-DC/Total AC  | Ratio of dicarboxy-acylcarnitines to total acylcarnitines        | Indicator of v-oxidation of fatty acids                                                                         |
| MUFA (PC)/SFA (PC)    | Ratio of mono-unsaturated to saturated glycerophosphocholines    | Measure of the activity of fatty acids desaturases and indicator of nutritional lipid composition               |
| PUFA (PC)             | Sum of poly-unstaturated glycerophosphocholines                  | Indicator of nutritional lipid composition                                                                      |
| PUFA (PC)/MUFA(PC)    | Ratio of poly-unsaturated to unstaturated glycerophosphocholines | Measure of the activity of fatty acids desaturases and indicator of nutritional lipid composition               |
| PUFA (PC)/SFA (PC)    | Ratio of poly-unsaturated to staturated glycerophosphocholines   | Measure of the activity of fatty acids desaturases and indicator of nutritional lipid composition               |
| SFA(PC)               | Sum of staured glycerophosphocholines                            | Indicator of nutritional lipid composition                                                                      |
| Total lysoPC/Total PC | Ratio of lysoglycerophosphocholines to glycerophosphocholines    | Indicator of phospholipase activity                                                                             |

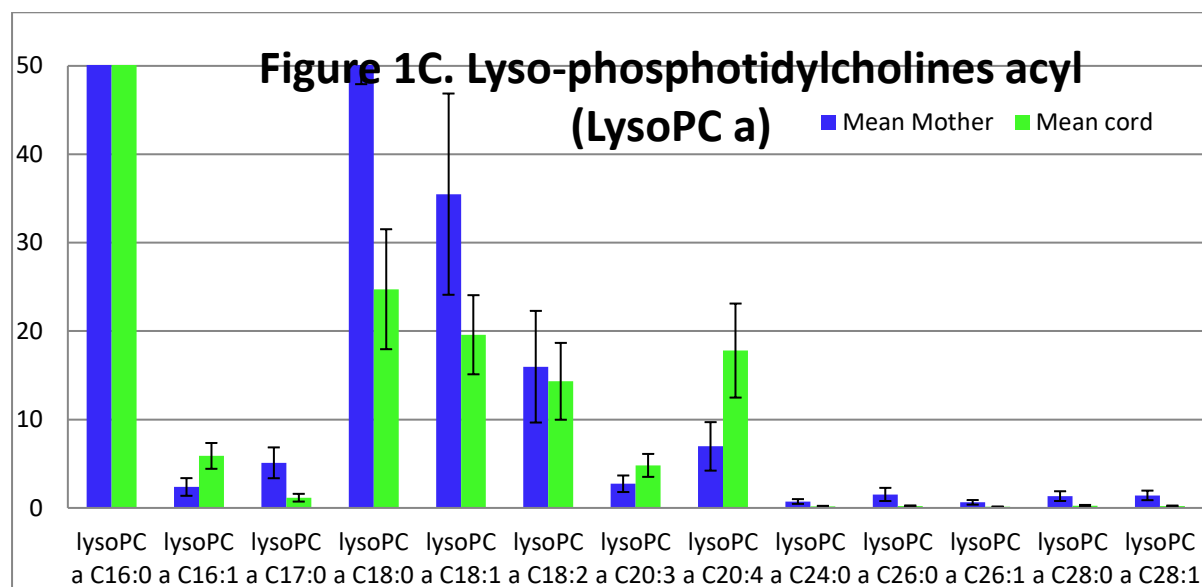
**Appendix Figure 1A** Bar graph comparing average concentrations of maternal and cord serum amino acids. X-axis represents individual amino acid metabolites and Y-axis represents their mean concentration in ( $\mu\text{M}$ ).



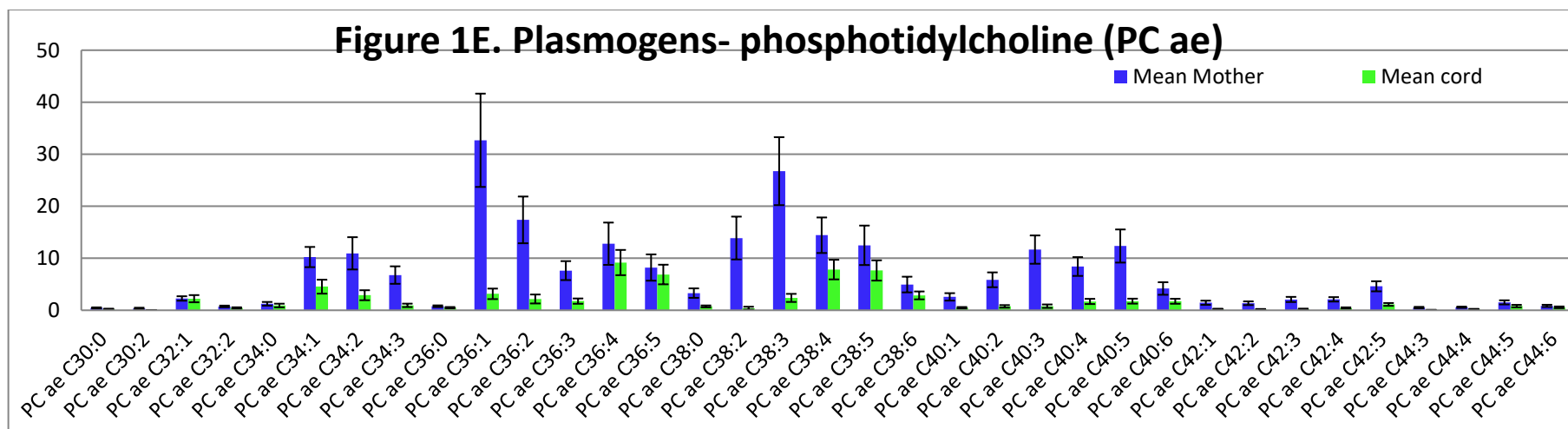
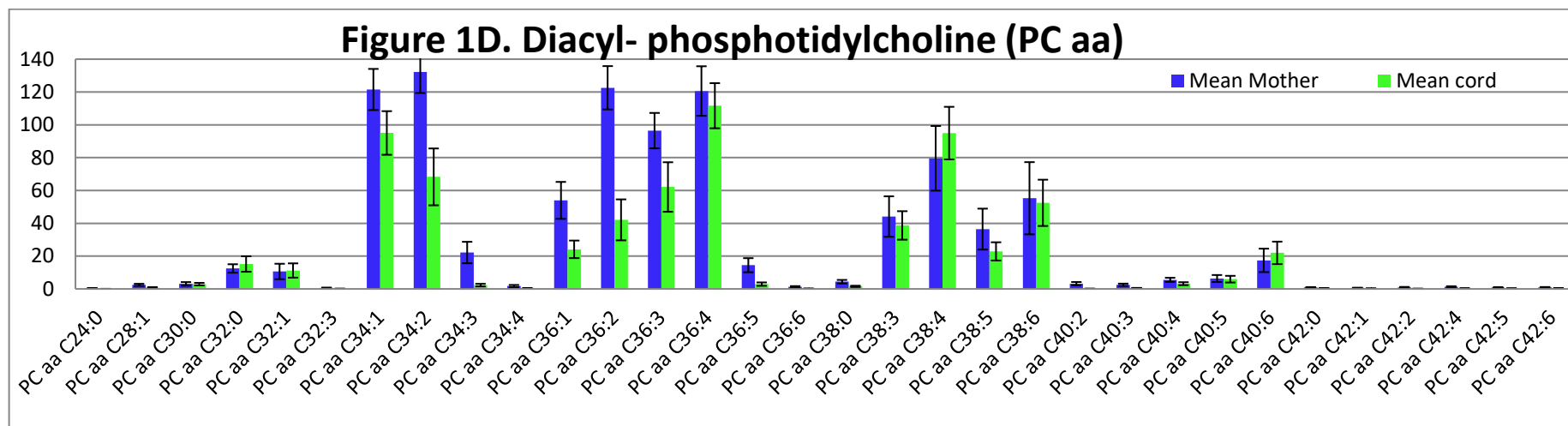
**Appendix Figure 1B** Bar graph comparing average concentrations of maternal and cord serum biogenic amines. X-axis represents individual biogenic amine metabolites and Y-axis represents their mean concentration in ( $\mu\text{M}$ ).



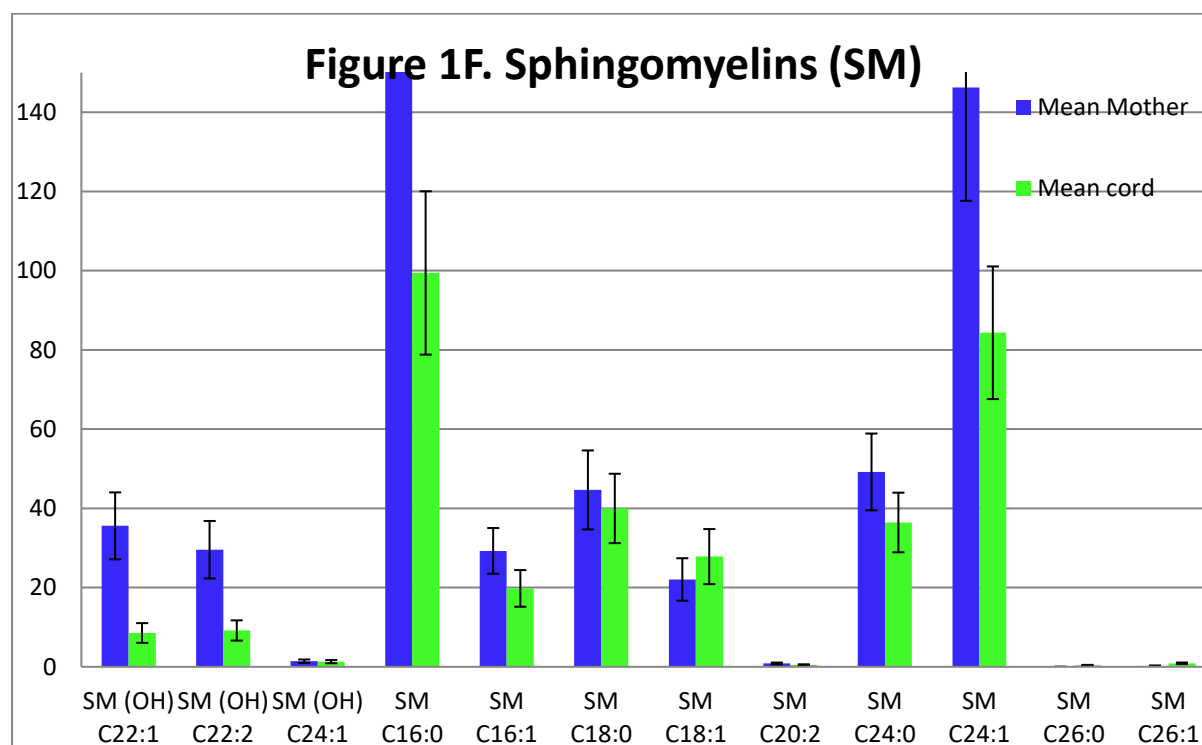
**Appendix Figure 1C.** Bar graph comparing average concentrations of maternal and cord serum Lyso-glycerophosphocholines (Lyso PC). X-axis represents individual Lyso-glycerophosphocholines metabolites and Y-axis represents their mean concentration in ( $\mu$ M).



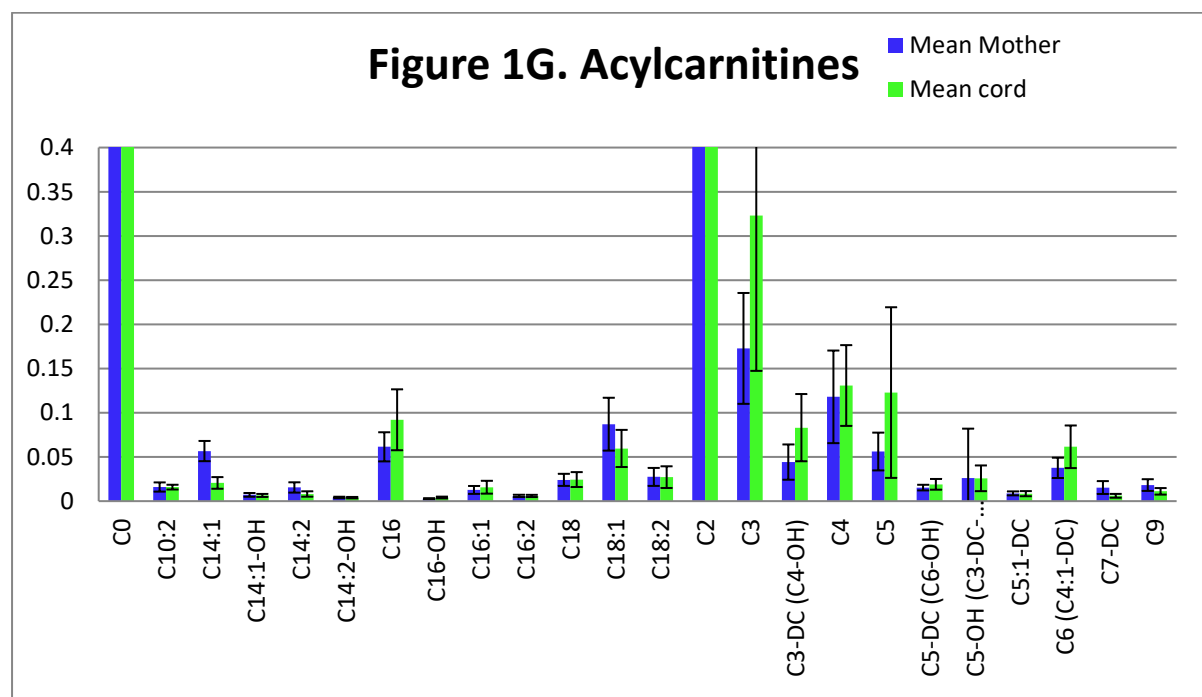
**Appendix Figure 1D & 1E.** Bar graph comparing average concentrations of maternal and cord serum Diacyl-glycerophosphocholines (PC aa) and Plasmogens-glycerphosphocholine (PC ae). X-axis represents individual metabolites and Y-axis represents their mean concentration in ( $\mu\text{M}$ ).



**Appendix Figure 1F. Bar graph comparing average concentrations of maternal and cord serum Sphingomyelins (SM). X-axis represents individual sphingomyelins and hydroxyl sphingomyelins (SM-OH) metabolites and Y-axis represents their mean concentration in ( $\mu\text{M}$ ).**



**Appendix Figure 1G. Bar graph comparing average concentrations of maternal and cord serum acylcarnitine. X-axis represents individual carnitine metabolites and Y-axis represents their mean concentration in ( $\mu\text{M}$ ).**



**Appendix Table 4 Spearman correlation Rho values and P values reported, metabolites significantly associated ( $P \leq 0.05$ ) with different POPs ( $\Sigma$ PCBs, HCB and DDE) in 1st trimester maternal serum in N=67 women with healthy pregnancy from the Rhea pregnancy cohort.**

| Metabolites      | $\Sigma$ PCBs        |         | HCB                  |         | DDE                  |         |
|------------------|----------------------|---------|----------------------|---------|----------------------|---------|
|                  | Spearman coefficient | P value | Spearman coefficient | P value | Spearman coefficient | P value |
| Citrulline       | -0.268               | 0.029   | -                    | -       | -                    | -       |
| Kynurenine       | -0.376               | 0.002   | -0.362               | 0.003   | -                    | -       |
| lysoPC a C16:0   | 0.275                | 0.024   | 0.29                 | 0.017   | -                    | -       |
| PC aa C28:1      | 0.251                | 0.04    | -                    | -       | -                    | -       |
| PC aa C30:0      | 0.241                | 0.049   | -                    | -       | -                    | -       |
| PC aa C32:1      | 0.274                | 0.025   | 0.283                | 0.02    | -                    | -       |
| PC aa C32:3      | -                    | -       | 0.267                | 0.029   | -                    | -       |
| PC aa C34:3      | 0.356                | 0.003   | 0.361                | 0.003   | -                    | -       |
| PC aa C34:4      | 0.36                 | 0.003   | 0.303                | 0.013   | -                    | -       |
| PC aa C36:3      | 0.262                | 0.032   | -                    | -       | -                    | -       |
| PC aa C36:4      | 0.282                | 0.021   | -                    | -       | -                    | -       |
| PC aa C36:5      | 0.36                 | 0.003   | 0.411                | 0.001   | -                    | -       |
| PC aa C36:6      | 0.311                | 0.01    | 0.432                | 0       | -                    | -       |
| PC aa C38:3      | 0.306                | 0.012   | 0.247                | 0.044   | -                    | -       |
| PC aa C38:4      | 0.288                | 0.018   | -                    | -       | -                    | -       |
| PC aa C38:5      | 0.318                | 0.009   | -                    | -       | -                    | -       |
| PC aa C38:6      | 0.249                | 0.043   | -                    | -       | -                    | -       |
| PC aa C40:4      | 0.304                | 0.012   | 0.298                | 0.014   | -                    | -       |
| PC aa C40:5      | 0.317                | 0.009   | -                    | -       | -                    | -       |
| PC aa C42:2      | -                    | -       | 0.255                | 0.038   | -                    | -       |
| PC aa C42:5      | -                    | -       | 0.281                | 0.021   | -                    | -       |
| PC ae C34:1      | 0.245                | 0.046   | 0.275                | 0.024   | -                    | -       |
| PC ae C34:3      | -                    | -       | -                    | -       | -0.288               | 0.018   |
| PC ae C36:0      | -                    | -       | 0.336                | 0.005   | 0.252                | 0.04    |
| PC ae C36:1      | 0.249                | 0.042   | 0.284                | 0.02    | -                    | -       |
| PC ae C38:0      | 0.254                | 0.038   | 0.357                | 0.003   | -                    | -       |
| PC ae C38:3      | 0.318                | 0.009   | 0.355                | 0.003   | -                    | -       |
| PC ae C40:1      | 0.275                | 0.024   | 0.347                | 0.004   | -                    | -       |
| PC ae C40:4      | 0.27                 | 0.027   | 0.295                | 0.015   | -                    | -       |
| PC ae C40:5      | 0.249                | 0.042   | 0.305                | 0.012   | -                    | -       |
| PC ae C42:4      | 0.243                | 0.048   | -                    | -       | -                    | -       |
| SM (OH) C14:1    | 0.258                | 0.035   | -                    | -       | -                    | -       |
| Kynurenine / Trp | -0.395               | 0.001   | -0.33                | 0.006   | -0.267               | 0.029   |

**Appendix Table 5. Spearman correlation Rho values and P values reported for metabolites significantly ( $P \leq 0.05$ ) associated with different POPs ( $\Sigma$ PCBs, HCB and DDE) in cord serum in N=86 neonates with healthy birth outcome from the Rhea pregnancy cohort.**

| Metabolites     | $\Sigma$ PCBs        |                | HCB                  |                | DDE                  |                |
|-----------------|----------------------|----------------|----------------------|----------------|----------------------|----------------|
|                 | Spearman coefficient | <i>P</i> value | Spearman coefficient | <i>P</i> value | Spearman coefficient | <i>P</i> value |
| Aspartate       | -                    | -              | -                    | -              | -                    | -              |
| Glycine         | -                    | -              | -                    | -              | -                    | -              |
| Citrulline      | -                    | -              | -0.249               | 0.021          | -                    | -              |
| Lysine          | -                    | -              | -0.209               | 0.050          | -                    | -              |
| Serine          | -                    | -              | -0.210               | 0.050          | -                    | -              |
| Met-SO          | -                    | -              | 0.231                | 0.033          | -                    | -              |
| C3-DC (C4-OH)   | 0.24                 | 0.030          | -                    | -              | -                    | -              |
| C16:1           | -                    | -              | -                    | -              | 0.216                | 0.046          |
| C16-OH          | -                    | -              | -0.217               | 0.045          | -                    | -              |
| lysoPC a C28:0  | -                    | -              | -                    | -              | -                    | -              |
| PC aa C34:2     | 0.215                | .046           | -                    | -              | -                    | -              |
| PC aa C36:2     | 0.229                | .034           | -                    | -              | -                    | -              |
| PC aa C40:6     | -                    | -              | -0.219               | 0.043          | -                    | -              |
| PC ae C42:3     | -                    | -              | -0.218               | 0.044          | -                    | -              |
| Tyr/Phe         | -                    | -              | -                    | -              | 0.266                | 0.013          |
| Met-So/Met      | -                    | -              | 0.225                | 0.037          | -                    | -              |
| Total DMA / Arg | -0.213               | 0.050          | -                    | -              | 0.298                | 0.005          |

Appendix Table 6. List of 120 chemical standards detected

| Superclass | Class                        | Standard                               | monoisotopic mass | [M-H] <sup>-</sup> | Detected on RP Negative mode (ULTIMA QtoF) RT in water | [M+H] <sup>+</sup> | p value | Concentration in urine (according to HMDB) | Detected on RP Positive mode (ULTIMA QtoF) RT in water |
|------------|------------------------------|----------------------------------------|-------------------|--------------------|--------------------------------------------------------|--------------------|---------|--------------------------------------------|--------------------------------------------------------|
|            |                              | pantothenic acid, calcium salt hydrate | 115.13            | 114.1227           | 3.08                                                   | 116.1373           |         |                                            | 2.96                                                   |
|            |                              | hippurate                              | 179.0582432       | 178.0509432        | 4.08                                                   | 180.0655432        | 0.31    | 217 (28-610) umol/mol creatin              | 3.97                                                   |
|            |                              | inosine                                | 268.0807695       | 267.0734695        | 2.45                                                   | 269.0880695        | -2.1    | 0.28 +/- 0.15 umol/mol creatin             | 2.37                                                   |
|            |                              | isoleudine                             | 131.0946287       | 130.0873287        |                                                        | 132.1019287        | -1.73   | 3.05-16.78 umol/mol creatinin              | 1.52                                                   |
|            |                              | kynurenic acid                         | 189.0426          | 188.03529          | 3.78                                                   | 190.04989          | 1.16    | 2.6(0.9-4.1) umol/mol creatinin            | 3.77                                                   |
|            |                              | L-Leudine                              | 131.0946287       | 130.0873287        |                                                        | 132.1019287        | -1.52   | 4.888 (1.842-8.421) umol/mol               | 1.69                                                   |
|            |                              | N-acetyl-DL-aspartic acid              | 175.0480724       | 174.0407724        | 1.12                                                   | 176.0553724        | -0.79   | 10.3 (6.1-15.4) umol/mol crea              | 1.01                                                   |
|            |                              | N-acetyl-L-cysteine                    | 163.0303138       | 162.0230138        | 2.47                                                   | 164.0376138        | -0.03   | 0-0.13 umol/mol creatinine                 | 2.45                                                   |
|            |                              | phenylalanine                          | 165.0789786       | 164.0716786        | 2.75                                                   | 166.0862786        | -1.38   | 6.4 (3.5-11.2) umol/mol creati             | 2.64                                                   |
|            |                              | s-adenosyl-L-homocysteine              | 384.1215885       | 383.1142885        |                                                        | 385.1288885        | -2.37   |                                            | 2.08                                                   |
|            |                              | tryptophan                             | 204.0898776       | 203.0825776        | 3.44                                                   | 205.0971776        | -1.06   | 2.45-29.40 umol/mol creatinin              | 3.31                                                   |
|            |                              | tyrosine                               | 181.0738932       | 180.0665932        | 1.77                                                   | 182.081932         | -2.26   | 15.1 (6.1-23.2) umol/mol crea              | 1.59                                                   |
|            |                              | gamma-glutamyl-cysteine                | 250.0623423       | 249.0550423        |                                                        | 251.0696423        | -2.53   | 0.46 +/- 0.14 umol/mol creatinin           | 1.27                                                   |
|            |                              | glutathione, oxidised                  | 612.1519619       | 611.1446619        |                                                        | 613.1592619        |         |                                            | 1.89                                                   |
|            |                              | glutathione, reduced                   | 307.083806        | 306.076506         | 1.12                                                   | 308.091106         | -4.9    | 0.030 (0.005-0.085) umol/mol               | 1.02                                                   |
|            | Diazines                     | Thymine                                | 126.0429274       | 125.0356274        |                                                        | 127.0502274        | -0.62   | 0.10 (0.045-0.16) umol/mol cr              | 2.27                                                   |
|            | Diazines                     | Uracil                                 | 112.0272774       | 111.0199774        |                                                        | 113.0345774        | -1.07   | 12.0 (2.0-22.0) umol/mol crea              | 1.05                                                   |
|            | Pyridine                     | Niacinamide                            | 122.0480128       | 121.0407128        |                                                        | 123.0553128        | -0.37   | 0.30 +/- 0.51 umol/mol creatin             | 1.06                                                   |
|            | Pyridines and derivatives    | Pyridoxine                             | 169.0738032       | 168.0665932        | 1.4                                                    | 170.081932         | -0.77   | <1 mmol/mol creatinine                     | 1.27                                                   |
| Aromatic   | Diazines                     | 1,3-Dimethyluracil                     | 140.0585775       | 139.0512775        |                                                        | 141.0658775        |         |                                            | 3.02                                                   |
|            | flavonoids                   | Rutin                                  | 610.153381        | 609.146081         | 4.56                                                   | 611.160681         |         |                                            | 4.55                                                   |
|            | Imidazopyrimidin es          | Hypoxanthine                           | 136.0385108       | 135.0312108        | 1.35                                                   | 137.0458108        | -1.11   | 7.2 (1.8-24.1) umol/mol creati             | 1.33                                                   |
|            | Imidazopyrimidin es          | Uric Acid                              | 168.02834         | 167.02104          |                                                        | 169.03564          | -2.17   | 1244.00 (565.5-1923.00) umol/l             | 1.25                                                   |
|            | Imidazopyrimidin es          | Xanthine                               | 152.0334254       | 151.0261254        |                                                        | 153.0407254        | -0.73   | 2.9 (1.3-4.6) umol/mol creatin             | 1.62                                                   |
|            | Indoles                      | 5-Hydroxy-3-indoleacetic acid          | 177.0789786       | 176.0716786        | 1.62                                                   | 178.0862786        | 0.91    |                                            | 3.96                                                   |
|            | Indoles                      | Indoxyl Sulfate potassium              | 250.965454        | 249.958154         | 4.44                                                   | 251.972754         |         |                                            | 4.45                                                   |
|            | Pteridines and derivatives   | riboflavin                             | 376.1382844       | 375.1309844        |                                                        | 377.1455844        | -1.46   | 0.18 +/- 0.08 umol/mol creatin             | 4.2                                                    |
|            | Benzoic acid and derivatives | 3-Hydroxanthranilic Acid               | 153.042587        | 152.035287         |                                                        | 154.049887         | 0.81    | 3.28 (0.46-5.19) umol/mol cre              | 3.22                                                   |
|            | Benzoic acid and derivatives | phthalic acid                          | 166.0266087       | 165.0193087        | 4.14                                                   | 167.0339087        | 0.73    | 30.5 (22.8-98.3) umol/mol cre              | 4.13                                                   |
|            | Benzoic acid and derivatives | Salicylic Acid                         | 138.0316941       | 137.0243941        | 5.44                                                   | 139.0389941        | 2.26    | 1.1 umol/mol creatinine                    | 5.44                                                   |
|            | Benzoic acid and derivatives | Vanillic Acid                          | 168.0422587       | 167.0349587        |                                                        | 169.0495587        | 1.43    | 1.6 (0.6-3.3) umol/mol creatin             | 4.21                                                   |
|            | Cinnamic acid                | 3-Hydroxycinnamic Acid                 | 164.0473441       | 163.0400441        | 4.95                                                   | 165.0546441        | 1.71    | 0.032 +/- 0.059 umol/mol crea              | 4.83                                                   |
|            | Phenethylamines              | Tyramine                               | 137.084064        | 136.076764         |                                                        | 138.091364         | -0.14   | 0.237 (0.2-0.275) umol/mol cr              | 1.89                                                   |
|            | Phenols                      | 4-hydroxy-3-methoxymandelic            | 198.0528234       | 197.0455234        | 2.67                                                   | 199.0601234        | 0.94    | 10.0 (0.3-20.0) umol/mol crea              | 2.51                                                   |

|               |                   |                                      |             |             |           |             |       |                          |                     |       |
|---------------|-------------------|--------------------------------------|-------------|-------------|-----------|-------------|-------|--------------------------|---------------------|-------|
|               | Phenols           | homovanillic acid                    | 182.0579088 | 181.0506088 | 4.32      | 183.0652088 | 0.33  | 7.5 (3.6-14.6)           | umol/mmol creati    | 4.3   |
|               | phenylacetic acid | Homogentisic acid                    | 168.0422587 | 167.0349587 | 2.94      | 169.0495587 | 0.86  | 1.0 (0.0-2.0)            | umol/mmol creatin   | 2.92  |
|               | phenylacetic acid | o-hydroxyphenylacetic acid           | 152.0473441 | 151.0400441 | 4.56      | 153.0546441 | 0.85  | 2.0 (0.9-4.5)            | umol/mmol creatin   | 4.55  |
|               | phenylacetic acid | p-Hydroxyphenylacetic Acid           | 152.0473441 | 151.0400441 |           | 153.0546441 | 0.75  | 6.0 (2.4-9.7)            | umol/mmol creatin   | 3.42  |
|               | Monosaccharides   | N-acetyl-DL-glutamic acid            | 189.0637225 | 188.0564225 | 1.61      | 190.0710725 | -0.67 | 0.5 (0.0-1.0)            | umol/mmol creatin   | 1.58  |
| Lipid         | fatty acid        | 2-Hydroxy-2-methylbutyric Acid       | 118.0629942 | 117.0556942 | 3.14      | 119.0702942 | 0.13  | 0.8 (0.4-1.5)            | umol/mmol creatini  | 3.12  |
|               |                   | ethylmalonic acid                    | 132.0422587 | 131.0349587 | 3.11      | 133.0495587 | 0.43  | 4.8 (0.1-17.9)           | umol/mmol creati    | 2.96  |
|               |                   | fumaric acid                         | 116.0109586 | 115.0036586 | 1.53?     | 117.0182586 | 0.46  | 1.0 (0.0-2.0)            | umol/mmol creatin   | 1.51  |
|               |                   | Glutaconic Acid                      | 130.0266087 | 129.0193087 |           | 131.0339087 | 0.05  | 3.1 (1.2-3.1)            | umol/mmol creatin   | 2.6   |
|               |                   | 3-Methylglutaric Acid                | 146.0579088 | 145.0506088 | 3.44      | 147.0652088 | 0.33  | 1.0 (0.1-2.8)            | umol/mmol creatin   | 3.42  |
|               |                   | Palmitic acid (sodium salt)          | 278.4       | 277.3927    |           | 279.4073    |       |                          |                     | 8.97  |
|               |                   | C3; Propionylcarnitine Chloride      | 217.1314    | 216.1241    |           | 218.1387    | -2.27 | 0.07 (0.01-0.20)         | umol/mmol cre       | 2.18  |
|               |                   | isoC4; isobutyryl-L-Carnitine Chlori | 231.1470582 | 230.1397582 |           | 232.1543582 |       |                          |                     | 2.98  |
|               |                   | C4; Butyryl-L-Carnitine Chloride     | 231.1470582 | 230.1397582 |           | 232.1543582 | -2.1  | 0.76 (0.12-1.81)         | umol/mmol cre       | 3.05  |
|               |                   | isoC5; isovaleryl-L-Carnitine Chlori | 245.1627082 | 244.1554082 |           | 246.1700    | -1.9  | 0.220 (0.037-0.440)      | umol/mmol           | 3.79  |
|               |                   | C6; Hexanoyl carnitine               | 259.17835   | 258.17105   |           | 260.18565   | -1.68 | 0.05 (0.03-0.07)         | umol/mmol cre       | 4.53  |
|               |                   | C12; Dodecanoylcarnitine             | 343.2722    | 342.2649    |           | 344.2795    | 0.03  | 0.02 (0.01-0.04)         | umol/mmol cre       | 6.84  |
|               |                   | C14; Myristoyl-L-Carnitine Chloric   | 371.3035    | 370.2962    |           | 372.3108    | 0.86  | 0.048 (0.001-0.088)      | umol/mmol           | 7.38  |
|               |                   | C16; Palmitoyl-L-Carnitine Chloric   | 399.33485   | 398.32755   |           | 400.34215   | 1.77  | 0.0022 (0.0020-0.0030)   | umol/n              | 7.83  |
|               |                   | C18; 1; Oleoylcarnitine              | 425.350509  | 424.343209  |           | 426.357809  |       |                          |                     | 7.94  |
|               |                   | C18; Stearoyl-L-Carnitine Chlorid    | 427.3661    | 426.3588    |           | 428.3734    |       |                          |                     | 8.24  |
|               |                   | cytidine                             | 243.0855205 | 242.0782205 | 0.99      | 244.0928205 | -2.51 | 1.50 +/- 1.34            | umol/mmol creatin   | 1     |
|               |                   | thymidine                            | 242.0902716 | 241.0829716 | 2.87      | 243.0975716 | -0.93 | 2.2 (0.7-3.9)            | umol/mmol creatin   | 2.74  |
|               |                   | uridine                              | 244.0685361 | 243.0622361 |           | 245.0768361 | -1.98 | 0.994 (0.637-1.351)      | umol/mmol           | 1.65  |
| Organic acids | aromatic acid     | -3-hydroxy-3-phenylpropionic a       | 166.0629942 | 165.0556942 | 4.59      | 167.0702942 | 1.83  | 1.00 (0.00-2.00)         | umol/mmol cre       | 4.58  |
|               | carboxylic acids  | citrate                              | 192.0270026 | 191.0197026 | 1.24/1.52 | 193.0343026 | -1.64 | 972.5 (89.5-1647.7)      | umol/mmol           | 1.09  |
|               |                   | Sebacic Acid                         | 202.1205091 | 201.1132091 | 5.67      | 203.1278091 | 1.93  | 9.5 +/- 12.8             | umol/mmol creatinic | 5.66  |
|               |                   | Suberic Acid                         | 174.0882089 | 173.0819089 | 4.68      | 175.0965089 | 1     | 16.85 +/- 2.73           | umol/mmol creati    | 4.66  |
|               |                   | trans-aconitic acid                  | 174.0164379 | 173.0091379 | 2.07      | 175.0237379 | -0.41 | 6.6 (1.8-25.1)           | umol/mmol creati    | 2.05  |
|               |                   | adipic acid                          | 146.0579088 | 145.0506088 | 3.37      | 147.0652088 | 0.08  | 9.2 (0.1-19.9)           | umol/mmol creati    | 1.08  |
|               |                   | Azelaic Acid                         | 188.104859  | 187.097559  | 5.19      | 189.112159  | 1.57  | 2.5 (0.1-12.3)           | umol/mmol creati    | 5.187 |
|               |                   | Glutaric Acid                        | 132.0422587 | 131.0349587 | 2.62      | 133.0495587 | -0.29 | 3.58 +/- 0.87            | umol/mmol creatin   | 2.6   |
|               |                   | Methylmalonic Acid                   | 118.0266087 | 117.0193087 | 1.98      | 119.0339087 | 0.17  | 0.64 +/- 0.39            | umol/mmol creatin   | 1.94  |
|               |                   | Methylsuccinic Acid                  | 132.0422587 | 131.0349587 | 2.98      | 133.0495587 | 0.13  | 2.8 umol/mmol creatinine |                     | 2.96  |
|               | Hydroxy Acids     | 2-Hydroxybutyric Acid                | 104.0473441 | 103.0400441 |           | 105.0546441 | -0.16 | 2.6 (2.6-9.9)            | umol/mmol creat     | 1.97  |
|               |                   | 4-Hydroxybutyric acid Na             | 126.09      | 125.0827    | 1.45      | 127.0973    |       |                          |                     | 1.32  |
|               |                   | 2-Oxovaleric acid                    | 116.0473441 | 115.0400441 | 2.95      | 117.0546441 | 0.35  | 0.45 +/- 0.02            | umol/mmol creatin   | 2.8   |

|            |  |                            |           |           |      |           |      |                     |                   |           |
|------------|--|----------------------------|-----------|-----------|------|-----------|------|---------------------|-------------------|-----------|
| Bile acids |  | Cholic acid                | 408.2875  | 407.2802  |      | 409.2948  | 2.02 | 2.2 (0.13-6.6)      | umol/mmol creati  | 6.99      |
|            |  | Deoxycholic acid           | 392.29265 | 391.28535 |      | 393.29995 | 3.5  | 0.055 (0.05-0.060)  | umol/mmol c       | 7.84      |
|            |  | Glycochenodeoxycholic acid | 449.3141  | 448.3068  |      | 450.3214  |      |                     |                   | 7.07      |
|            |  | Glycocholic acid           | 465.309   | 464.3017  |      | 466.3163  | 1.65 | 0.42 +/- 0.36       | umol/mmol creatin | 6.33      |
|            |  | Glycodeoxycholic acid      | 449.3141  | 448.3068  |      | 450.3214  |      |                     |                   | 7.2       |
|            |  | Hyocholic acid             | 408.2875  | 407.2802  |      | 409.2948  | 2.72 |                     |                   | 6.77      |
|            |  | Hyodeoxycholic acid        | 392.29265 | 391.28535 |      | 393.29995 | 3.08 | 0.0056 (0.0-0.011)  | umol/mmol c       | 7.1       |
|            |  | Lithocholic acid           | 376.2977  | 375.2904  |      | 377.305   | 4.38 |                     |                   | 8.91      |
|            |  | Taurochenodeoxycholic acid | 499.2967  | 498.2894  |      | 500.304   | 1.38 | 0.00065 +/- 0.0019  | umol/mmol c       | 6.8       |
|            |  | Taurocholic acid           | 515.2916  | 514.2843  | 6.49 | 516.2989  | 0.79 |                     |                   | 6.06-6.1  |
|            |  | Taurodeoxycholic acid      | 499.2967  | 498.2894  |      | 500.304   |      |                     |                   | 6.93-7.00 |
|            |  | Taurolithocholic acid      | 483.3018  | 482.2945  |      | 484.3091  |      |                     |                   | 7.9-8.02  |
|            |  | Ursodeoxycholic acid       | 392.29265 | 391.28535 |      | 393.29995 | 3    | 0.022 (0.013-0.030) | umol/mmol         | 7.1       |