# The discovery and evaluation of metabolic and epigenetic biomarkers in the diagnosis of epithelial ovarian cancer

**Thesis submitted by**

Kayleigh Rebecca Davis

**For the degree of Doctor of Philosophy of Imperial College London and the Diploma of Imperial College London**

**Imperial College London**

**Department of Surgery & Cancer**

**2017**

## Declaration of Originality

I certify that this thesis, and the research to which it refers, are the product of my own work, unless otherwise stated and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

## 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 no alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

## Abstract

*Background:* Epithelial ovarian cancer (EOC) is often diagnosed late in the advanced stages of the disease with an associated poor prognosis. A large scale multicentre randomised control screening trial, the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS), failed to show a significant reduction in mortality in the screening groups using trans-vaginal ultrasound and cancer antigen 125 (CA 125) compared to the no screening group. More sensitive biomarkers for diagnosis of ovarian cancer are a clinical priority. Molecular profiling of 'liquid biopsies' has the potential to detect changes associated with the tumour in non-invasively collected blood samples. This study focusses on the use of epigenomics, specifically DNA methylation and metabonomics to identify tumour-specific changes and their performance as diagnostic and prognostic biomarkers in non-invasively collected biofluid.

*Methods:* Differential methylation analysis using the Illumina Infinium® Methylation BeadChip was interrogated to identify candidate methylation biomarkers specific to EOC tumour DNA compared to peripheral blood mononuclear cells (PBMCs). Bisulphite pyrosequencing was used to characterise these biomarkers in tumour and blood.

Metabolic profiling techniques including nuclear magnetic resonance (NMR) and mass spectrometry (MS) of blood and urine were used to identify metabolic changes associated with EOC diagnosis and survival.

*Results:* Nine loci differentially methylated between tumour and PBMC DNA, were identified. Methylation of 5/9 candidate loci were shown to be significantly different in prospectively collected plasma samples from patients with high grade serous ovarian cancer (HGSOC) compared to benign/healthy controls with one marker achieving an area under the curve (AUC) of 0.901.

Metabonomic analysis of plasma samples identified alterations in amino acid, ketone body and lipid metabolism associated with EOC. Several of these metabolites including tryptophan and beta-hydroxybutyrate were also shown to be significantly associated with overall survival (OS) with hazard ratios of 0.39 (95% CI 0.18-0.83) and 2.76 (95% CI 1.28-5.97) respectively.

*Conclusions:* EOC tumour-specific DNA methylation can be detected in tumour and plasma of EOC cancer patients, providing insight into the biology of the disease and potential as novel diagnostic biomarkers. Metabolic profiling has uncovered a number of candidate EOC biomarkers associated with prognosis. Prospective validation of these biomarkers could help EOC treatment stratification in the future.

## Acknowledgments

Firstly, I would like to thank my supervisors, Professor Jeremy Nicholson for providing this opportunity through the StratiGrad programme and Professor Bob Brown for his continuous guidance and support. I would like to acknowledge the StratiGrad programme for funding this PhD studentship.

I would also like to thank my fellow StratiGrad students within CSM and members of the Epigenetics unit, past and present. In particular to Dr Kirsty Flower, my post-doc mentor for always going above and beyond this role and to Alun and Nair for continuously providing comic relief, advice and support.

I also acknowledge and thank collaborators in this project, specifically the National Phenome Centre, Dr James Flanagan, members of Dr Hector Keun and Dr Sadaf Ghaem-Maghami's groups and the ECMC and ICHTB, specifically Naina Patel, Nona Rama, Nahal Masrour and Katherine Costelloe for all your support in sample and clinical data collection. Thank you to Dr Paula Cunnea and Dr Ed Curry for all your help and guidance during the project and to Dr Katherine Nixon, Dr David Phelps and Dr John Wahba for sample collection.

A special mention to the women, volunteers and patients at Hammersmith/Queen Charlotte's and Chelsea Hospitals as without your selfless donation, this and other clinical research is simply not possible. Thank you.

Thank you to my new colleagues at RBHT, especially Dr Jackie Donovan for giving me the time to complete this thesis.

Thank you to my family and friends and the Goodwin family for continuous encouragement and support, who I have no doubt will be pleased to see me detached from my laptop. To Mum and Dad for continuously encouraging me to go further, making me believe that I can achieve it and helping me be the person I am today. Lastly to Ned, for unwavering support even in the hardest of times and longest of days, your kindness and patience is indefatigable.

# Table of contents

Declaration of Originality ..... 2
Copyright Declaration ..... 2
Abstract ..... 3
Acknowledgments ..... 4
Table of contents ..... 5
Abbreviations ..... 9
List of tables ..... 10
List of figures ..... 12
CHAPTER 1. Introduction ..... 16
1.1 General Introduction ..... 17
1.2 Epidemiology ..... 17
1.3 Classification of tumours ..... 18
1.3.1 Staging ..... 18
1.3.2 Grading ..... 19
1.4 Clinical challenges ..... 20
1.4.1 Diagnosis and screening ..... 20
1.4.2 Prognostic and predictive biomarkers ..... 22
1.5 Cancer systems biology – the 'omics' era ..... 23
1.5.1 Circulating tumour DNA; the 'liquid biopsy' ..... 25
1.5.2 Epigenomics ..... 27
1.5.3 Metabonomics ..... 32
1.6 Summary and Hypothesis ..... 53
1.7 Aims ..... 53
1.71 Methylation of cell-free circulating DNA in plasma ..... 53
1.72 Metabolic profiling of plasma and urine ..... 53
CHAPTER 2. Materials and Methods ..... 54
Methods ..... 55
2.1 Study design and strategy ..... 55
2.2 Systematic literature search for metabolites identified in metabonomics-based analysis of ovarian cancer patients ..... 55
2.3 Clinical data and sample collection ..... 56
2.3.1 TCGA HGSOC dataset & Mayo clinic – 'the discovery cohort' ..... 56
2.3.2 HH cohort & PBMC cohort – 'the verification cohort' ..... 56
2.3.3 Charité and New HH cohort HGSOC dataset – 'the validation cohort' ..... 57

2.3.4 Blood and urine samples ..... 57
2.4 Epigenomics ..... 58
2.4.1 Clinical material processing ..... 58
2.4.2 Pyrosequencing ..... 59
2.4.3 Real time-PCR ..... 61
2.5 Metabonomics ..... 63
2.5.1 $^{1}$H Nuclear Magnetic Resonance (NMR) Spectroscopy ..... 63
2.5.2 Targeted MS (Biocrates AbsoluteIDQ® p150 Kit) ..... 65
2.6 Statistical analysis ..... 67
2.6.1 Multivariate statistical analysis ..... 67
2.6.2 Univariate statistical analysis ..... 67
2.6.3 Survival analysis ..... 68
Materials ..... 69
2.7 Chemicals, kits and buffers ..... 69
DNA extraction ..... 69
DNA quantification ..... 69
Electrophoresis ..... 69
Bisulphite conversion ..... 69
PCR ..... 69
Pyrosequencing ..... 69
Real-Time PCR ..... 70
$^{1}$H NMR ..... 70
MS ..... 70
2.8 Equipment & plasticware ..... 70
DNA quantification ..... 70
Electrophoresis ..... 70
Pyrosequencing ..... 71
Real-Time PCR ..... 71
$^{1}$H NMR ..... 71
MS ..... 71
General Equipment & Plasticware ..... 71
CHAPTER 3. Cell-free circulating DNA methylation as a blood based diagnostic biomarker of epithelial ovarian cancer ..... 73
3.1 Introduction ..... 74

3.2 Background ..... 74
3.3 Results ..... 75
3.3.1 Discovery of differentially methylated loci between HGSOC and PBMCs ..... 75
3.3.2 Verification of differentially methylated candidate loci ..... 80
3.3.3 Development of pyrosequencing assays for the differentially methylated candidates ..... 84
3.3.4 Pyrosequencing of differentially methylated candidates in an independent cohort of tumour and PBMC samples ..... 87
3.3.5 Pyrosequencing of differentially methylated candidates in prospectively collected blood samples ..... 90
3.3.6 Exploring the relationship between the 9 candidate CpG sites ..... 95
3.3.7 Correlation of methylation and expression of CpG Island associated ZNF154 and HOXA9 ..... 96
3.3.8 Methylation of ZNF154 (cg21790626) as a pan-cancer biomarker ..... 98
3.3.9 Methylation of ZNF154 (cg21790626) and the association with survival ..... 100
3.4 Discussion ..... 103
3.4.1 Differentially methylated loci as diagnostic biomarkers of serous ovarian cancer ..... 103
3.4.2 Differentially methylated loci between serous tumours and benign ovarian tissue and their association with gene expression ..... 103
3.4.3 Bisulphite pyrosequencing is an appropriate analytical technique for methylation detection in solid tissue ..... 104
3.4.4 ZNF154 as a pan-cancer diagnostic and prognostic biomarker ..... 105
3.4.5 Biological function of candidate diagnostic biomarkers ..... 106
3.4.6 Bisulphite pyrosequencing of ctDNA methylation in plasma can diagnose high grade serous ovarian cancer ..... 107
Summary ..... 109
CHAPTER 4. $^{1}$H NMR-based metabonomic profiling of plasma and urine to identify metabolites associated with ovarian cancer ..... 110
4.1 Introduction ..... 111
4.2 Background ..... 111
4.3 Results ..... 111
4.3.1 Characteristics of the patients in the study ..... 111
4.3.2 $^{1}$H CPMG NMR analysis of plasma samples from EOC patients and controls ..... 113
4.3.3 Survival analysis ..... 133
4.3.4 $^{1}$H NOESY NMR analysis of plasma samples from EOC patients and controls ..... 137
4.3.5 LipoFIT® analysis of lipoprotein subclasses in plasma samples from EOC patients and controls ..... 154

4.3.6 $^{1}$H NOESY NMR analysis of urine samples from EOC patients and controls ..... 158
4.4 Discussion ..... 160
4.4.1 $^{1}$H NMR metabolic signatures in the diagnosis of epithelial ovarian cancer ..... 160
4.4.2 Beta-hydroxybutyrate as a diagnostic and prognostic biomarker of EOC ..... 162
4.4.3 Altered amino acid metabolism in ovarian cancer ..... 163
4.4.4 Altered lipid metabolism in ovarian cancer ..... 164
4.5 Summary ..... 165
CHAPTER 5. Targeted MS-based quantitative profiling of plasma to identify metabolites associated with epithelial ovarian cancer ..... 166
5.1 Introduction ..... 167
5.2 Background ..... 167
5.3 Results ..... 168
5.3.1 Characteristics of the patients in the study ..... 168
5.3.2 Analysis of plasma samples from EOC patients and controls using the Biocrates AbsoluteIDQ® p150 Kit ..... 168
5.3.3 Survival analysis ..... 185
5.3.4 Predictive capacity of combining metabolites from analysis of plasma using $^{1}$H NMR and the Biocrates AbsoluteIDQ® p150 targeted assay for the diagnosis of EOC ..... 192
5.4 Discussion ..... 194
5.4.1 Metabolic signature for the diagnosis of epithelial ovarian cancer ..... 194
5.4.2 Metabolites associated with cancers of non-ovarian origin ..... 196
5.4.3 Acetyl carnitine is associated with EOC, stage of disease and survival ..... 197
5.4.4 Altered lipid metabolism in ovarian cancer and association with survival ..... 197
Altered amino acid metabolism in ovarian cancer and association with survival ..... 198
5.5 Summary ..... 199
CHAPTER 6. General Discussion ..... 201
6.1 Circulating tumour DNA methylation biomarkers and future directions ..... 202
6.1.1 Limitations and future work ..... 202
6.2 Metabolic biomarkers and future directions ..... 203
6.2.1 Limitations and future work ..... 205
6.3 Biomarker study limitations ..... 206
6.4 Concluding remarks ..... 207
Chapter 7. References ..... 208
Chapter 8. Appendix 1 ..... 225
Appendix 2 ..... 228

## Abbreviations

| | |
|---|---|
| AIC | Akaike information criterion |
| AUC | Area under the curve |
| BOT | Benign ovarian tumour |
| CA 125 | Cancer Antigen 125 |
| cfDNA | Cell free circulating DNA |
| CGI | CpG island |
| CNS | Central nervous system |
| CNV | Copy number variation |
| CPMG | Carr-Purcell-Meiboom-Gill |
| CpG | C (cytosine) base followed by G (guanine) base in DNA linked by a phosphodiester bond (p) |
| CRC | Colorectal cancer |
| ctDNA | Circulating tumour DNA |
| CV | Coefficient of variation |
| D2HG | D-2-Hydroxyglutarate |
| DMCpG | Differentially methylated CpG |
| DNA | Deoxyribonucleic acid |
| EOC | Epithelial ovarian cancer |
| EGFR | Epidermal growth factor receptor |
| FACHS | Family and Community Health Study |
| FDR | False discovery rate |
| FFPE | Formalin-fixed paraffin-embedded |
| FIA | Flow injection analysis |
| FN | False negative |
| FP | False positive |
| GC/-MS | Gas chromatography/-mass spectrometry |
| GSEA | Gene set enrichment analysis |
| HBA | Hydroxybutyric acid |
| HE4 | Human epididymis protein 4 |
| HGSOC | High grade serous ovarian cancer |
| HH | Hammersmith Hospital |
| HMDB | Human Metabolome Database |
| HR | Hazard ratio |
| HRT | Hormone replacement therapy |
| ICHTB | Imperial College Healthcare Tissue Bank |
| IDH | Isocitrate dehydrogenase |
| LC/-MS | Liquid chromatography/-mass spectrometry |
| LOD | Limit of detection |
| LOOCV | Leave-one-out cross-validation |
| LTR | Long term reference |
| MALDI/-TOF | Matrix-assisted laser desorption ionisation/-time of flight |
| MGMT | $O^6$-methylguanine-DNA methyltransferase |
| miRNA | Micro RNA |
| MMS | Multimodal screening |
| MRM | Multiple reaction monitoring |
| mRNA | Messenger RNA |
| MS | Mass spectrometry |
| MSP | Methylation-specific PCR |
| NMR | Nuclear magnetic resonance |
| NOESY | Nuclear Overhauser effect spectroscopy |
| NPC | National Phenome Centre |
| NPV | Negative predictive value |
| OC | Ovarian cancer |
| OPLS/-DA | Orthogonal projections to latent structures/-discriminant analysis |
| OS | Overall survival |
| PBMC | Peripheral blood mononuclear cells |
| PCA | Principal component analysis |
| PCR | Polymerase chain reaction |
| PFS | Progression free survival |
| PLS/-DA | Partial least squares/-discriminant analysis synonymous with projection to latent structures |
| Ppm | Parts per million |
| PPV | Positive predictive value |
| QC | Quality control |
| ROCA | Risk of ovarian cancer algorithm |
| ROC | Receiver operating characteristic |
| RD | Residual disease |
| RNA | Ribonucleic acid |
| RT-PCR | Real-time PCR |
| SR | Sample reference |
| STOCSY | Statistical total correlation spectroscopy |
| TCGA | The Cancer Genome Atlas |
| TN | True negative |
| TOF | Time of flight |
| TPR | True positive rate |
| UKCTOCS | UK Collaborative Trial of Ovarian Cancer Screening |
| VIP | Variable importance in projection |

# List of tables


| | | Page |
|---|---|---|
| Table 1.1. | The FIGO staging system used in ovarian cancer | 19 |
| Table 1.2. | Table of publications and summary of findings of metabonomics-based analysis in body fluid to identify potential biomarkers of epithelial ovarian cancer | 43-52 |
| Table 2.1. | Primer sequences of candidate loci for PCR and pyrosequencing | 59 |
| Table 2.2. | Primer sequences of real time PCR | 62 |
| Table 2.3. | Table for calculation of the performance of a classification model, also known as a confusion matrix | 68 |
| Table 3.1. | Clinical characteristics of study participants used in the discovery cohort | 77 |
| Table 3.2. | Details of the nine differentially methylated candidate cg sites in tumour (TCGA) and PBMCs (Mayo clinic) datasets | 78 |
| Table 3.3. | Clinical characteristics of study participants in the verification cohort | 81 |
| Table 3.4. | Details of the nine differentially methylated candidate cg sites in the verification cohort | 82 |
| Table 3.5. | Clinical characteristics of study participants used in the candidate validation study | 88 |
| Table 3.6. | Details of the nine differentially methylated candidate cg sites in the Charité cohort, New HH cohort and PBMCs using bisulphite pyrosequencing | 89 |
| Table 3.7. | Clinical characteristics of study participants used in the candidate validation study of prospectively collected blood samples from patients undergoing gynaecological surgery at Hammersmith Hospital and non-affected volunteers | 91 |
| Table 4.1. | Clinical characteristics of participants used in the ovarian NMR study | 112 |
| Table 4.2. | Table of peaks from CPMG $^{1}$H NMR spectra of plasma samples that were integrated, the chemical shift and upper and lower boundaries for integration | 124 |
| Table 4.3. | Table of the %CV values for the peak integration in the CPMG $^{1}$H NMR spectra of the QC samples | 125 |
| Table 4.4. | Confusion matrix for the performance of the cross-validated step-wise logistic regression model for the 3 metabolite integrals from CPMG-$^{1}$H NMR data with significantly different levels in plasma from patients with EOC and without | 130 |
| Table 4.5. | Coefficient of variation (%) between biological replicates for significantly different metabolite integrals from CPMG-$^{1}$H-NMR data used in stepwise logistic regression model | 132 |
| Table 4.6. | Table of peaks from NOESY $^{1}$H NMR spectra of plasma samples that were integrated, the chemical shift and upper and lower boundaries for integration | 147 |
| Table 4.7. | Table of the %CV values for the peak integration in the NOESY $^{1}$H NMR spectra of the QC samples | 148 |
| Table 4.8. | Confusion matrix for the performance of the cross-validated step-wise logistic regression model for the 2 metabolite integrals from NOESY-presat-$^{1}$H NMR data with significantly different levels in plasma from patients with EOC and without | 151 |
| Table 4.9. | Coefficient of variation (%) between biological replicates for significantly different metabolite integrals from NOESY-presat $^{1}$H-NMR data used in stepwise logistic regression model | 153 |

| | | |
|---|---|---|
| Table 5.1. | Clinical characteristics of participants used in the ovarian Biocrates study | 168 |
| Table 5.2. | Confusion matrix for the performance of the cross-validated step-wise logistic regression model for 5 metabolites with significantly different levels in plasma of patients with EOC and without of the Biocrates data | 182 |
| Table 5.3. | Coefficient of variation (%) between biological replicates for significantly different metabolites used in stepwise logistic regression model of the Biocrates data | 184 |
| Table 5.4. | Summary of 6 metabolites from the Biocrates Absolute*IDQ*® p150 targeted assay that are associated with survival | 189 |
| Table 5.5. | Confusion matrix for the performance of the cross-validated step-wise logistic regression model for the 5 metabolites in the combined model consisting of histidine and unknown.1.81ppm integrals from CPMG-$^{1}$H NMR, glycoprotein acetyls integral from NOESY-presat-$^{1}$H NMR and PC.ae.C44.4 and SM.C20.2 from the Biocrates Absolute*IDQ*® p150 Kit with significantly different levels in plasma from patients with EOC and without | 193 |
| Supplementary Table 2.1. | List of metabolites from the Biocrates Absolute*IDQ*® p150 kit | 228-231 |
| Supplementary Table 2.2. | List of lipoprotein subfractions from $^{1}$H NMR analysis of plasma samples | 232-234 |

# List of figures


| | | Page |
|---|---|---|
| Figure 1.1. | The 'omics-cascade' | 24 |
| Figure 1.2. | Signalling pathways that regulate cancer metabolism | 33 |
| Figure 1.3. | Simplified diagram of an NMR spectrometer | 36 |
| Figure 1.4. | Workflow for experimental and data analysis of samples by 1H NMR Spectroscopy | 41 |
| Figure 2.1. | Flowchart of project study design | 55 |
| Figure 3.1. | Flowchart of study design for cfDNA methylation biomarker identification | 75 |
| Figure 3.2. | Flowchart of candidate probe selection in the discovery cohort | 76 |
| Figure 3.3. | Box and whisker plots of the candidate cg sites in the discovery cohort | 79 |
| Figure 3.4. | Box and whisker plots of the candidate cg sites in the verification cohort | 83 |
| Figure 3.5. | Scatterplot of the candidate cg sites in DNA extracted from cell-lines using bisulphite pyrosequencing | 84 |
| Figure 3.6. | Graphical representation of methylation of serially diluted DNA using bisulphite pyrosequencing | 86 |
| Figure 3.7. | Correlation of cg21790626 methylated detected by bisulphite pyrosequencing and Illumina Infinium methylation array in the HH cohort | 87 |
| Figure 3.8. | Box and whisker plots of the candidate cg site methylation in the validation cohort of HGSOC tumours and PBMCs from patients with ovarian tumours using bisulphite pyrosequencing | 90 |
| Figure 3.9 | Box and whisker plots of the differentially methylated candidate cg sites in prospectively collected blood samples from patients with HGSOC (cancer) and benign pathologies or non-affected volunteers (controls) | 93 |
| Figure 3.10. | Box and whisker plots of the significantly differentially methylated candidate cg sites in prospectively collected blood samples from patients with HGSOC | 94 |
| Figure 3.11. | Correlation of 9 candidate hits | 96 |
| Figure 3.12. | Scatterplot showing the correlation of DNA methylation of cg21790626 in the *ZNF154* gene determined by bisulphite pyrosequencing and expression of *ZNF154* relative to *GAPDH* determined by RT-qPCR | 97 |
| Figure 3.13. | Scatterplot showing the correlation of DNA methylation of cg27009703 in the *HOXA9* gene determined by bisulphite pyrosequencing and expression of *HOXA9* relative to *GAPDH* determined by RT-qPCR | 98 |
| Figure 3.14. | Box and whisker plots of methylation of *ZNF154* (cg21790626) across all tissue types from TCGA | 99 |
| Figure 3.15. | Kaplan-Meier survival plots for OS and PFS in HH dataset of all primary tumours | 102 |
| Figure 4.1. | Comparison of $^{1}$H CPMG NMR plasma spectra | 113 |
| Figure 4.2. | Averaged $^{1}$H NMR CPMG plasma spectra highlighting differences in lipid resonances (0.84ppm) and beta-hydroxybutyrate (1.2ppm) | 114 |
| Figure 4.3. | Averaged $^{1}$H NMR CPMG plasma spectra highlighting differences in glycoprotein acetyls at 2.04ppm and beta-hydroxybutyrate doublet of doublets at 2.31ppm and 2.41ppm | 114 |

| | | |
|---|---|---|
| Figure 4.4. | Unsupervised PCA scores plot of all plasma samples (n=200) and IQC samples: SR (n=8) and LTR (n=8) in the $^{1}$H NMR CPMG data | 115 |
| Figure 4.5. | Unsupervised PCA scores plot of all samples (n=200) and IQC samples (n=16) coloured by rack number (3-6) in the $^{1}$H NMR CPMG data | 116 |
| Figure 4.6. | Unsupervised PCA scores plot of plasma samples coloured according to class in the $^{1}$H NMR CPMG data | 117 |
| Figure 4.7. | Unsupervised PCA scores plot of plasma samples from patients with EOC (n=67), benign ovarian disease (n=55) and healthy volunteers (n=14) in the $^{1}$H CPMG NMR data | 118 |
| Figure 4.8. | Supervised OPLS-DA scores plot of plasma samples from patients with EOC (n=67) and the 'Normal' group (n=69) combining patients with benign gynaecological disease (n=55) and healthy volunteers (n=14) in the $^{1}$H NMR CPMG data | 119 |
| Figure 4.9. | Permutation plot of 100 permutations of OPLS model between EOC and 'Normal' groups in the $^{1}$H CPMG NMR data of plasma samples | 120 |
| Figure 4.10. | Correlation coefficient plot corresponding to $^{1}$H CPMG NMR OPLS-DA model of EOC vs. Normal controls of plasma samples | 121 |
| Figure 4.11. | STOCSY analysis of a driver peak at δ1.20 | 122 |
| Figure 4.12. | ROC curve for 'Normal' versus EOC by metabolic signature based on $^{1}$H NMR CPMG data of plasma samples | 123 |
| Figure 4.13. | Metabolite integrals from CPMG-$^{1}$H NMR data that are significantly different between the EOC and 'Normal' groups, FDR-adjusted | 126 |
| Figure 4.14. | Metabolite integrals from CPMG-$^{1}$H NMR data that are significantly different between the EOC and 'Normal' groups | 127 |
| Figure 4.15. | Box and whisker plots of metabolite integrals from CPMG-$^{1}$H NMR data that show an association between stage (Low = Stage I,II and High = Stage III,IV) and metabolite integral, in the EOC group only | 128 |
| Figure 4.16. | Correlation plot of the 13 metabolite integrals from CPMG-$^{1}$H NMR data between the EOC and 'Normal' groups | 129 |
| Figure 4.17. | ROC curve for non-malignant 'Normal' versus EOC by cross-validated prediction score based on step-wise logistic regression model of 3 metabolites from the CPMG-$^{1}$H NMR data | 131 |
| Figure 4.18. | Kaplan-Meier survival plots for OS and PFS in the NMR ovarian study of patients with EOC | 134-135 |
| Figure 4.19. | Kaplan-Meier survival curve for OS dependent on beta-hydroxybutyrate concentration | 136 |
| Figure 4.20. | Kaplan-Meier survival curve for PFS dependent on the peak integration of metabolite feature at 2.39ppm | 137 |
| Figure 4.21. | Comparison of $^{1}$H NOESY NMR plasma spectra | 138 |
| Figure 4.22. | Averaged $^{1}$H NMR NOESY plasma spectra highlighting differences in glycoprotein acetyls (2.04ppm) and choline phospholipids (3.215ppm) | 139 |
| Figure 4.23. | Unsupervised PCA scores plot of all samples (n=199) and IQC samples; SR (n=8) and LTR (n=8) in the $^{1}$H NMR NOESY data | 140 |
| Figure 4.24. | Unsupervised PCA scores plot of all samples (n=199) and IQC samples (n=16) coloured by rack number (3-6) in the $^{1}$H NMR NOESY data | 141 |
| Figure 4.25. | Unsupervised PCA scores plot of plasma samples coloured according to class in the $^{1}$H NMR NOESY data | 142 |
| Figure 4.26. | Unsupervised PCA scores plot of plasma samples from patients with EOC (n=67), benign ovarian disease (n=55) and healthy volunteers (n=14) in the $^{1}$H NOESY NMR data | 143 |
| Figure 4.27. | Supervised OPLS-DA scores plot of plasma samples from patients with | 144 |

| | | |
|---|---|---|
| | EOC (n=67) and 'Normal' group (n=69) combining patients with benign gynaecological disease and healthy volunteers in the $^{1}$H NMR NOESY data | |
| Figure 4.28. | Permutation plot of 100 permutations of OPLS-DA model between EOC and 'Normal' groups in the $^{1}$H NOESY NMR study of plasma samples | 145 |
| Figure 4.29. | Correlation coefficient plot corresponding to $^{1}$H NOESY NMR OPLS-DA model of EOC vs. Normal controls of plasma samples | 146 |
| Figure 4.30. | ROC curve for 'Normal' versus EOC by metabolic signature based on $^{1}$H NMR NOESY data | 147 |
| Figure 4.31. | Metabolite integrals from NOESY-presat-$^{1}$H NMR data that are significantly different between EOC and 'Normal' groups, FDR adjusted | 149 |
| Figure 4.32. | Box and whisker plots of metabolite integrals from NOESY-presat-$^{1}$H NMR data that show an association between stage (Low = Stage I,II and High = Stage III,IV) and metabolite integral, in the EOC group only | 150 |
| Figure 4.33. | ROC curve for non-malignant 'Normal' versus EOC by cross-validated prediction score based on step-wise logistic regression model of 2 metabolites from NOESY-presat-$^{1}$H NMR data | 152 |
| Figure 4.34. | Metabolite integrals from lipoprotein sub-fractions data that are significantly different between EOC and 'Normal' groups | 155 |
| Figure 4.35. | Box and whisker plots of lipoprotein sub-fractions data that show an association between stage (Low = Stage I,II and High = Stage III,IV) and metabolite integral, in the EOC group only | 156 |
| Figure 4.36. | Correlation plot of the 35 lipoprotein sub-fractions that were significantly different between the EOC and 'Normal' groups | 157 |
| Figure 4.37. | Supervised OPLS-DA scores plot of urine samples from patients with EOC (n=18) and 'Normal' group (n=32) combining patients with benign gynaecological disease and healthy volunteers in the $^{1}$H NMR NOESY data | 159 |
| Figure 4.38. | Permutation plot of 20 permutations of the OPLS-DA model between the EOC and 'Normal' groups in the $^{1}$H NOESY NMR data of urine samples | 160 |
| Figure 5.1. | Unsupervised PCA scores plot of all samples (n=246), standards (n=4), zero samples (n=12) and IQC samples (n=44) of the Biocrates data | 169 |
| Figure 5.2. | Unsupervised PCA scores plot of all samples and IQC samples coloured by plate number (3-6) of the Biocrates data | 170 |
| Figure 5.3. | Unsupervised PCA scores plot of plasma samples coloured according to class of the Biocrates data | 171 |
| Figure 5.4. | Unsupervised PCA scores plot of plasma samples from patients with EOC, benign gynaecological disease and healthy volunteers of the Biocrates data | 172 |
| Figure 5.5. | Supervised PLS-DA scores plot of plasma samples from patients with EOC and benign ovarian disease and Healthy volunteers, grouped to 'Normal' of the Biocrates data | 173 |
| Figure 5.6. | Supervised PLS-DA loadings plot of plasma samples from patients with EOC and benign ovarian disease and Healthy volunteers, grouped to 'Normal' of the Biocrates data | 174 |
| Figure 5.7. | Variable importance in the projection (VIP) plot summarising the importance of the variables to explain X and to correlate to Y (EOC/Normal) in the PLS-DA model of the Biocrates data | 175 |

| | | |
|---|---|---|
| Figure 5.8. | ROC curve for non-malignant 'Normal' and EOC groups by metabolic signature based on the Biocrates Absolute*IDQ*® kit of 122 metabolites | 176 |
| Figure 5.9. | Metabolites from Biocrates targeted assay that are significantly different between EOC and Benign groups | 177 |
| Figure 5.10. | Metabolites from Biocrates targeted assay that are significantly different between EOC and normal groups | 178 |
| Figure 5.11. | Box and whisker plots of metabolites that show an association between stage (Low = Stage I, II and High = Stage III, IV) and metabolite concentration, in the EOC group only of the Biocrates data | 179 |
| Figure 5.12. | Box and whisker plots of metabolites that are show an association between grade (Low = Grade I,II and High = Grade III, Undifferentiated and metabolite concentration, in the EOC group only of the Biocrates data | 180 |
| Figure 5.13. | Correlation plot of the 30 significantly different metabolites between the EOC and normal groups of the Biocrates data | 181 |
| Figure 5.14. | ROC curve for non-malignant 'Normal' versus EOC by cross-validated prediction score based on step-wise logistic regression model of 5 metabolites of the Biocrates data | 183 |
| Figure 5.15. | Kaplan-Meier survival plots for OS and PFS in the Biocrates ovarian study of patients with EOC. | 187-188 |
| Figure 5.16. | Kaplan-Meier survival curves for OS dependent on metabolite concentration of the Biocrates Absolute*IDQ*® p150 targeted assay | 190-191- |
| Figure 5.17. | Kaplan-Meier survival curve for PFS dependent on lysoPC.a.C18.0 concentration | 192 |
| Figure 5.18. | ROC curve for non-malignant 'Normal' versus EOC by cross-validated prediction score based on step-wise logistic regression model of 5 metabolites from CPMG-$^{1}$H NMR, NOESY-presat-$^{1}$H NMR and the Biocrates Absolute*IDQ*® p150 Kit | 194 |
| Figure 6.1. | A schematic illustrating the metabolism of hydroxybutyric acid metabolites and succinate in metabolic and signal transduction pathways | 204 |
| Supplementary Figure 1.1. | Kaplan-Meier curves for OS and PFS in the TCGA dataset | 225 |
| Supplementary Figure 1.2. | Kaplan-Meier curves for OS and PFS in the Charité dataset | 226 |
| Supplementary Figure 1.3. | Kaplan-Meier curves for OS and PFS in the New HH dataset | 227 |

# CHAPTER 1. Introduction

## 1.1 General Introduction

Ovarian cancer (OC) is the sixth most common cancer occurring in women in the UK and the fifth most common cause of cancer death in women. It is the most lethal of all gynaecological malignancies. The average 5 year survival rate for women diagnosed with ovarian cancer in the UK is 46% however this drastically drops to just 4% for patients diagnosed with Stage IV disease (Cancer Research Uk). More women are diagnosed with late stage disease (55-58% of patients at stage III/IV) than early stage (Cancer Research Uk). This late presentation and associated poor survival can be attributed to the subtle and non-specific symptoms that only manifest in the late stages of the disease and the lack of effective non-invasive screening strategies (Badgwell and Bast, 2007).

The identification of robust biomarkers for early detection, when therapeutic intervention is most effective, is a clinical priority to reduce morbidity and mortality from the disease. The rapid development of systems oncology and our understanding of the 'omics-cascade' including genomics, epigenomics and metabolomics has provided the opportunity to use highly sensitive large-scale profiling techniques to further our understanding of systems oncology and identify specific biomarkers for disease diagnosis, prognosis, monitoring response to treatment and detection of relapse.

## 1.2 Epidemiology

In the UK in 2014, 7,378 women were diagnosed with OC and it accounted for 4,128 deaths in the same year (Cancer Research Uk). Incidence of ovarian cancer is strongly linked to age, with the higher incidences occurring among older women. The peak incidence rate is in those aged 75-79, with an average of 53% of cases diagnosed in women aged 65 and over (Cancer Research Uk).

Tumours of the ovary can be grouped into three distinct histological categories; surface epithelial, sex cord and germ cell tumours, although each of these has their own histological sub-types (Chen *et al.*, 2003). Sex cord tumours are thought to originate from the theca cells, granulosa and sertoli and leydig cells. They account for 7% of all malignant ovarian tumours (Chen *et al.*, 2003) and are often associated with endocrine manifestations and therefore present with abnormal vaginal bleeding. Approximately 90% of patients are diagnosed at Stage I and the reported 10 year survival rate is 86-96% (Chen *et al.*, 2003).

Germ cell tumours such as dysgerminoma, immature teratomas and yolk sac tumours are thought to originate from the primordial germ cells. They are rare and account for 3-7% of malignant ovarian tumours (Chen *et al.*, 2003). Surface epithelial tumours were previously thought to originate from surface epithelium of the ovary and account for 90% of all malignant ovarian tumours and 60% of all ovarian tumours including benign and borderline tumours (Chen *et al.*, 2003). High grade serous

ovarian cancer (HGSOC) is now thought to originate from epithelium of the fallopian tube (Kim *et al.*, 2012). Benign tumours or masses, lack invasive behaviour and profuse cell proliferation. Borderline tumours are generally atypically proliferating and have low malignant potential (Chen *et al.*, 2003). Surface epithelial tumours can be further subdivided into 5 major subtypes: serous, mucinous, endometrioid, clear cell and transitional cell (or Brenner type) and some patients can have a mixed subtype. Tumours can also be classified as undifferentiated if they lack any specific differentiation. The most common type of surface epithelial ovarian cancer is the serous histological subtype that present late in life with associated poor survival (Rosen *et al.*, 2010). It can be further subdivided based on the extent of nuclear atypia and mitosis into low and high grade disease (Malpica *et al.*, 2004).

There are a number of risk factors associated with ovarian cancer and these are dependent on histology and subtype. Of the cases of OC diagnosed in the UK every year, 21% are linked to major lifestyle and other risk factors such as smoking, use of post-menopausal hormone replacement therapy (HRT) and reproductive history (breastfeeding) (Parkin *et al.*, 2011). The most important risk factors of OC development are age and known genetic mutations. Ovarian cancer risk is up to 65% higher in women with *BRCA1* mutation and up to 35% higher in women with *BRCA2* mutation (Ingham *et al.*, 2013). The majority of hereditary cases of ovarian cancer are linked to *BRCA1/2* mutations and 5-15% of OC cases are due to inherited conditions (Lynch *et al.*, 2013). Lynch syndrome, also known as hereditary non-polyposis colorectal cancer (HNPC), is another risk factor for development of OC; approximately 10-15% of hereditary ovarian carcinoma cases are due to this inherited disorder that is caused by alterations in mismatch repair genes such as *MLH1, MSH2, MSH6* and *PMS2 (Lynch et al., 2009)*. OC risk can also be affected by reproductive factors that affect the number of (and breaks between) ovulations. Structural changes caused by ovulation can lead to development of OC and therefore factors that decrease the number of ovulations such as multiparity (having more children), breastfeeding, lower age at menopause and oral contraceptive pill (OCP) use can be protective against OC (Tsilidis *et al.*, 2011). Associated hormonal factors may compound this or have independent effects (Lukanova and Kaaks, 2005); an estimated 1% of ovarian cancer cases in the UK are linked to HRT use (Parkin *et al.*, 2011).

## 1.3 Classification of tumours

### *1.3.1 Staging*

The extent of tumoural spread, also referred to as stage of disease is established by radiological evaluation and surgery. Surgical management of the disease in the UK consists of 'debulking' of the tumour through resection of both of the ovaries and fallopian tubes (bilateral salpingoophrectomy),

removal of the uterus (total abdominal hysterectomy), the omentum (omentectomy) with or without removal of the pelvic and paraaortic lymph nodes (lymphadenectomy). This can occur before chemotherapy (primary debulking), during chemotherapy treatment (interval debulking) or after completion of chemotherapy or to treat recurrence (secondary debulking). Currently in the UK, the FIGO (Federation of gynaecology and obstetrics) staging system is used, shown in Table 1.1.

Table 1.1 The FIGO staging system used in ovarian cancer Adapted from (Prat, 2015).

| **FIGO stage** | **Description** |
|---|---|
| **1** | **Tumour limited to ovaries** |
| 1A | - Tumour limited to one ovary (capsule intact) or fallopian tube.<br>- No tumour on surface of ovary.<br>- No malignant cells in ascites or peritoneal washings. |
| 1B | - Tumour limited to both ovaries (capsules intact) or fallopian tubes.<br>- No tumour on surface of ovary.<br>- No malignant cells in ascites or peritoneal washings. |
| IC | - Tumour limited to one or both ovaries or fallopian tubes with any of the following:<br>o Capsule ruptured<br>o Tumour on surface of ovary<br>o Malignant cells in ascites or peritoneal washings. |
| **II** | **Tumour involves one or both ovaries or fallopian tubes with pelvis extension or primary peritoneal cancer** |
| IIA | - Extension and/or implants on uterus and/or fallopian tube(s).<br>- No malignant cells in ascites of peritoneal washings. |
| IIB | - Extension to other pelvic intraperitoneal tissues.<br>- No malignant cells in ascites or peritoneal washings. |
| IIC | - Pelvic extension (IIA or IIB) with malignant cells in ascites or peritoneal washings. |
| **III** | **Tumour involves one or both ovaries or fallopian tubes or primary peritoneal with microscopically confirmed peritoneal metastasis outside the pelvis and/or regional lymph node metastasis.** |
| IIIA | - Microscopic peritoneal metastasis beyond pelvis. |
| IIIB | - Macroscopic peritoneal metastasis (<2cm in greatest dimension) beyond pelvis. |
| IIIC | - Peritoneal metastasis (>2cm in greatest dimension) beyond pelvis and/or regional lymph node metastasis. |
| **IV** | **Distant metastasis (excluding peritoneal metastasis)** |

### *1.3.2 Grading*

Low and high grade ovarian carcinomas vary at the genomic and molecular levels; high grade serous tumours exhibit mutations in *p53* whereas these are rare in low-grade carcinoma (Singer *et al.*, 2005). The grade of disease refers to differentiation of the tumour or the extent to which tumours

resemble the normal tissue. Tumour grade is determined microscopically and is included in the assessment of histology and stage of the tumour. Microscopic examination can predict behaviour of the tumour which can guide therapeutic strategies. Tumours that are well differentiated are Grade 1 (G1), moderately differentiated are Grade 2 (G2), poorly differentiated are Grade 3 (G3) or undifferentiated (G4).

## 1.4 Clinical challenges

### *1.4.1 Diagnosis and screening*

The development of robust, non-invasive, sensitive and specific biomarkers for early detection and prognostication of ovarian cancer remains a clinical challenge. Since the introduction of serum based monitoring of CA 125 as a biomarker for epithelial ovarian cancer in 1981, monitoring of this high molecular weight glycoprotein is now considered the gold standard (Bast *et al.*, 1981). CA 125 levels of less than 35 U/ml are considered normal (Bast *et al.*, 1983) and it is raised in approximately 90% of patients with advanced stage epithelial ovarian cancer, but in only 50% of patients with Stage I disease (Bast *et al.*, 2002). Reported sensitivity and specificity of CA 125 therefore varies but is estimated as 61% to 90% and 35% to 91% respectively for distinguishing benign from malignant masses (Pepin *et al.*, 2014). CA 125 is also raised in several benign conditions such as endometriosis, liver cirrhosis and peritonitis (Sarojini *et al.*, 2012). The use of CA 125 is therefore suggested for monitoring patients with a confirmed diagnosis of ovarian cancer; however the majority of tests requested are in patients for evaluation with or without a suspicious mass. Surprisingly, the survival benefit associated with CA 125 either before or after diagnosis has never been reported (Ueland, 2017). Nevertheless, advances in the use of CA 125 have taken place with the development of a risk of ovarian cancer algorithm (ROCA) that identifies a change-point in CA 125 based on longitudinal serial measurements of the biomarker (Skates *et al.*, 2001). To assess the impact of this risk algorithm as a screening tool for early detection of ovarian cancer, it would need to be included in a prospective screening trial for ovarian cancer.

An effective screening test would improve early stage detection as surgical resection could potentially cure the disease ultimately leading to a reduction in mortality. To address this issue a randomised controlled trial; the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS) was initiated. Between 2001 and 2005, 202,638 post-menopausal women were recruited and randomly allocated to multimodal screening (MMS) with annual serum CA 125 interpreted with the risk of ovarian cancer algorithm (ROCA), USS (annual transvaginal ultrasound screening) or to no screening in a 1:1:2 ratio (Menon *et al.*, 2015). Median follow up was 11.1 years for all groups. During the trial, they reported that the use of ROCA within the MMS group (comparing CA 125 levels to the patient's

baseline level) compared to a fixed cut-off doubled the number of screen-detected invasive epithelial ovarian cancers (Menon *et al.*, 2015).

The relative reduction in mortality was 15% (95% CI -3 to 30, p=0.1) in the MMS group and 11% (-7 to 27, p=0.21) in the USS group. Although not statistically significant in the primary analysis, authors admitted that initial study design would have benefited from an extended follow-up as well as altered statistical design due to the late effect of screening on mortality seen in other trials of this nature (Jacobs *et al.*, 2015). In addition, for each ovarian and peritoneal cancer detected by screening, an additional 2 women in the MMS group and 10 women in the USS group had false positive surgery. For firm conclusions to be reached regarding implementation of a screening test for ovarian cancer, further evidence on the efficacy and cost-effectiveness are needed.

A very recent publication from this group using data from UKCTOCS has assessed the within-trial cost-effectiveness of an NHS ovarian cancer screening programme however, the results are inconclusive as the results required estimation of the cost of the ROCA test within the NHS, extrapolation of the mortality benefit and extrapolation over the lifetime of the patients (Menon *et al.*, 2017). Until the true mortality benefit of screening post-menopausal women for ovarian cancer is known, implementation of a public health programme for ovarian cancer screening is still to be realised.

OVA1, developed by Vermillion Inc and licensed to Quest Diagnostics, is a panel of biomarkers approved in the US for triaging patients with a pelvic mass that combines CA 125, transthyretin, apolipoprotein A1, beta-2 microglobulin and transferrin. It has been shown to outperform CA 125 in terms of increased sensitivity and negative predictive value (which are important for screening markers), but the specificity and positive predictive value decreased (Ware Miller *et al.*, 2011). Another multivariate index assay, the Risk of Ovarian Malignancy Algorithm (ROMA®) consisting of CA 125 and human epididymis protein 4 (HE4) were also FDA-approved for the same indications (Moore *et al.*, 2011). In 2016, another multivariate index assay was approved combining markers from both ROMA® and OVA1, the Overa® test combines CA 125, HE4, apolipoprotein A1, follicle stimulating hormone (FSH) and transferrin. The result is a more sensitive marker with improved specificity compared to OVA1 (Coleman *et al.*, 2016). However, these tests are only approved once a mass has been detected and the need for surgery has already been agreed. This ensures the assay causes no tangible harm through unnecessary surgery due to false positive results and is used to determine the extent of surgery required and if a gynaecological oncologist is necessary, therefore improving the quality of patient care and ultimately survival. The trend of multivariate index assays looks set to continue with a recent publication of a 4 marker panel including CA 125 and HE4 in a

cohort of patients with early stage disease achieving a highest sensitivity of 83.2% with 98% specificity (Simmons *et al.*, 2016).

### *1.4.2 Prognostic and predictive biomarkers*

The two most important prognostic factors for epithelial ovarian cancer are stage of disease and presence or absence of residual disease after treatment intervention. It is widely accepted that volume of residual disease following surgical intervention influences overall survival (Shih and Chi, 2010). This appears to be independent of type of chemotherapy and dose administered (Bristow *et al.*, 2002). The definition of optimal cytoreduction, defined by the Gynaecologic Oncology Group (GOG) states that optimal cytoreduction is residual disease less than or equal to 1cm and suboptimal refers to residual disease greater than 1cm (Hoskins *et al.*, 1994). However, increasing evidence suggests that this definition of 'optimal' should refer to the absence of any residual disease post-surgery as total macroscopic debulk is considered one of the most important factors (Chi *et al.*, 2006, Shih and Chi, 2010). Wimberger *et al.*, showed a significantly improved OS in patients with complete resection in Stage IV disease compared to visible tumour including minimal residual disease. Median OS in the patients with complete resection was 54.6 months whereas patients with residual tumour of 1-10mm had a median OS of 25.8 months and patients with >10mm residual tumour had a median OS of 23.9 months (Wimberger *et al.*, 2010). There appears to be no significant difference in the association between residual disease status and survival among histological subtypes (Melamed *et al.*, 2017).

Although volume of residual disease, specifically total macroscopic debulk, will remain as one of the most important prognostic factors, this can only be determined after surgical intervention. Prognostic markers, those that can estimate the changes of disease progression, or predictive markers, those associated with response to a particular therapy, would aid clinicians in identifying patients who would most benefit from radical surgery and therefore facilitate treatment choice on an individual personalised basis.

CA 125 has been the most extensively studied and clinically utilised tumour marker. It is widely accepted that CA 125 is a predictive and prognostic factor in CA 125 positive ovarian cancers. An exhaustive number of studies have assessed the utility of CA 125 as a prognostic biomarker of survival; assessing the marker pre- and post-operatively, pre-and post-chemotherapy, during chemotherapy, the half-life, the nadir, the time to reach nadir and longitudinal CA 125 values (comprehensively evaluated in (Gupta and Lis, 2009)). This comprehensive evaluation on the role of CA 125 in predicting ovarian cancer survival concluded that although results from a number of studies were sometimes contradictory, CA 125 levels are a strong prognostic factor of OS and

progression free survival (PFS) in ovarian cancer (Gupta and Lis, 2009). However, a definitive method for prediction of OS and PFS in ovarian cancer using CA 125 is yet to be agreed upon and there is still no robust mechanism to decide which patients will benefit from a particular treatment.

The emergence and rapid growth of studies using genomics, proteomics and other 'omics' has led to a new era of biomarker research with the discovery of hundreds of new prognostically valuable EOC biomarkers (Huang *et al.*, 2010). The most relevant of which include biomarkers of angiogenesis and related signalling pathways such as VEGF (Paley *et al.*, 1997) and EphA2 (Thaker *et al.*, 2004), however the number of clinically useful biomarkers is still pitifully small (Huang *et al.*, 2010). Therefore new biomarkers either alone or in combination are warranted to increase sensitivity and specificity and improve ovarian cancer diagnosis and prognosis.

### 1.5 Cancer systems biology – the 'omics' era

Systems oncology or cancer systems biology aims to understand cancer as an integrated system of complex networks and interactions rather than an isolated system of genes, proteins and metabolites (Archer *et al.*, 2016). The field of –omics based analysis of the complex mechanisms of cancer covers analysis of the genome through high-throughput DNA and RNA sequencing to mass spectrometry analysis of proteins and metabolites which together culminate in an individual phenotype (Buescher and Driggers, 2016). This 'omics-cascade' is shown in Figure 1.1. It utilises rapidly developing analytical and computational methods to gain insight into these complex biological mechanisms and future development will be crucial in identifying biomarkers and implementing targeted patient therapy to those who are most likely to benefit, based on their tumour biology (Werner *et al.*, 2014).

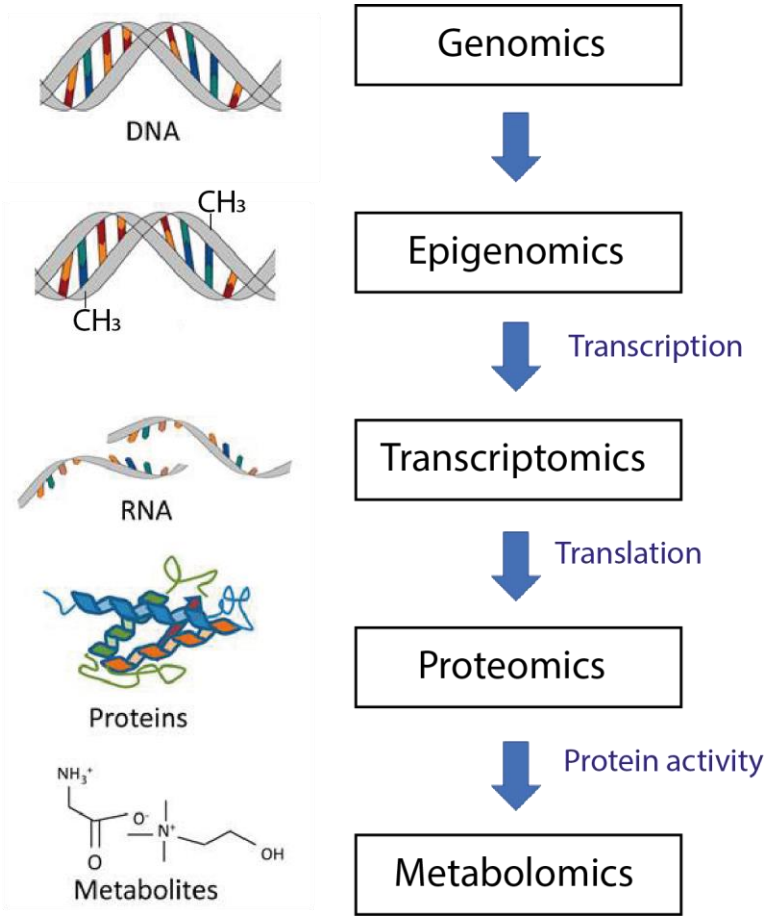



Figure 1.1. 'The omics-cascade'. Adapted with permission from (Haukaas *et al.*, 2017)

Heterogeneity of tumours, both between patients but also within a single tumour is an undeniable barrier to accurate diagnosis and targeted therapy. Sequential multiple biopsies of tumours or metastases is invasive, expensive and time consuming and not without risk. Lung biopsies are associated with a significantly morbidity (Kaarteenaho, 2013) and therefore multiple biopsies are not feasible. A recent study from TRACERx (Tracing Non-Small-Cell Lung Cancer Evolution Through Therapy (Rx)) showed that phylogenetic profiling of tumours DNA within the blood, so-called circulating tumour DNA (ctDNA) could track the subclonal dynamics of relapsing non-small cell lung cancer (NSCLC) and metastasis and identify adjuvant chemotherapy resistance (Abbosh *et al.*, 2017). Findings from this study indicate that monitoring of residual disease and adjuvant treatment response is not only possible but it can be guided by ctDNA from the bloodstream, which can be assessed at multiple times relatively simply and non-invasively during the course of treatment. Development of therapies could therefore target emerging subclones before clinical recurrence occurs.

### *1.5.1 Circulating tumour DNA; the 'liquid biopsy'*

The first report of DNA circulating in the blood was in 1948 (Mandel and Metais, 1948) and specifically in the blood of cancer patients in 1977 (Leon *et al.*, 1977). It wasn't until 1994 that it was proven to originate from the tumour (Sorenson *et al.*, 1994, Vasioukhin *et al.*, 1993). However, the first practical use of monitoring DNA in the blood was in prenatal screening of foetal DNA within the maternal bloodstream (Lo *et al.*, 1997). This has revolutionised the field of non-invasive prenatal testing of conditions such as trisomy 21 in Down syndrome (Norton *et al.*, 2015).

Tumour DNA, in contrast, is more difficult to detect due to the variable and potentially very small fraction of the total circulating DNA found in the blood, sometimes less than 0.01% (Diehl *et al.*, 2005). The amount may be dependent on type and stage of the cancer. Highly sensitive techniques are therefore required to detect and quantify small amounts of DNA. Furthermore, circulating tumour cells (CTCs) and exosomes have recently been identified to be a source of tumour DNA and are attractive as potential biomarkers; in particular tumour exosomes which can harbour tumour-specific mutations and have been shown to play a role in metastasis (Costa-Silva *et al.*, 2015).

Over the previous 10 years, highly sensitive techniques have been developed to monitor ctDNA in blood. Diehl *et al*., combined real time polymerase chain reaction (PCR) to measure the total DNA concentration and bead-based digital PCR in emulsions termed 'BEAMing' to measure the percentage of mutant DNA, however, this requires previous knowledge of the mutations harboured by the tumour via a biopsy (Diehl *et al.*, 2006).

Mutations in the tumour suppressor gene, *TP53* have been shown to occur in >97% of cases of high grade serous ovarian cancer (Ahmed *et al.*, 2010). Further developments in sequencing led Nitzen Rosenfeld's group to use targeted deep sequencing ('Tam-Seq) to track multiple mutations in *TP53* in circulating DNA (Forshew *et al.*, 2012). Although appropriate for advanced cancers, this cost-effective sequencing method lacks sensitivity for routine clinical use for ctDNA detection.

A large study by Bottegowda *et al*. used digital PCR to assess the ability to detect ctDNA in a cohort of patients covering a number of different malignancies including pancreatic, ovarian, CRC, head and neck. ctDNA was detected in >75% of patients with ovarian cancer. In metastatic CRC, sensitivity for testing clinically relevant mutations in *KRAS* was 87.2% and specificity was 99.2% in patients with metastatic CRC proving its utility as a non-invasive alternative to tumour biopsy. They showed that ctDNA from tumours confined to the central nervous system (CNS) were frequently undetectable suggesting that the blood brain barrier may prevent crossing of ctDNA to the bloodstream. Ninety-six percent of patients who responded to epidermal growth factor receptor (EGFR) blockage but subsequently relapsed were found to have developed one or more mutations in genes involved in

the mitogen-activated protein kinase pathway (Bettegowda *et al.*, 2014). This study suggests longitudinal studies of ctDNA to monitor patients could identify resistance mutations and personalise therapeutic approaches.

Circulating tumour DNA detection has shown to be more sensitive than the current clinically used protein biomarkers such as CA153 for breast cancer (Dawson *et al.*, 2013). This study also showed that ctDNA detection was more sensitive than detecting circulating tumour cells (CTCs). Furthermore, ctDNA has a very short half-life within the circulation, approximately 2 hours, providing a clearer view of the current state of the tumour compared to most protein biomarkers which can persist in the circulating for weeks (Diehl *et al.*, 2008, Riedinger *et al.*, 2006).

Exome sequencing is an alternative, unbiased approach to genotyping ctdna, requiring no prior knowledge of the tumour. Murtaza *et al*. sequenced the complete exome of patients with lung, ovarian and breast cancers and tracked markers of resistance through the course of treatment (Murtaza *et al.*, 2013). However, it is currently a very expensive approach to sequence every sample to the depth required to detect rare mutant fragments.

Newman *et al*., used a combined approach they termed CAPP-seq; a next generation sequencing approach covering multiple classes of somatic mutations to a high depth at a reasonable cost. CtDNA was detected in 100% of patients with Stage II-IV NSCLC but in only 50% of patients with Stage I (Newman *et al.*, 2014). These results suggest this method may have utility as a monitoring tool but the amount of circulating tumour DNA in early stage disease may limit its use as a screening tool. The question remains whether a more sensitive technique would provide the necessary sensitivity to detect the small number of ctDNA fragments within the bloodstream or are there just simply not enough fragments in the circulation to detect early stage disease.

Most recently, Phallen *et al*. applied their method, termed TEC-Seq (targeted error correction sequencing) to address this question in patients with Stage I or II colorectal, breast, lung and ovarian tumours as well as a population of healthy individuals. They were able to detect somatic mutations in ctDNA of 71, 59, 59 and 68% of patients respectively and no alterations were found in the plasma of the healthy individuals. It is unclear from the manuscript what percentage of Stage I patients in each cohort were accurately diagnosed. The specificity of their method is once again very high but the sensitivity is yet to improve substantially on the current methods and reach a level appropriate for routine clinical use (Phallen *et al.*, 2017).

Despite this, there are a number of other applications of the liquid biopsy that can be used to improve patient outcome including, but not limited to: assessment of molecular heterogeneity of an

individual's disease, monitoring tumour dynamics, identification of genetic determinants for targeted therapy, evaluation of response to therapy, monitoring minimal residual disease, assessment of evolution of resistance and detection of recurrence (Diaz and Bardelli, 2014). Introduction of these applications into routine clinical use will require evidence from clinical trials of the impact on patient care. The true utility will be answered when we can intervene based on ctDNA monitoring with more effective targeted therapies. Trials such as TracerX are providing a deeper understanding of how NSCLC is evolving which will hopefully provide evidence for development of these targeted therapies in the future.

Although analysis of somatic point mutations may be one of the most studied alterations in circulating DNA due to its high specificity, other alterations reflecting changes present in the tumour can also occur and therefore be utilised to detect tumour DNA in blood or other fluids including urine, saliva and cerebrospinal fluid (Fleischhacker and Schmidt, 2007). Tumour-derived rearrangements and copy number variation (Leary *et al.*, 2012, Mcbride *et al.*, 2010, Chan *et al.*, 2013), fragment size (Jiang *et al.*, 2015a) as well as epigenetic changes such as DNA methylation (Wong *et al.*, 1999) are such examples. Furthermore, the term 'liquid biopsy' shouldn't be confined to the idea of just alterations in DNA. The liquid biopsy represents changes in the bloodstream that reflect changes present in the tumour or as a consequences of it, which as discussed in Section 1.5, includes anything from the genome to the metabolome.

### *1.5.2 Epigenomics*

The term 'epigenetics' was first coined by Conrad Waddington in the early 1940s to describe changes to the cellular phenotype that were independent of changes to the DNA sequence (Waddington, 1942). This definition has been subsequently refined and we now understand that by adapting chromatin, the physiological form of DNA through epigenetic mechanisms heritable changes in gene expression can occur (Allis and Jenuwein, 2016).

Chromatin is made up of repeating units of nucleosomes, which consist of 147 bp of DNA wrapped around an octamer of four core histone proteins (H3, H4, H2A and H2B). Epigenetic mechanisms that modify chromatin structure include DNA methylation, covalent histone modifications, nucleosome positioning and non-coding RNAs including microRNAs (miRNAs) (Sharma *et al.*, 2010). These modifications collectively referred to as the 'epigenome', primarily regulate the compactness and accessibility of chromatin and are therefore key determinants of cell diversity and gene activity. Disruption or failure of the epigenetic mechanisms can result in altered gene function and inappropriate activation or inhibition of various signalling pathways and lead to disease states such as cancer (Egger *et al.*, 2004, Jones and Baylin, 2002). Global changes in the epigenome are a

hallmark of cancer. Once traditionally seen as a genetic disease, it is now accepted that cancer also involves epigenetic alterations (Sharma *et al.*, 2010).

#### 1.5.2.1 DNA methylation

One of the most extensively studied epigenetic processes is DNA methylation. DNA methylation occurs by the covalent modification of a methyl group, catalysed by DNA methyltransferases (DNMTs) to position 5 on carbon on the cytosine residues in CpG dinucleotides. CpG dinucleotides are clustered in short CpG-rich DNA stretches called 'CpG islands' (CGIs) and occupy approximately 60% of human gene promoters (Portela and Esteller, 2010).

It is widely understood that CGIs are unmethylated in transcriptionally active genes whilst those that are silent are characterised by hypermethylation at the promoter (Bird, 1986). The majority of CGIs remain unmethylated during development and in differentiated cells (Suzuki and Bird, 2008). However, methylation of CGI promoters during development can result in long-term transcriptional silencing such as in the case of X-chromosome inactivation and imprinted genes (Bird, 2002).

There are a number of proposed mechanisms for gene silencing by DNA methylation including acting as a physical barrier to transcription factors at promoter binding sites or recruitment of methyl-CpG binding domain proteins (MBDs) or 'readers' that are associated with an inactive compacted chromatin state via recruitment of histone deacetylases or 'erasers' (Deaton and Bird, 2011).

##### 1.5.2.1.1Aberrant DNA methylation and cancer

DNA methylation is involved in transcriptional regulation of genes and is essential for development and cell functioning, it is therefore not surprising that defects in this mechanism may lead to diseases such as cancer (Esteller, 2008). The first reports of epigenetic abnormalities in cancer were reported in the early 1980s. A reduction in DNA methylation was observed in specific genes in human colon cancer cells compared with normal tissues (Feinberg and Vogelstein, 1983) and Gama-Sosa et al. observed a global reduction in 5-methylcytosine content of DNA from tumour samples (Gama-Sosa *et al.*, 1983).

DNA hypomethylation or loss of DNA methylation of normally inactivated regions such as satellite sequences or repetitive sequences can lead to mutagenesis and genomic instability. Despite this global hypomethylation in cancer, localised hypermethylation of CGIs in regulatory regions can lead to loss of gene expression and this can occur in tumour suppressor genes (TSGs). For example, DNA methylation mediated silencing of two cell cycle related genes, *CDKN2A* and *CDKN2B*, involved in control of G1 progression, occurs in a number of different cancers (Drexler, 1998). Furthermore, various genes associated with DNA repair processes such as *MLH1* (Fleisher *et al.*, 1999) and *MGMT*

(0-6-methylguanine-DNMT) (Esteller *et al.*, 1999) have also been shown to be associated with hypermethylation in cancers.

#### 1.5.2.2 DNA methylation as a biomarker of cancer

As discussed, there is a clinical need for more sensitive and specific biomarkers not only for early detection of cancer but also prognosis, prediction of therapeutic response, monitoring of response to therapy and detection of relapse. There are a number of advantages of DNA methylation that make it a potential candidate for use as a biomarker. Changes in DNA methylation can occur early in tumour development and potentially even in precancerous lesions (Kanai, 2010) making it a candidate as an early detection biomarker or even to assess the risk of progression to malignancy. As well as early detection biomarkers, methylation could be used for prognostic or predictive biomarkers in response to treatment. As previously mentioned detection of these potential biomarkers in blood rather than multiple biopsies is non-invasive, cost-effective and associated with far less complications. However, to avoid false positives the tumour-specific methylation must be undetectable in not only normal tissue or benign tumours but also in peripheral blood mononuclear cells (PBMCs) or white blood cells which make up the majority of the DNA detected in the bloodstream. Lastly, compared to detection of mutation analysis, DNA methylation events are well defined and therefore provide a simpler prospect for detection and analysis. Whereas a number of mutations inactivating tumour-suppressor genes could be distributed across a number of sites throughout the gene, requiring more sensitive analysis, DNA methylation requires less sensitive methods to detect whether the gene is methylated or not and to what degree (Kulis and Esteller, 2010).

Current examples of DNA methylation biomarkers for various cancers include: *SEPT9* for colorectal cancer (Devos *et al.*, 2009), *CDKN2A/p16* (Belinsky *et al.*, 1998) and *SHOX2* for lung cancer (Schmidt *et al.*, 2010). However, many studies report DNA methylation in sputum and bronchiolar lavages and not in blood. Detection of methylation of $O^6$-methylguanine-DNA methyltransferase (MGMT), a DNA repair gene is one of the only predictive markers that has reached the clinic. In patients with glioblastoma, methylation of this gene is associated with susceptibility to the alkylating drug, temozolomide resulting in a survival benefit (Hegi *et al.*, 2005). Lastly, methylation biomarkers don't have to necessarily contribute to tumourigenesis but if it could be shown to be associated with tumour development it could act as a putative biomarker for diagnosis of cancer. One example of this is *Vimentin* in colorectal cancer. Methylation of this gene occurs in colorectal cancer but not in normal colorectal epithelial cryt cells and although it is not expressed in both cases (Chen *et al.*, 2005) it is associated with Duke's stage (Shirahata *et al.*, 2009).

Methylation of the promoter of *IFFO1* was identified as a promising sensitive diagnostic biomarker of ovarian cancer with a reported AUC of 0.95, however only 16 ovarian cancer patients and 2 normal controls were assessed (Campan *et al.*, 2011).

Despite this, there are still a small number of clinically available methylation biomarkers compared to the number of publications and this is in part due to the lack of consensus for an agreed methodology for DNA methylation detection for routine clinical use. Another barrier to implementation is the relatively small number of DNA molecules from the tumour in the blood. Compared to mutational analysis, where the mutation is present in only the DNA arising from the tumour, non-tumoural DNA can also be methylated. A signal from non-tumoural DNA will therefore be present in the blood, of which the majority will be from lysis of peripheral blood mononuclear cells (PBMCs). This could be overcome if a candidate CpG site was unmethylated in PBMCs compared to tumour DNA. For implementation into clinical use, larger studies with more patients and prospective trials are required to fully validate and prove clinical utility of these epigenetic biomarkers.

For the investigation of differential methylation in a number of tissues, advances in highly sensitive technologies to detect differential methylation at a genome wide and site-specific level have been developed.

#### 1.5.2.3 Technology for DNA methylation analysis

There are three main possible methods for methylation detection: 1) endonuclease digestion with methylation sensitive restriction enzymes, 2) affinity enrichment with chromatin immunoprecipitation assays and 3) bisulphite conversion with an appropriate detection method.

Endonuclease digestion utilises methylation sensitive restriction enzymes that can digest the target sequence if it is either methylated or not methylated, depending on the enzyme used, examples include: HpaII, SacII, NotI and BstUI. However, the method is restricted to CpG sites that are localised within the enzyme restriction site.

The affinity enrichment approach uses specific antibodies against MBD proteins (methyl-CpG-binding proteins) which specifically recognise 5-methylcytosines or against 5-methylcytosines specifically, allowing these methylated fragments to be immuno-precipitated for further hybridisation to a DNA microarray. This is a genome wide screening tool providing a map of the global distribution of methylation CpG sites however, large amounts of DNA are needed and it is restricted by the specificity of the antibody.

Bisulphite conversion is the most commonly used method to assess methylation. Bisulphite treatment of DNA converts only unmethylated cytosines to uracil, whereas methylated cytosines stay unchanged. PCR amplification converts uracil to thymine providing a different sequence for downstream detection methods using further amplification, sequencing or hybridisation.

Amplification methods such as methylation-specific PCR (MSP) (Herman *et al.*, 1996) can only detect only a signal from a small number of CpG sites in a region and the result is qualitative, development of quantitative MSP such as MethyLight has improved the sensitivity of the assay but cannot quantify methylation at individual CpG sites (Eads *et al.*, 2000). However, sequencing methods such as pyrosequencing can quantitate methylation at individual CpG sites.

Pyrosequencing utilises biotin labelled PCR products that are immobilised to streptavidin sepharose beads. After denaturation, the biotinylated single stranded DNA amplicon hybridises with a sequencing primer and is incubated with the enzymes: DNA polymerase, ATP sulfurylase, luciferase and apyrase and their substrates. Each addition of a deoxynucleotide (dNTP) complementary to the base in the template strand is accompanied by the release of pyrophosphate (PPi). This is converted to ATP which drives the conversion of luciferin to oxyluciferin which generates visible light which is detected by sensors and is proportional to the amount of ATP. This generates a peak in the raw data output (Pyrogram®) and the height is proportional to the number of nucleotides incorporated. Addition of dNTPs is performed sequentially and incorporated into the complementary DNA strand and then degraded prior to the next nucleotide dispensation. The nucleotide sequence is determined from the peak signals in the Pyrogram trace. Pyrosequencing software calculates the ratio of unmethylated cytosines (thymines) to methylated cytosines at each individual CpG site in the sequence, which is usually restricted to approximately 100bp. The EpiTYPER® and MassARRAY® system from Agena Biosciences uses MALDI-TOF mass spectrometry to quantitate methylation in amplicons up to 600bp.

For genome wide analysis other methods such as the array based Illumina Infinium® Methylation BeadChip which can assess methylation of 27 thousand (27K), 450 thousand (450K) and most recently development of the EPIC Beadchip which can analyse 850,000 methylation sites across the genome at single nucleotide resolution. This works on bisulphite converted whole-genome amplified DNA that hybridises to bead chips that are either methylated or unmethylated. Each chipis scanned and fluorescence intensity of the labelled nucleotides allows quantitation of methylation at individual CpG sites.

These advances in technology have developed our understanding of differential methylation in human health and disease and with efforts from consortia such as the NIH Epigenome Roadmap and TCGA, publically available genome wide methylation data will aid in our understanding of basic biology which will lead to translational research into therapeutic targets and clinically relevant biomarkers.

### *1.5.3 Metabonomics*

Metabonomics is one of the newest members of 'omics' science and refers to the study of small molecular weight compounds or metabolites and their relationship within a biological system. Endogenous metabolites play a key role in biological systems and represent potential targets for understanding disease initiation and progression. Metabonomics has advanced our understanding of diseases such as cancer and is increasingly being used in the discovery of diagnostic biomarkers and identification of pathways involved in cancer that could ultimately lead to identification of potential drug targets, monitoring response to therapy and in our understanding of disease mechanisms (Wishart, 2016).

#### 1.5.3.1 Alterations in cancer cell metabolism or the metabolic response to cancer

Cancer cells undergo profound metabolic reprogramming to sustain increased energy demand for growth and proliferation and it is considered a hallmark of cancer (Hanahan and Weinberg, 2011, Pavlova and Thompson, 2016). Alterations in the metabolism of cancer cells to meet the bioenergetics, biosynthetic and redox demands of malignant cells are extensive; a brief summary of some of the most notable and well characterised changes are discussed here.

The classical and most well-known metabolic reprogramming in cancer is the Warburg effect. This refers to the increase in glycolysis seen in cancer cells even in the presence of oxygen (Warburg, 1956). Glycolytic intermediates are then able to feed subsidiary pathways to fulfil to the metabolic needs of proliferating cells (Lunt and Vander Heiden, 2011). TCA cycle intermediates can also contribute to the biosynthetic needs of the cell. Other mechanisms are therefore required to replenish the carbon sources in the cycle and any pathway that contributes to this is termed an anaplerotic pathway (Lunt and Vander Heiden, 2011). Examples of anaplerotic pathways include glutaminolysis, which provides alpha-ketoglutarate from glutamine, pyruvate carboxylation which provides oxaloacetate from glucose/pyruvate and oxidation of some branched chain amino acids such as isoleucine and valine into acetyl-CoA (Owen *et al.*, 2002).

In normal cells, growth factors stimulate the PI3K-AKT-mTOR pathway that in turn activates HIF-1 (hypoxia-inducible factor) and SREBP (sterol regulatory element binding protein (SREBP) to promote glycolytic flux and fatty acid synthesis (Dibble and Manning, 2013). Aberrant activation at any of the

points in this pathway, which frequently occurs in a number of cancers, can therefore achieve constitutive signalling with minimal growth factor dependence thereby reprogramming metabolism to support growth (Yuan and Cantley, 2008).

In addition, gain of function of *MYC* supports anabolic growth through increased expression of enzymes and transporters involved in glycolysis, glutaminolysis, fatty acid synthesis and serine metabolism (Stine *et al.*, 2015). Figure 1.2 highlights some of the signalling pathways that regulate cancer metabolism. Other well-known oncogenes and tumour suppressor genes such as *Kras* and *p53* have also been shown to promote tumorigenesis through regulation of metabolism (Kruiswijk *et al.*, 2015).

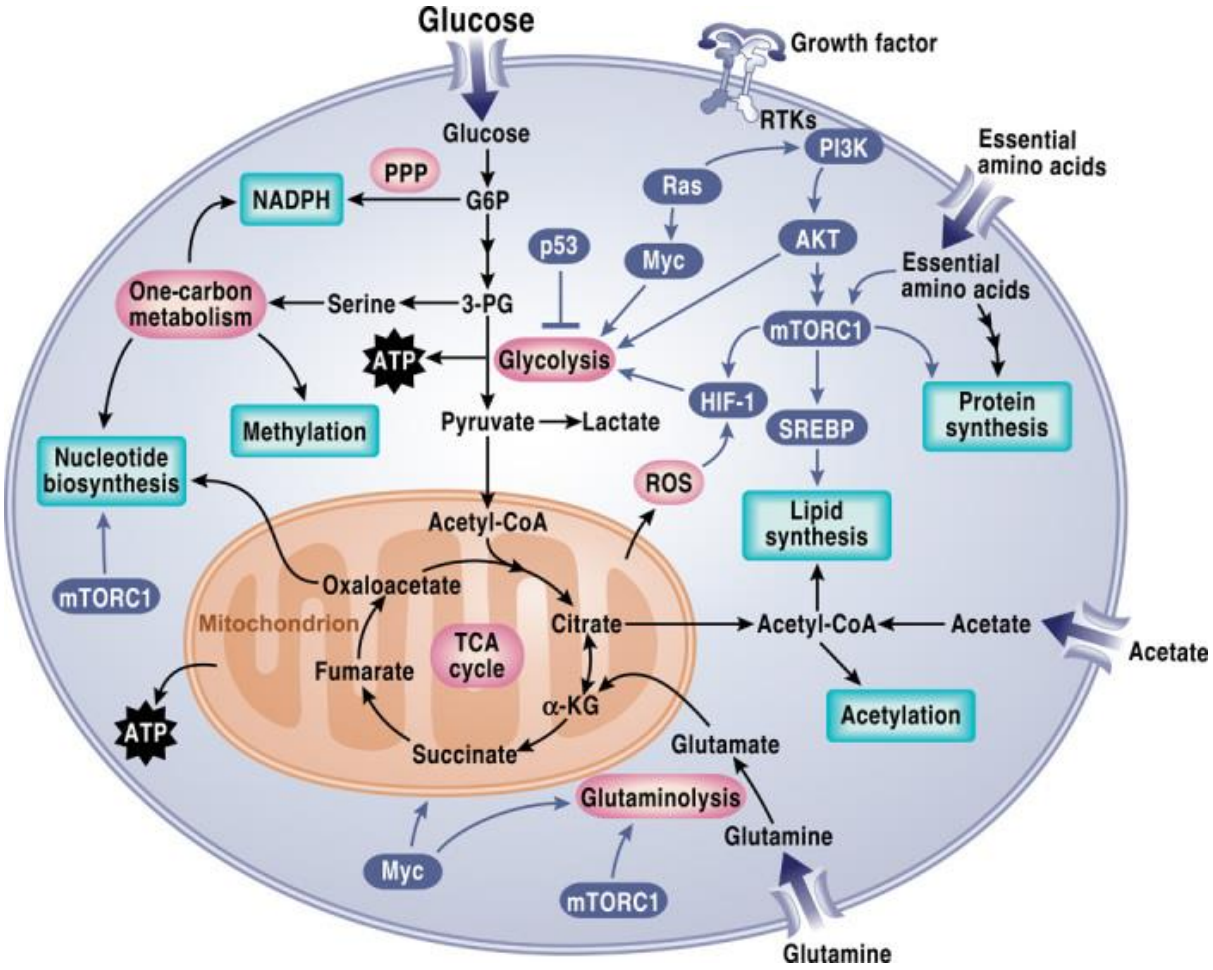



Figure 1.2 Signalling pathways that regulate cancer metabolism. Reproduced with permission from (Deberardinis and Chandel, 2016). Aberrant activation of mTORC1 in tumour cell induces an anabolic growth program resulting in synthesis of nucleotides, proteins, and lipids. Loss of tumour suppressors like *p53* or activation of oncogenes like *MYC* further promotes anabolism through transcriptional regulation of metabolic genes. Metabolism controls signalling through regulating reactive oxygen species (ROS), acetylation, and methylation. PPP, pentose phosphate pathway; G6P,

glucose-6-phosphate; 3-PG, 3-phosphoglycerate; ATP, adenosine 5´-triphosphate; mTORC1, mechanistic target of rapamycin complex 1; α-KG, α-ketoglutarate; RTK, receptor tyrosine kinase.

To support growth and proliferation, there is a huge degree of metabolic flexibility to changes in fuels supplied in the tumour microenvironment (Boroughs and Deberardinis, 2015). This constant change in metabolic reprogramming may provide greater challenges in identifying novel targets for therapy and biomarkers of disease. However, mutational deregulation of metabolic enzymes such as isocitrate dehydrogenase (IDH) 1 or 2 that produce elevated levels of one or more metabolites such as D-2-hydroxyglutarate (D2HG) could prove to be effective biomarkers of disease. Indeed mutations in *IDH1/IDH2* occur in gliomas, acute myeloid leukaemias (AML) and other cancers (Yan *et al.*, 2009, Mardis *et al.*, 2009, Kang *et al.*, 2009) and concentrations of D2HG can be used to identify these patients both in biofluid (Lombardi *et al.*, 2015) and in tissue (Voelxen *et al.*, 2016).

Identification of these so called 'oncometabolites' such as D2HG are rare, indeed the list is short. However, patterns of metabolic reprogramming of more than one metabolite may have utility as a biomarker of disease. For example, ovarian cancer cells take up fatty acids from neighbouring adipocytes to maintain the ATP/ADP ratio in nutrient and oxygen poor environments and fatty acids are a key component of membranes. Other macromolecules such as proteins can also be taken up from the extracellular space and degraded through autophagy to supply essential amino acids for the cell (Galluzzi *et al.*, 2015). These interactions with the extracellular space and the vasculature as well as oncometabolites could lead to specific patterns of metabolites in the blood stream that could have potential as biomarkers for cancer detection and disease progression. Such identification requires highly sensitive techniques to characterise and quantify the metabolites within a biological sample.

#### 1.5.3.3 Technology

Metabolomics technologies seek to characterise and quantify the metabolites in a biological sample. No single method can quantify the entire metabolome and therefore the two most common techniques used are nuclear magnetic resonance (NMR) and mass spectrometry (MS) which provide the most complete assessment of the metabolic phenotype. Methods for sample extraction, derivatisation, separation and detection can further increase the number and class of compounds that can be assessed using these techniques. Advanced chemometric multivariate data analysis tools are required to analyse and interpret the complex biochemical mixtures.

There is currently an increased emphasis on quantification and automation in metabonomics with the introduction of commercial kits such as the Absolute*IDQ®* p180 Kit from Biocrates Life Sciences AG which uses internal standards and reference standards to provide absolute quantification of

various metabolites including amino acids, acylcarnitines and glycerophospholipids. Automation and quantification can increase throughput, reliability and reproducibility which are essential for validation and ultimate adoption into a clinical setting however disadvantages include only having a limited view of the metabolome and reliance on commercial partners which could lead to a reduction in method innovation in the field.

To ensure, reliability and reproducibility in the field, strict protocols for sample handling, quality control and standardised procedures for data processing and interpretation are required. Much like its other 'omics' counterparts there is a need for a repository for data generated from metabonomics studies ideally with associated clinical meta data that can be interrogated with data from other 'omics' such as genomics or epigenomics.

#### 1.5.3.3.1 Nuclear magnetic resonance (NMR)

NMR spectroscopy can be used to detect a wide range of low molecular weight compounds (<1000 Daltons) in biofluids and solid tissue (Beckonert *et al.*, 2007). It is a powerful tool in metabonomics as it is highly reproducible and does not destroy the sample providing a 'snapshot' of the metabolism at the time of sampling (Dona *et al.*, 2014).

NMR spectroscopy is a technique that exploits the magnetic properties of various atomic nuclei, including $^{1}H$, $^{13}C$, $^{15}N$ and $^{31}P$, to obtain information about molecular structure and therefore the identity of molecules within a biological sample (Bothwell and Griffin, 2011). In this study, the magnetic property of the proton nuclei ($^{1}H$) was assessed.

At the centre of the $^{1}H$ NMR spectrometer is a superconducting magnet that the sample is placed into which causes the proton nuclei to spin. At the base of the magnet is a radiofrequency (RF) coil that sends radiofrequency pulses to excite the protons (Figure 1.3). The coil also collects the oscillating electromagnetic signals as the protons relax back to equilibrium called the free induction decay (FID). This waveform of intensity versus time is fourier transformed to produce $^{1}H$ NMR spectra of intensity versus frequency. The resonance frequency of the nucleus is related to the resonance frequency of an internal standard; usually trimethylsilylpropanoic acid (TSP); this is called the chemical shift (δ). The chemical shift (δ) is dependent on the chemical environment of the proton which is directly related to the molecular structure.

The data is represented as a series of peaks with the chemical shift (δ) in parts per million (ppm) on the x axis and intensity on the y axis. Some molecules contain a number of protons which can experience different environments and can exert an effect on the magnetic field producing different chemical shifts depending on their location, this effect is termed spin-spin coupling. The result is a

number of peaks with various intensities along the spectrum corresponding to the same molecule and the splitting of peaks termed multiplicity.

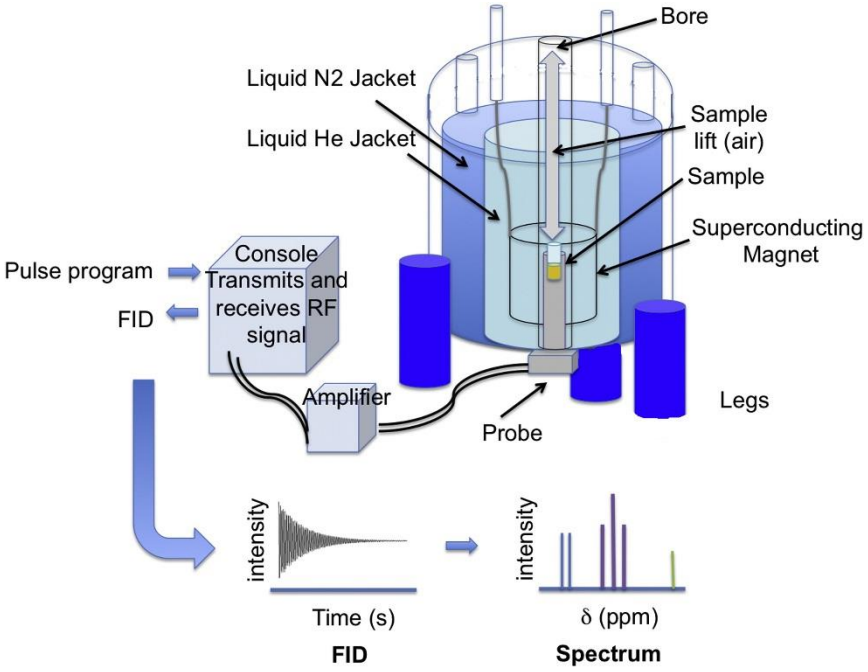



Figure 1.3 Simplified diagram of an NMR spectrometer (Reproduced with permission from (Rankin *et al.*, 2014). A superconducting magnet within the NMR spectrometer causes the proton nuclei in the sample to spin. At the base of the magnet is an RF coil that sends radiofrequency pulses to excite the protons. The coil collects the oscillating electromagnetic signal as the protons relax back to equilibrium called the free induction decay (FID). This waveform of intensity versus time is fourier transformed to produce $^{1}$H NMR spectra of intensity versus frequency.

Different pulse sequences can be applied from the RF coil to enhance or attenuate different signals. In this study, the Carr-Purcell-Meiboom-Gill (CPMG) pulse program was applied to the samples to identify smaller molecular weight molecules through attenuation of signals from proteins and lipids (Meiboom and Gill, 1958). This is achieved through an extended delay between the excitatory pulse and the acquisition period where nuclei from the larger molecules will have stopped spinning and the resonances are therefore not collected and so are not visible in the spectra. The other pulse sequence within the study is the nuclear overhauser effect spectroscopy (NOESY) presaturation sequence which suppresses the water peak without compromising signals from the other molecules within the sample (Nicholson and Foxall, 1995). Examples of spectra from samples included in the study are shown in Chapter 4, Figure 4.1 (CPMG) and Figure 4.21 (NOESY).

*1.5.3.3.1.1 LipoFIT – Lipoprotein subfraction analysis using NMR*

Lipoproteins are made up of lipids and proteins that function as carriers of hydrophobic lipid molecules such as cholesterol and triglycerides through the aqueous bloodstream. Lipoproteins can be classified into five major groups: chylomicrons, very low, low, intermediate and high density lipoproteins where chylomicrons are the largest and least dense particles. They can be further classified according to the lipid composition itself: triglycerides, phospholipids and cholesterol as well as the associated apoprotein; apoA, apoB, apoC, apoD and apoE. Lipoprotein fractionation and quantification have been of particular interest in the investigation of cardiovascular disease development and risk (Carmena *et al.*, 2004) however recent interest in other human diseases such as multiple sclerosis (MS) have also increased (Jorissen *et al.*, 2017) due to their role in regulation of the inflammatory response.

NMR Spectroscopy can quantify lipoproteins and their subfractions due to the fact that they have different chemical compositions and sizes and will therefore experience different chemical environments and produce distinct signals between the 1.4 and 0.6 ppm spectral resonances corresponding to the methyl and methylene groups of the lipids (Jeyarajah *et al.*, 2006). NMR spectroscopy can therefore identify and quantify lipoprotein and subfractions in the blood, as the amplitude of the signal is proportional to the concentration. Several commercial companies have developed protocols for lipoprotein subfraction analysis from NMR spectra including LipoProfile® (LipoScience Inc) and AXINON® lipoFIT® (Numares AG, Regensburg, Germany (Aru *et al.*, 2017).

1.5.3.3.2 Mass Spectometry (MS)

Mass spectrometry is a technique based on the measurement of the mass to charge (m/z) ratio of molecular species after they have been ionised. Before detection, molecular species can be chromatographically separated using techniques such as: capillary electrophoresis (CE), gas-chromatography (GC) and liquid chromatography (LC) where the species are eluted based on their affinity for the column in use. The packing material of the column can determine what class of compounds are separated and multiple columns can be used to gain complementary information from biological samples.

Before samples enter the mass spectrometer they need to be ionised, i.e. positively or negatively charged. The two most common forms of ionisation are electrospray ionisation and atmospheric pressure chemical ionisation (APCI). Electrospray ionisation uses electrical energy to assist the transfer of ionic species from solution into the gaseous phase. Once the solvent has been evaporated the ions enter the mass spectrometer which is under high vacuum. Ions are then

separated based on their mass using a variety of techniques before they reach the detector where a mass spectrum displays the relative intensities of the signals according to their m/z ratio.

For untargeted metabonomic analysis, that is the global profiling of the metabolome in a biological sample, the time-of-flight (TOF) mass spectrometers tend to be most widely used. Once the ions enter the mass spectrometer they travel through the time of flight tube and their velocity is dependent on their mass; a smaller molecule which reach the detector quicker than a larger molecule, provided they have the same charge. This has become popular in untargeted metabonomics due to the high sensitivity and wide range m/z scan coupled with lock-mass correction for improved mass accuracy. However, other mass spectrometers, such as the quadrupole mass spectrometers have become popular in targeted analyses of specific compounds and are currently used in the clinical laboratory to measure immunosuppressant drugs and screen for inborn errors of metabolism. To this end, to increase the accessibility and provide validation of the potential biomarkers identified in metabonomics, targeted mass spectrometric analysis of biomarkers is becoming increasingly used.

Quadrupole mass spectrometers have two pairs of cylindrical rods with equal and opposite charge which cause an oscillating electric field that the ions pass through before they reach the detector. It is common place to have an additional mass analyser connected through a collision cell, termed tandem mass spectrometry, where ions are fragmented before detection in the second mass spectrometer. This allows the precursor ions of interest to be selected before entering the collision cell and the fragmented or daughter ions to be further selected in the second quadrupole, this is termed selected or multiple reaction monitoring (MRM). The targeted Absolute*IDQ*® p150/p180 Kits from Biocrates Life Sciences AG enables mass spectrometry based, quantitative analysis of up to 200 endogenous metabolites using triple quadrupole mass spectrometry. One of the main advantages of the 'kit' based targeted quantitative analysis over untargeted global profiling is that it eliminates the need for feature extraction, complex computational and statistical analysis and metabolite identification, which are some of the most challenging aspects of metabonomic analysis.

#### 1.5.3.4 Multivariate statistical analysis for metabonomic data

Data generated from untargeted metabonomics can include hundreds of metabolites or features and comprehensive evaluation of the data is required to identify metabolic mechanisms and biomarkers of human diseases. Multivariate data analysis is suitable for analysis of large datasets, where the number of variables is larger than the number of samples. These techniques can simplify these complex multivariate data-sets for visualisation of the similarities and differences between

samples. There are two main approaches; unsupervised and supervised analyses which are discussed below.

#### 1.5.3.4.1 Unsupervised multivariate statistical analysis methods

Unsupervised analysis of high dimensional datasets can give an overall idea of the structure of the data and can reveal inherent trends and outliers.

Principal components analysis (PCA) is a multivariate dimensionality reduction method designed to extract and graphically represent the systematic variation in a data matrix without *a priori* knowledge of the sample classification. This method aims to find a smaller number of uncorrelated variables, or principal components (PCs) that can best explain the total variation in the original dataset, whilst still retaining most of the information (Wold *et al.*, 1987). If a variable shows large differences between samples, this can dominate the variation and therefore the PCA model.

A scaling method can be used to minimise this effect and adjust the importance of each variable. Scaling methods include unit variance; which gives the same importance to all variables or Pareto scaling; which takes into account the difference in variation in variables and minimises these differences (Van Den Berg *et al.*, 2006).

The PCs explain the proportion of variability or % variance in the data and although the number of PCs is equal to the number of variables, only a limited number are used as they explain progressively less information. When two PCs have been derived, they define a plane into which the observations are projected onto to improve visualisation of the data. The observations on this plane are called scores which explain the variability between the samples called the scores plot. A Hotellings T2 ellipse is used to define a 95% tolerance region and observations outside this region suggest they deviate from normality. The associated loadings plot illustrates the weight of the variables responsible for this separation between the samples in the scores plot.

#### 1.5.3.4.2 Supervised multivariate statistical analysis methods

The purpose of supervised multivariate statistical analysis methods is to determine the association between the response variable; a continuous matrix and the covariates; the spectral intensities. In supervised methods, the response variable is known and is used to guide the training of the model; making it appropriate for biomarker discovery and predictions.

The partial least squares or projection to latent structures (PLS) method (Wold *et al.*, 1996) finds the best set of X variables (metabolic data) that can explain the variation in Y (class information). To determine the set of X variables that could predict the variation between 2 classes, such as case vs. control the PLS-discriminant analysis (DA) can be used where Y is a vector representing class

membership (Boulesteix, 2004). In addition, another variation of the method is the orthogonal projections to latent structures (OPLS) (Trygg and Wold, 2002) which removes systematic variation in X that is orthogonal to Y thereby enhancing the interpretation of the supervised model.

There are a number of ways to ensure the validity of the supervised models created. These include assessment of the $R^2$ and $Q^2$. The $R^2$ refers to the 'goodness of fit' of the model and $Q^2$ refers to the predictive power of the model. A seven-fold cross validation was also used to ensure that the models were not over-fitted. This is done by leaving out $1/7^{th}$ of the data at a time, this is used to train the model and this information is used to predict the other data. This is repeated until all the data has been predicted. The robustness of the model can then be determined using the permutation test where the class variable of a Y matrix is deliberately altered a set number of times and a new model generated. The permutated model should have a lower $R^2$ and $Q^2$ compared to the original model, if it is robust.

The loadings plot or the loadings correlation coefficient plot, in the case of NMR data, for the (O)PLS DA model represents the metabolites that are responsible for the classification. In the case of NMR data, a pseudo NMR spectrum is created where the correlation values are superimposed onto the spectrum to show the specific metabolites causing the separation (Cloarec *et al.*, 2005b). Once they have proven to be robust as discriminating classes they must next be identified.

#### 1.5.3.4.3 Metabolite identification

There are a number of methods that can be used to identify metabolites or features belonging to metabolites identified in multivariate statistical analysis. The literature and databases of metabolites (e.g. HMDB) can then be interrogated to identify the peaks. However, many peaks in an NMR spectrum can overlap depending on the structure and environment of the molecule, making it hard to confirm an assignment.

Another method is statistical total correlation spectroscopy (STOCSY) that can identify correlation of the intensities of peaks across the entire spectrum. This method can identify other spectral intensities relating to that feature thereby helping to elucidate the identity of the original feature but also identify features from other metabolites which may be related mechanistically to the original feature (Cloarec *et al.*, 2005a).

Another method for increased certainty in the identification of the metabolite is via spiking experiments, where the pure standard is spiked into a sample and subsequently analysed to see if the peak(s) of interest change in intensity.

A diagram of the workflow for the experimental and data analysis of samples using $^{1}$H NMR is shown in Figure 1.4.

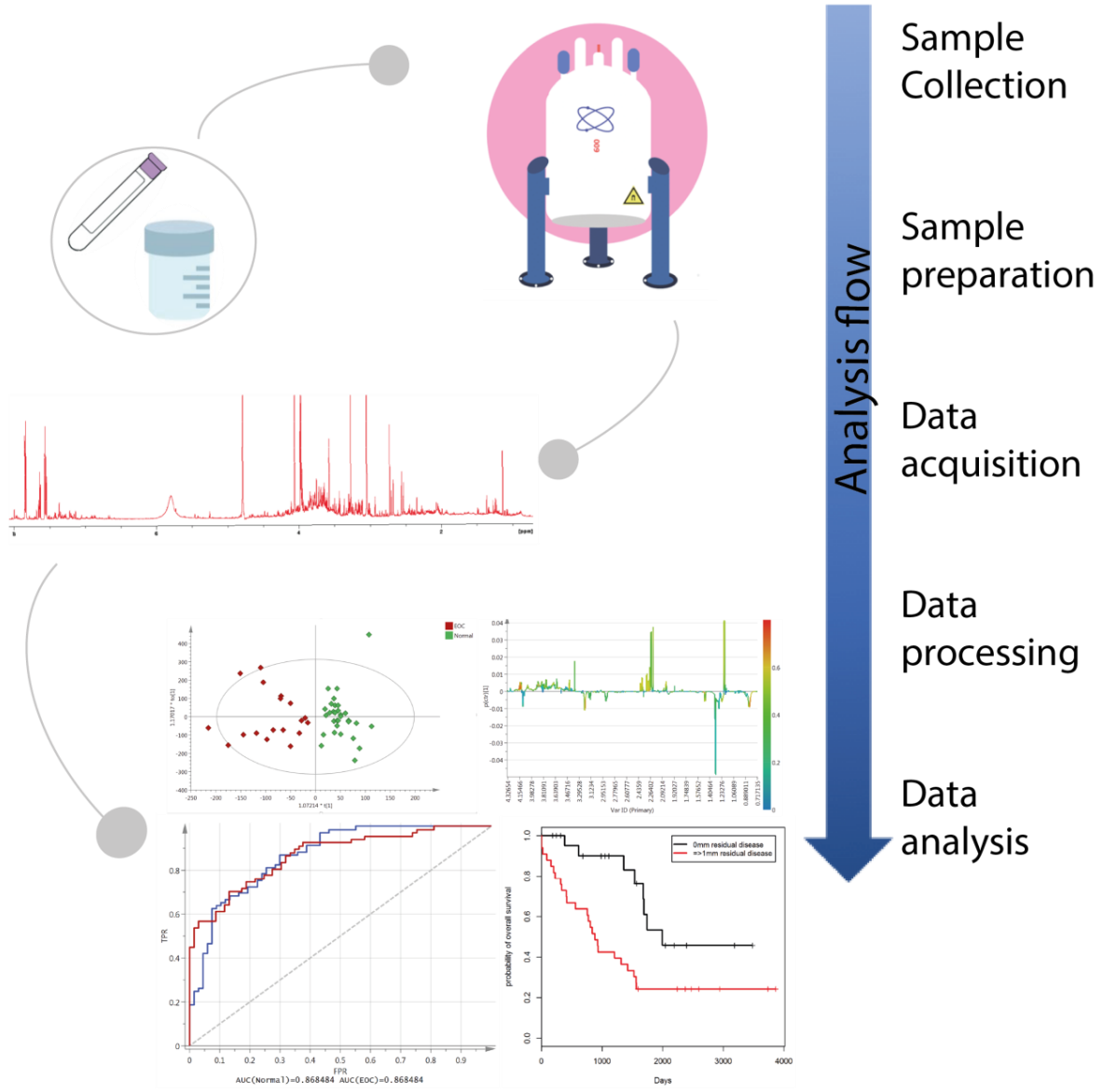



Figure 1.4 Workflow for experimental and data analysis of samples by $^{1}$H NMR Spectroscopy. Plasma and urine samples were analysed by $^{1}$H NMR spectroscopy. The data was pre-processed and analysed using unsupervised and supervised multivariate analysis using SIMCA software. The diagnostic and prognostic performance of the global profile as well as individual spectral regions, were assessed. NMR image adapted with permission from (Ahmad *et al.*, 2017).

#### 1.5.3.5 Systematic review of altered metabolism in ovarian cancer

In order to explore and summarise potential biomarkers of altered metabolism in ovarian cancer, a systematic literature search was conducted covering any studies that had explored the metabolome of patients with ovarian cancer see Materials and Methods: Section 2.2.

A summary of the results from this review are shown in Table 1.2. However, at the same time another group also published a review of the current literature on use of metabolomics for biomarker discovery in ovarian cancer (Turkoglu *et al.*, 2016).

Pathways relating to lipid, amino acid, nucleotide and ketone metabolism were all shown to be altered in ovarian cancer compared to control populations. Some of the most notable and consistent alterations were found in metabolites relating to cellular respiration where increased levels of intermediates in the TCA cycle were identified. Increases in the amino acids: glutamine, glycine, cysteine and threonine were consistent features whereas there was enhanced degradation of tryptophan, histidine and phenylalanine. Interestingly elevated levels of ketone bodies, such as 3-hydroxybutyrate were also consistently reported. Kyneurine, a metabolite of tryptophan as well as glycerophospholipids were also identified as potential biomarkers associated with survival in patients with EOC.

Metabonomic analysis of non-invasively sampled body fluid from cancer patients and associated controls provides an opportunity to develop diagnostic and prognostic biomarkers; however the clinical utility of these biomarkers has remained limited. A lack of associated clinical and other 'omics' data has prevented further elucidation of the mechanisms and pathways involved in the pathogenesis of the disease. Furthermore, the lack of validation studies and small numbers of study participants has thwarted the progression of these biomarkers to the clinic.

Table 1.2. Table of publications and summary of findings of metabonomics-based analysis in body fluid to identify potential biomarkers of epithelial ovarian cancer.

| Disease/control | Cohort size | Histology | Grading/Staging | Specimen type | Upregulated metabolites | Downregulated metabolites | Analytical method | Ref |
|---|---|---|---|---|---|---|---|---|
| EOC<br>Benign Cyst<br>Healthy women | 38<br>12<br>53 (21 pre*, 32 post*) | 30 papillary serous<br>2 clear cell<br>2 endometrioid<br>1 mucinous<br>1 undifferentiated<br>2 other | 1 Stage IB<br>1 Stage IC<br>34 Stage IIIC<br>2 Stage IV | Serum | 2.77ppm, 2.04ppm, 3.7ppm (sugar hydrogens), 2.25ppm, 1.18ppm (beta-hydroxybutyrate) | | $^{1}$H NMR | (Odunsi *et al.*, 2005) |
| **Research aim:** Examine the utility of the metabonomic approach for discriminating sera from women with EOC from healthy controls. | | | | | | | | |
| Ovarian carcinoma<br><br>Borderline tumour | 66<br><br>9 | 50 serous<br>16 non-serous<br>5 serous<br>4 mucinous | 1 G1<br>23 G2<br>38 G3<br>12 Stage I<br>4 Stage II<br>47 Stage III<br>3 Stage IV | Fresh frozen tumour samples | Glycerolphosphate alpha, Uracil, Hypoxanthine, Pyrazine-2,5-bishydroxy, Inositol-2-phosphate, Phosphoric acid, Glutamic acid, Glycine, Malic acid<br>γ-aminobutyric acid, α-Tocopherol 2, Glucose-1-phosphate degr., Proline, Fumaric acid, Creatine, Cysteine, Butyric acid 2-hydroxy, Glycine minor, Glutamine, Threonine, Asparagine | Nonadecanoic acid, Stearic acid, Heptadecanoic acid, Benzoic acid, Lactic acid | GC-TOF-MS | (Denkert *et al.*, 2006) |
| **Research aim:** Characterise the molecular changes in ovarian tissue through metabolic profiling of quantitative signatures of primary metabolites. | | | | | | | | |

| Disease/control | Cohort size | Histology | Grading/Staging | Specimen type | Upregulated metabolites | Downregulated metabolites | Analytical method | Ref |
|---|---|---|---|---|---|---|---|---|
| EOC<br>Breast Cancer<br>Healthy volunteers | 50<br>48<br>72 | | 10 Stage I<br>2 Stage II<br>32 Stage III<br>2 Stage IV<br>4 Undocumented | Urine | | <u>EOC vs. healthy</u><br>Creatine<br>Acetate<br>Succinate<br>Levoglucosan<br>Lactate<br>Pyroglutamate<br>Formate<br>Isoleucine<br>Sucrose<br>Trigonelline<br>Leucine<br>Asparagine<br>Urea<br>Glucose<br>Ethanolamine<br>Dimethylamine<br>4-hydroxyphenylactate<br>Creatinine<br>Alanine<br>Hippurate<br>1-methylnicotinamide<br>Uracil<br>Valine<br>Trans-Aconitate | $^{1}$H NMR | (Slupsky *et al.*, 2010) |
| **Research aim:** To apply metabolomics to find urinary biomarkers of breast and ovarian cancer. | | | | | | | | |
| EOC (early stage I/II)<br>Age-matched controls<br><br><br>RCC female patients | 120 & 50 (median age 49)<br>132 & 50 (median age 49)<br><br><br>30 (median age 66) | 45 endometrioid<br>21 serous<br>17 clear cell<br>26 mucinous<br>11 mixed<br><br>26 clear cell<br>4 papillary | 43 G1<br>38 G2<br>26 G3<br>4 G4<br>86 & 35 Stage I<br>33 & 15 Stage II<br><br>15 Stage I<br>15 Stage II-IV | Serum | Beta-hydroxybutyrate, acetone, acetoacetate | Alanine, choline moiety of phospholipids, creatine/creatinine, valine, LDL1, VLDL, VLDL1 | $^{1}$H NMR | (Garcia *et al.*, 2011) |
| **Research aim:** Explore the use of $^{1}$H NMR based metabonomics of serum for the diagnosis of early stage (I/II) EOC using a microflow NMR probe. | | | | | | | | |

| Disease/control | Cohort size | Histology | Grading/Staging | Specimen type | Upregulated metabolites | Downregulated metabolites | Analytical method | Ref |
|---|---|---|---|---|---|---|---|---|
| Advanced EOC<br>12 without recurrence<br>26 with recurrence | 38 (15 <50 years, 23>50 years) | 23 serous<br>15 others | 3 G1<br>9 G2<br>26 G3<br>36 Stage III<br>2 Stage IV | Plasma | Kynurenine, bilirubin | L-tryptophan, lyso PC(14:0), lysoPE (18:2) | UPLC-QTOF-MS | (Zhang *et al.*, 2015) |
| **Research aim:** Identify metabolic markers associated with EOC recurrence. | | | | | | | | |
| Pre-op EOC<br><br><br><br>Benign ovarian tumour<br><br><br><br>Healthy controls<br><br><br>Post-op EOC | 40 (21 post, 19 pre)<br><br><br>62 (14 post, 35 pre, 13 undocumented)<br><br><br>54 (5 post, 45 pre, 4 undocumented)<br>18 (9 pre, 9 post) | 17 serous<br>6 others<br>8 undocumented<br><br><br><br><br><br><br><br>9 serous<br>3 others<br>6 undocumented | 0 G1<br>1 G2<br>19 G3<br>20 undocumented<br>7 Stage I<br>5 Stage II<br>20 Stage III<br>2 Stage IV<br>6 undocumented<br><br><br><br>0 G1<br>1 G2<br>9 G3<br>8 G4<br>5 Stage I<br>1 Stage II<br>10 Stage III<br>1 Stage IV<br>1 Undocumented | Urine | <u>EOC vs. healthy</u><br>Imidazole-5-yl-pyruvate, N-acetylneuraminate 9-phosphate, N4-acetylcytidine, urate-3-ribonucleoside, pseudouridine, succinic acid, (S)-reticuline, taurine, N-acetylneuraminic acid, 3-sialyl-*N*-acetyllactosamine, β-nicotinamide mononucleotide, prolylhydroxyproline, selenocystathionine, 3'-sialyllactose, N-acetylgalactosamine 4-sulfate<br><u>EOC vs. BOT</u><br>3-dehydroquinic acid, 3-indolelactic acid, LPA (P-16:0e/0:0),(S)-reticuline, N-acetylneuraminic acid, 3-sialyl-*N*-acetyllactosamine, selenocystathionine, 3'-sialyllactose<br><u>Pre vs. post op</u><br>N4-acetylcytidine, urate-3-ribonucleoside, pseudouridine, Nα-acetyl-L-glutamine, succinic acid | <u>EOC vs. healthy</u><br>L-histidine, Nα-acetyl-L-glutamine, prasterone sulfate, glycodeoxycholate<br><u>EOC vs. BOT</u><br><br><u>Pre vs. post op</u><br>3-indolelactic acid, 3-sialyl-*N*-acetyllactosamine, 3'-sialyllactose | UPLC-QTOF-MS | (Zhang *et al.*, 2013) |
| **Research aim:** Evaluate the application of urinary metabolomics on discovering potential biomarkers of epithelial ovarian cancer. | | | | | | | | |

| Disease/control | Cohort size | Histology | Grading/Staging | Specimen type | Upregulated metabolites | Downregulated metabolites | Analytical method | Ref |
|---|---|---|---|---|---|---|---|---|
| Ovarian cancer<br><br>Benign ovarian tumour<br><br>Healthy women | 22 (15 pre, 7 post)<br><br>29 (26 pre, 3 post)<br><br>25 (22 pre, 3 post) | 17 serous<br>1 clear cell<br>4 endometrioid<br>8 serous<br>19 mucinous<br>1 endometrioid<br>1 other | 5 Stage I<br>5 Stage II<br>10 Stage III<br>2 Stage IV | Urine | <u>EOC vs. BOT/Healthy</u><br>Pseudouridine, fragment of pseudouridine, phytosphingosine, | <u>EOC vs. BOT/Healthy</u><br>Hippuric acid, fragment of hippuric acid, homovanillic acid sulfate | HILIC-MS<br>RPLC-MS | (Chen *et al.*, 2012) |
| **Research aim:** To identify the differences of urinary hydrophilic and hydrophobic metabolic profiling between healthy women, benign ovarian tumour and ovarian cancer. | | | | | | | | |
| EOC<br><br>Benign ovarian tumour<br><br>Healthy women | 29 (15 pre, 14 post) (206)<br><br>28 (21 pre, 7 post) (199)<br><br>27 (15 pre, 12 post)(361) | 19 serous<br>1 mucinous<br>2 clear cell<br>6 endometrioid<br>1 other<br>8 serous<br>19 mucinous<br>1 other | 14 Stage I<br>3 Stage II<br>12 Stage III<br>0 Stage IV | Serum | <u>EOC vs. BOT/Healthy</u><br>Phenylalanine,<br>Phe-phe,<br>Propionylcarnitine,<br>LysoPC(16:0),<br>Glycocholic acid,<br>27-nor-5β-cholestane-3,7,12,24,25 pentol glucuronide (CPG) | <u>EOC vs. BOT/Healthy</u><br>LysoPC(18:2) | UPLC-QTOF-MS | (Chen *et al.*, 2011) |
| **Research aim:** To use a two steps strategy metabolomics to screen/identify and validate novel metabolic biomarker(s) for epithelial ovarian cancer. | | | | | | | | |
| EOC<br><br>Normal | 80 (27 pre, 50 post)<br>93 (73 pre, 14 post) | 32 serous<br>1 mucoid<br>3 endometrioid<br>44 unknown | 2 G1<br>9 G2<br>38 G3<br>31 Unknown<br>6 Stage I<br>7 Stage II<br>40 Stage III<br>3 Stage IV<br>24 Unkown | Plasma | Ganglioside, Lysophospholipids, Ceramides, Phytosphingosine, N'-Formylkynurenine | Demethylphylloquinone | UPLC-QTOF-MS | (Fan *et al.*, 2012) |
| **Research aim:** To discover potential biomarkers in distinguishing newly diagnosed EOC from normal individuals and validate the discrimination performance of the selected biomarkers in an independent cohort. | | | | | | | | |
| EOC<br><br>Benign ovarian tumour | 80 (30 pre, 50 post)<br>90 (67 pre, 23 post) | 32 serous<br>1 mucoid<br>3 endometrioid<br>10 Other<br>34 Undocumented | 2 G1<br>9 G2<br>38 G3<br>31 Undocumented | Plasma | | 2-Piperidinone, L-tryptophan, LysoPC(18:3)<br>LysoPC(14:0) | UPLC-QTOF-MS | (Zhang *et al.*, 2012) |

| Disease/control | Cohort size | Histology | Grading/Staging | Specimen type | Upregulated metabolites | Downregulated metabolites | Analytical method | Ref |
|---|---|---|---|---|---|---|---|---|
| **Research aim:** To apply metabolomics to differentiate EOC from BOT and identify biomarkers to discriminate these two groups. | | | | | | | | |
| EOC<br>BOT<br>Uterine fibroid | 140 (58/82)<br>158 (113/45)<br>150 (116/34) | 63 serous<br>4 mucoid<br>4 endometrioid<br>5 clear cell<br>3 undifferentiated<br>1 transitional cell<br>6 mixed<br>37 other<br>17 undocumented | 10 G1<br>16 G2<br>67 G3<br>47 Undocumented | Plasma | EOC vs BOT/UF<br>3,5-Tetradecadiencarnitine, *Cis*-5-Tetradecenoylcarnitine, Dodecanoylcarnitine, Elaidiccarnitine, Tetradecanoylcarnitine, *Trans*-2-Dodecenoylcarnitine Kynurenine, LysoPC(14:0) in localized EOC, LPC(18:2(9Z,12Z)) in localized EOC, LPC(18:3(9Z,12Z,15Z)) in localized EOC, LPE(0:0/16:0) in localized EOC, LPE(20:2(11Z,14Z)/0:0) in localized EOC, PhePhe, L-(-)-3-Phenyllactic acid, L-tyroxine, Hydroxyphenyllactic acid, Pseudouridine | EOC vs BOT/UF<br>L-tryptophan, 3-indolepropionic acid, 5-Hydroxyindoleacetaldehyde, LysoPC(14:0) in metastatic EOC, LPC(18:2(9Z,12Z)) in metastatic EOC, LPC(18:3(9Z,12Z,15Z)) in metastatic EOC, LPE(0:0/16:0) in metastatic EOC, LPE(20:2(11Z,14Z)/0:0) in metastatic EOC, 3b-hydroxy-5-cholenoic acid, Glycoursodeoxycholic acid, Deoxycholic acid, Androsteroneglucuronide, Prasterone sulfate, Piperine, 2-piperidinone, Hippuric acid, γ-CEHC, δ-CEHC, L-histidine | UPLC-QTOF-MS | (Ke *et al.*, 2014) |
| **Research aim:** To systematically investigate ovarian cancer metabolism through metabolic profiling of plasma samples related to ovarian cancer in both positive and negative mode UPLC-MS. | | | | | | | | |
| EOC<br>Benign | 50<br>50 | 50 serous | Stage I 5<br>Stage II 3<br>Stage III 39<br>Stage IV 3 | Plasma | | C52 H79 N O5 S<br>PS (O-18:0/0:0)<br>Alanine | UPLC-QTOF-MS Lipidomics & targeted aqueous profiling | (Buas *et al.*, 2016) |
| **Research aim:** To screen for differentially abundant plasma metabolites between cases with serous ovarian carcinoma and controls with benign serous ovarian tumours using two independent LC-MS based metabolomics platforms, including a recently optimised in-house global lipidomics approach. | | | | | | | | |

| Disease/control | Cohort size | Histology | Grading/Staging | Specimen type | Upregulated metabolites | Downregulated metabolites | Analytical method | Ref |
|---|---|---|---|---|---|---|---|---|
| EOC<br>Benign<br>Healthy | 21<br>17<br>20 | Not given. | Stage I 1<br>Stage II 3<br>Stage III 16<br>Stage IV 1 | Serum | Arabitol (Ca vs. B/H & B vs. H)<br>Maltose (Ca vs. B/H)<br>Maltotriose (Ca vs. B/H)<br>Raffinose (Ca vs. B/H)<br>Mannitol (Ca vs. B/H)<br>Glucose (Ca vs. H)<br>Erythritol (Ca vs. B) | Inosine (Ca vs. B/H & B vs. H)<br>Ribose (Ca vs. B/H) | HILIC-MS/MS | (Cheng *et al.*, 2016) |
| **Research aim:** To study the endogenous carbohydrates in the serum to find some biomarker metabolites to differentiate the malignant ovarian cancer from the normal healthy controls, or malignant ovarian cancer from the benign tumour in samples with an elevated CA 125. | | | | | | | | |
| Early stage EOC<br>Normal healthy controls | 46<br>49 | 46 serous | Stage IA 6<br>Stage IB 2<br>Stage IC 11<br>Stage IIA 6<br>Stage IIB 7<br>Stage IIC 13<br><br>G1 5<br>G2 5<br>G3 29<br>Unknown 7 | Serum | 552.2327 m/z / 0.7min<br>Cortisone<br>Lysophatidylinositol(18:1)<br>Aspartyl-glutamic acid<br>Ceramide(d18:1/16:0)<br>365.3413 m/z, 16.97min<br>Lysophosphatidylethanolamine(22:6)<br>2-hydroxyl nonanoic acid | 16-(6-butoxy-3-hydroxy-4,5-dimethylcyclohex-1-en-1-yl)-6,10-dihydroxy-2,6,10,14-tetramethyl hexadecanoic acid<br>Iso-1,2-octadecanediol<br>3-hydroxyl dodecanedioic acid<br>Phosphatidylinositol(20:4/18:1)<br>7,9,13-trihydroxyoctacosa-16.22-dienoic acid<br>429.2997 m/z,11.49 min<br>451.2275 m/z, 1.58 min<br>129.0909 m/z, 1.14 min | UPLC-MS/MS | (Gaul *et al.*, 2015) |
| **Research aim:** To interrogate the serum metabolome of early-stage ovarian cancer patients and age-matched control women. | | | | | | | | |
| HGSOC<br>Control (non-malignant) | 158<br>100 | 158 serous | Stage I 3<br>Stage II 5<br>Stage III 101<br>Stage IV 32<br>NA 17<br><br>G3 158 | Serum | 3,4-dihydroxybutyric acid<br>Adipic acid<br>Erythritol<br>3-Hydroxyisovaleric acid (β-hydroxy β-methylbutyric acid)<br>2,4-Dihydroxybutyric acid<br>4-Hydroxyphenyllactic acid | | GCxGC-TOF-MS | (Hilvo *et al.*, 2016) |
| **Research aim:** To perform metabolomic profiling of serum and tumour tissue from patients with HGSOC and control patients with benign or non-neoplastic lesions to identify novel diagnostic and prognostic biomarkers associated with tumour burden and patient survival. | | | | | | | | |

| Disease/control | Cohort size | Histology | Grading/Staging | Specimen type | Upregulated metabolites | Downregulated metabolites | Analytical method | Ref |
|---|---|---|---|---|---|---|---|---|
| Malignant<br>Borderline<br>Benign | 10<br>5<br>8 | 1 clear cell<br>1 mixed<br>2 mucinous<br>6 serous | Stage IA 3<br>Stage IB 1<br>Stage IC 0<br>Stage IIA 1<br>Stage IIB 1<br>Stage IIIC 3<br>Stage IV 1<br><br>G2 5<br>G3 4<br>Unknown 1 | Cyst fluid | Lysine (malignant vs. borderline/benign) | Citrate (malignant vs. borderline/benign) | $^{1}$H-NMR | (Kyriakides *et al.*, 2016) |
| **Research aim:** To characterise the metabolic profiles of ovarian cyst fluid samples from benign, borderline, malignant ovarian tumours using $^{1}$H-NMR spectroscopy to shed more light into ovarian tumour and cancer development. | | | | | | | | |
| EOC<br>Healthy volunteers | 56<br>15 | 34 serous<br>8 mucinous<br>9 endometrial<br>4 clear cell<br>1 mixed epithelial | Stage I 10<br>Stage II 9<br>Stage III 31<br>Stage IV 6 | Urine | N-acetylneuraminic acid-9-phosphate<br>5'-Methylthioadenosine<br>Uric acid-3-nucleoside<br>Pseudouridine<br>L-Valine<br>Succinic acid<br>L-proline<br>B-Nicotinamide mononucleotide | | UPLC-QTOF-MS | (Jiang *et al.*, 2015b) |
| **Research aim:** To investigate the correlation between urine metabolites and clinical staging in patients with ovarian cancer. | | | | | | | | |
| EOC<br>Benign<br>Unaffected controls | 27<br>27<br>11 | Not given | Not given | EDTA-plasma | Cancer vs. Benign & Normal Controls<br>16:0 LPC<br>18:1 LPC<br>20:3 LPC<br>20:4 LPC<br>22:6 LPC<br>16:0/18:1 PC<br>16:0/18:2 PC<br>18:0/18:2 PC<br>18:0/20:5 PC<br>18:2/18:2 PC<br>18:2/18:2/16:0 TG | | UPLC-QTOF-MS Lipidomics | (Zhang *et al.*, 2016) |
| **Research aim:** To apply a nontargeted lipidomics approach based on UPLC-QTOF-MS combined with multivariate data analysis to investigate potential diagnostic biomarkers in plasma of OC patients. | | | | | | | | |

| Disease/control | Cohort size | Histology | Grading/Staging | Specimen type | Upregulated metabolites | Downregulated metabolites | Analytical method | Ref |
|---|---|---|---|---|---|---|---|---|
| Primary & Post-operative EOC<br><br><br><br><br><br>Relapsed EOC<br>Controls | 35<br><br><br><br><br><br>35<br>35 | 27 Serous<br>3 Mucoid<br>5 Endometrioid<br><br><br><br><br><br>25 Serous<br>2 Mucoid<br>8 Endometrioid | Stage III 31<br>Stage IV 4<br>G1 5<br>G2 11<br>G3 19<br><br><br><br><br><br>Stage III 31<br>Stage IV 4<br>G1 4<br>G2 14<br>G3 17 | Plasma | Kyneurine<br>Hypoxanthine<br>L-thyroxine<br>N-Undecanoylglycine<br>Hydroxyphenyllactic acid<br>4-Nitrophenol<br>B-Hydroxyisovaleric acid | EOC vs. control<br>LysoPC (18:2)<br>LysoPC (18:3)<br>LysoPC (14:0)<br>LysoPC (18:1)<br>LysoPC (16:1)<br>LysoPC (22:5)<br>LysoPC (20:3)<br>LysoPC (20:4)<br>LysoPE (18:2/0:0)<br>LysoPE (0:0/18:1)<br>L-Tryptophan<br>3-Indolepropionic acid<br>Tetracosahexaenoic acid<br>Aldosterone<br>Glycoursodeoxycholic acid<br>MG (24:6/0:0/0:0)<br>2-Octenoylcarnitine<br>Piperine<br>2-Piperidinone<br>Retinol<br>Dodecanedioic acid<br>12,13-diHODE<br>2-Octenoic acid<br>19,20-DiHDPA<br>γ-CEHC<br>p-Salicylic acid<br>Nutriacholic acid<br>Deoxycholic acid<br>L-Histidine<br>2,6,-Dihydroxybenzoic acid | RPLC-QTOF/MS | (Ke *et al.*, 2016) |
| **Research aim:** To investigate metabolic alterations that underlie the responses to advanced EOC, surgical intervention and disease relapse and to discover potential biomarkers that could monitor EOC surgical efficacy and disease relapse. | | | | | | | | |

| Disease/control | Cohort size | Histology | Grading/Staging | Specimen type | Upregulated metabolites | Downregulated metabolites | Analytical method | Ref |
|---|---|---|---|---|---|---|---|---|
| Early stage EOC<br>Healthy control | 21<br>31 | 7 serous<br>1 mucoid<br>2 endometrioid<br>11 others | All stage I/II<br>G1 1<br>G2 3<br>G3 7<br>Unknown 10 | EDTA-plasma | Pipericine<br>Adrenoyl Ethanolamide<br>395.2257 m/z /629.24s<br>402.3961 m/z /460.79s<br>LysoPC(16:0)<br>683.5415m/z /587.21s | 2-Piperidinone<br>4-(2-Aminophenyl)-2,4-dioxobutanoic acid<br>Trans-2-Dodecenoylcarnitine<br>MG (18:2)<br>Coniferin<br>LysoPE (16:0)<br>LysoPC(14:0)<br>Taurochenodesoxycholic acid<br>LysoPC (17:0)<br>LysoPC (18:3)<br>LysoPE (22:6)<br>548.2875 m/z /411.48s | UPLC-QTOF-MS | (Fan *et al.*, 2016) |
| **Research aim:** To identify diagnostic metabolic biomarkers for the early stages of EOC and explore the possible mechanisms of cancer development and progression. | | | | | | | | |

Predicting survival

| Disease/control | Cohort size | Histology | Grading/Staging | Specimen type | Increase in metabolites associated with survival | Decrease in metabolites associated with survival | Analytical method | Ref |
|---|---|---|---|---|---|---|---|---|
| EOC | 98 | 46 serous<br>10 others<br>42 undocumented | Stage I 9<br>Stage II 9<br>Stage III 56<br>Stage IV 7<br>Undocumented 17<br><br>G1 4<br>G2 11<br>G3 50<br>Undocumented 33 | EDTA-plasma | Kyneurine<br>Acetylcarnitine<br>PC (42:11) | LPE (22:0/0:0) | UPLC-MS | (Xie *et al.*, 2017) |
| **Research aim:** To investigate whether metabolites could be used to predict the survival of EOC through metabolic analysis. | | | | | | | | |
| HGSOC | 65 | 65 serous | Stage III 52 | Serum | | 43 glycerophospholipids | Biocrates | (Bachmayr- |

| Disease/control | Cohort size | Histology | Grading/Staging | Specimen type | Increase in metabolites associated with survival | Decrease in metabolites associated with survival | Analytical method | Ref |
|---|---|---|---|---|---|---|---|---|
| Benign/ Healthy controls | 62 | | Stage IV 13<br><br>G2 14<br>G3 51 | | | 5 amino acids (histidine, lysine, threonine, tryptophan, citrulline)<br>4 sphingolipids<br>1 biogenic amine | Absolute*IDQ*® p180 Kit UPLC-MS | Heyda *et al.*, 2017) |
| **Research aim:** To find survival relevant systemic metabolic changes in cancer patients and to define connections of these changes to the local microenvironment (ascites and the tumour) and to uncover systemic and local immune system and tumour cell characteristics associated with these metabolic changes. | | | | | | | | |

## 1.6 Summary and Hypothesis

As outlined, there is a clinical need for more sensitive biomarkers for diagnosis of epithelial ovarian cancer that could be sensitive enough to be used in screening a low risk population. The 'liquid biopsy' aims to detect changes in the plasma as a consequence of the tumour and various large scale profiling studies may identify novel biomarkers of disease. Understanding the mechanism of production of the biomarkers in disease pathogenesis could uncover possible targets for drug development. There is also a lack of predictive and prognostic biomarkers other than stage of disease and volume of residual disease which can only be assessed post-surgery. The ability to predict response to treatment based on the tumour biology before surgery could help guide treatment strategies and move us closer to achieving personalised medicine.

The first hypothesis addressed in the present study is that epigenomic analysis specifically DNA methylation of high grade serous ovarian tumours will identify candidates for diagnostic biomarkers that can be analysed in relatively non-invasively collected samples such as blood plasma. The second is that metabolic profiling of plasma and urine from patients with epithelial ovarian cancer and a control population will identify metabolic differences between the two groups and these can be further utilised as diagnostic and prognostic biomarkers of epithelial ovarian cancer.

The overall aim of the project is to discover and evaluate these potential diagnostic biomarkers.

## 1.7 Aims

### *1.71 Methylation of cell-free circulating DNA in plasma*

i. Identify differentially methylated loci between high grade serous ovarian cancer and peripheral blood mononuclear cells using genome-wide methylation studies.
ii. Verify findings in independent cohorts and identify if candidate loci are associated with survival.
iii. Develop assays for detection of methylation of candidate loci and validate in blood plasma of patients with high grade serous ovarian cancer and normal controls.
iv. Determine the diagnostic performance of these potential biomarkers in plasma.

### *1.72 Metabolic profiling of plasma and urine*

i. Identify potential diagnostic biomarkers of EOC from $^{1}$H NMR spectroscopy and targeted mass spectrometric analysis using the Biocrates Absolute*IDQ*® p150 Kit.
ii. Determine the diagnostic performance of these potential biomarkers in plasma.
iii. Determine the association of these potential biomarkers with disease characteristics such as stage of disease and survival.

# CHAPTER 2. Materials and Methods

# Methods

## 2.1 Study design and strategy

This biomarker discovery study focuses on two areas of systems oncology; epigenomics and metabonomics to identify biomarkers of epithelial ovarian cancer. The first utilises previously generated datasets in an attempt to identify biomarkers in cell-free circulating DNA in plasma and the second generates large scale metabolic datasets to identify metabolic biomarkers in plasma and urine. The study design is shown in Figure 2.1

**Biomarker discovery**

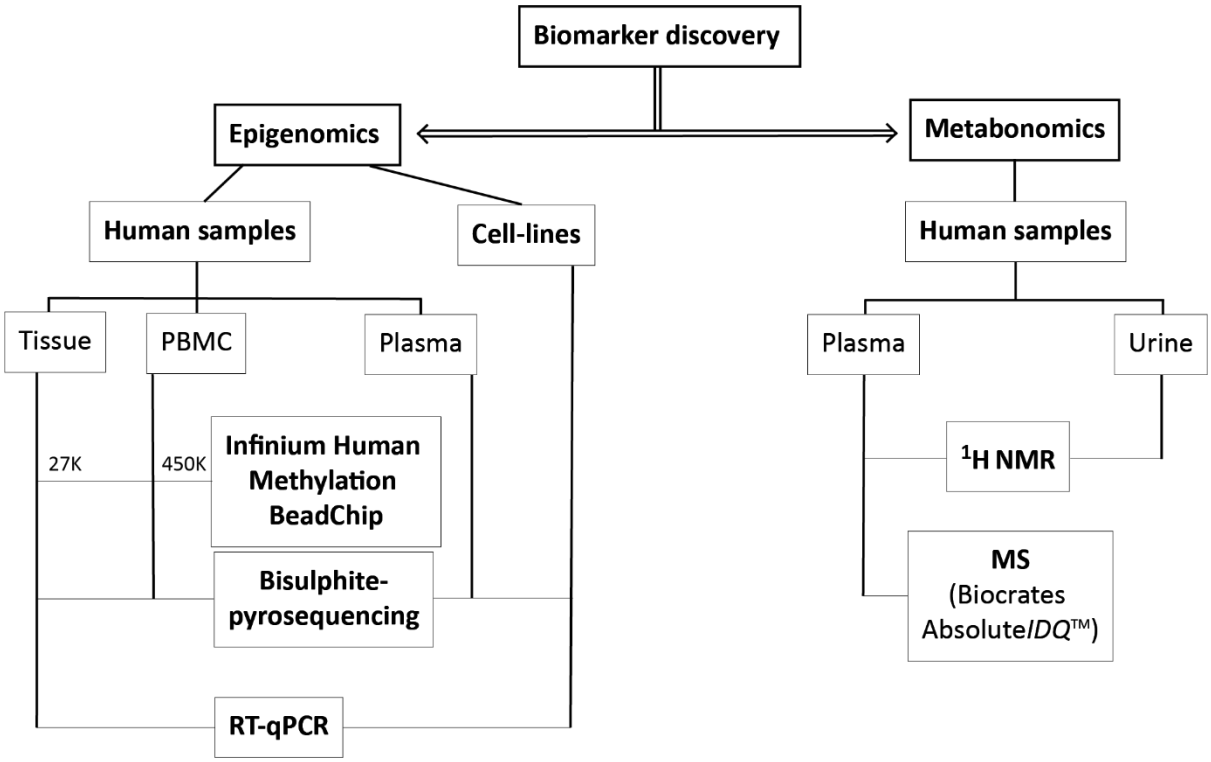



Figure 2.1 Flowchart of project study design

## 2.2 Systematic literature search for metabolites identified in metabonomics-based analysis of ovarian cancer patients

A systematic literature search was performed to identify published studies which investigated the altered metabolome associated with ovarian cancer using metabonomics in PUBMED using the search terms ((metabolomics) OR (metabonomics) AND (ovarian cancer). Only studies investigating altered metabolism in body fluid of ovarian cancer patients such as plasma or urine were assessed. The search yielded 81 articles from August 2003 to March 2017. Studies were excluded if not performed on human body fluid or in ovarian cancer.

## 2.3 Clinical data and sample collection

Previously generated genome wide methylation data-sets were used for the identification of cell-free circulating DNA methylation biomarkers (Chapter 3). These data-sets are summarised in Figure 3.1. Detailed clinical characteristics of the samples involved in each section of the study are shown in the relevant sections of the three results chapters; 3, 4 and 5. Brief descriptions of the cohorts used in this project are given below.

### *2.3.1 TCGA HGSOC dataset & Mayo clinic – 'the discovery cohort'*

#### 2.3.1.1 The Cancer Genome Atlas serous epithelial ovarian cancer (OVCA) dataset

The Cancer Genome Atlas (TCGA) is a large scale collaboration between the National Cancer Institute (NCI) and National Human Genome Research Institute (NHGRI) mapping the genomic changes in 33 types of cancer. Data is publically available and can be downloaded here: (https://portal.gdc.cancer.gov/). A total of 342 primary ovarian serous cystadenocarcinoma tumours with 10 corresponding 'normal' tissues containing level 3 methylation data from an Illumina Human Methylation27 BeadChip and associated annotated clinical data were downloaded. Methylation data had been previously summarised as beta-values, calculated as M/M+U were M is the signal from methylated beads and U is the signal from unmethylated beads at each targeted CpG site.

#### 2.3.1.2 The Mayo clinic PBMC data set

The second data set was from a study by Fridley *et al*., where PBMC samples were collected from patients at the Mayo Clinic in which a total of 276 PBMC samples with beta-value methylation data from an Illumina Human Methylation450 BeadChip were extracted. 121 of these samples were from patients with ovarian serous cystadenocarcinoma and 155 were from age-matched controls. Data was generously made available for use (Fridley *et al.*, 2014).

### *2.3.2 HH cohort & PBMC cohort – 'the verification cohort'*

#### 2.3.2.1 The HH cohort

A dataset previously generated within the lab for validation of analyses in TCGA was also generously made available for use; termed the HH cohort. DNA was extracted from fresh frozen or formalin-fixed paraffin embedded (FFPE) EOC tumour tissue and collected and stored in collaboration with the Imperial Experimental Cancer Medicine Centre (ECMC) and Imperial College Healthcare Tissue Bank (ICHTB). A total of 78 primary high grade serous samples were selected plus 8 additional control samples of fallopian tube and ovary from 4 patients receiving surgery for benign conditions with no previous history of EOC. Methylation data from Illumina Human Methylation27 BeadChip and well annotated clinical data were kindly made available for use (Bonito *et al.*, 2016).

#### 2.3.2.2 The PBMC cohort

A publically available dataset on methylation data from Infinium Human Methylation450 BeadChip of PBMC samples from 111 unaffected women from the Family and Community Health Study (FACHS) was downloaded from the Gene Expression Omnibus (GEO database under accession number GSE35069) and formed part of the verification cohort.

### *2.3.3 Charité and New HH cohort HGSOC dataset – 'the validation cohort'*

In addition, further DNA samples were available for validation studies. 89 serous tumour DNA samples were received from collaborators at Charité Universitätsmedizin, Berlin and kindly made available for use (Phelps *et al.*, 2017). A further 76 high grade serous samples were collected and stored in the ICHTB for validation studies and made available for use. Additional DNA samples stored in the ICHTB from PBMCs and benign ovarian tissue from the HH cohort were also made available for use as part of the validation cohort.

### *2.3.4 Blood and urine samples*

Ethics for sample collection was approved by the Imperial College NHS Trust Research and Ethics committee (formally Hammersmith and Queen's Charlotte's NHS Trust) reference 05/Q0406/178. Human samples used in this research project were obtained from the Imperial College Healthcare Tissue Bank (ICHTB). ICHTB is supported by the National Institute for Health Research (NIHR) Biomedical Research Centre based at Imperial College Healthcare NHS Trust and Imperial College London. ICHTB is approved by NRES to release human material for research (12/WA/0196), and the samples for this project R15018 were issued from sub-collection reference number Gyn_HG_12_060 and Gyn_HG_13_020. Confirmation of diagnosis, stage and grade were confirmed histologically by the Histopathology Department at Hammersmith Hospital. Patients were divided into groups based on their diagnosis. These included patients with histologically confirmed EOC, patients with benign gynaecological disease, patients with a confirmed malignancy not arising from the ovary or fallopian tube, termed the 'Malignant cohort' and unaffected female volunteers, termed the 'Healthy' cohort.

#### 2.3.4.1 Retrospectively collected blood and urine samples – *'the NPC ovarian study'*

Archived sodium heparin plasma and urine samples were collected and stored in the ICHTB to a defined standard operating procedure (SOP) for patients undergoing gynaecological surgery at Hammersmith Hospital between 2005 and 2013 and unaffected female volunteers. Blood was centrifuged at 1100g for 10 minutes and plasma removed and decanted into 1ml eppendorf tubes and stored at -80°C.

#### 2.3.4.2 Prospectively collected blood samples – '*the biofluid cohort*'

Due to PBMC contamination issues with archived plasma samples a prospective cohort of blood samples were collected and processed using published standard operating procedures (Page *et al.*, 2006). In the prospective blood samples, sodium heparin samples were collected and centrifuged at 1100g for 10 minutes at 4°C. The plasma was subsequently removed into a 15ml falcon tube and centrifuged for a second time at 2000g for 10 minutes at 4°C then transferred to labelled eppendorf tubes leaving all cell debris behind. The buffy coat layer containing peripheral blood mononuclear cells (PBMCs) was also removed from the original tube and transferred into labelled eppendorf tubes and both sample types were stored at -80°C.

## 2.4 Epigenomics

### *2.4.1 Clinical material processing*

#### 2.4.1.1 DNA extraction and quantification

PBMC DNA was extracted using the QIAamp DNA Blood Mini Kit (Qiagen) according to the manufacturer's protocol. Quantification of PBMC DNA was performed by Absorbance (OD A260nm) on the Nanodrop ND1000 instrument (Thermo Scientific). The DNA was eluted into 200µl buffer.

Plasma DNA was extracted using the QIAamp Circulating Nucleic Acid Kit (Qiagen). A starting volume of between 1-5 ml was extracted depending on volume of sample. The DNA was eluted in 50µl buffer. Due to the low concentrations of plasma DNA, quantification was performed using the Quant-iT PicoGreen dsDNA Assay kit (Invitrogen) as per the manufacturer's protocol. A standard curve was produced using standard solutions between 0.05 ng/µl and 20 ng/µl and fluorescence determined at an excitation filter of 485-12 and emission filter 520 using a LUMIstart optima microplate reader (BMG labtech). Standards and samples were run in triplicate and the average concentration calculated from the standard curve (fluorescence against concentration) using the linear regression equation after subtraction of the average blank reading.

#### 2.4.1.2 Bisulphite conversion

Extracted DNA was bisulphite converted using the EZ-DNA Methylation-Gold™ Kit (200rxns format) (Zymo Research) as per the manufacturer's protocol. Up to 500ng of genomic DNA was converted for each sample depending on the concentration following the extraction. For the conversion the following thermal cycler conditions were used; 98°C for 10 minutes, 64°C for 2.5 hours, 4°C storage up to 20 hours. During this process the DNA is temperature denatured and the cytosine nucleotides react with sodium bisulphite to convert the unmethylated cytosines into uracil. The temperature denatured bisulphite converted DNA is now bound to a column where it is desulphonated, washed and eluted in 15µl-50µl of buffer.

### *2.4.2 Pyrosequencing*

#### 2.4.2.1 Primers and PCR and electrophoresis

Primers were designed using PyroMark Assay Design software 2.0 to cover the CpG site of interest in the candidate loci and purchased from Invitrogen. The primer sequences are listed in Table 2.1.

Table 2.1. Primer sequences of candidate loci for PCR and pyrosequencing

*Biotin labelled primer

| Gene & Location | Primer | Sequence |
|---|---|---|
| FAM113B cg05596756 | Forward | AGGAAGATGTTGTTTTTGTATTGA |
| | Reverse | ACAAACRACACAAATCAACATTTAATTTCT* |
| | Sequence | TTTTTATATATGTTAAAGTTTTGAA |
| | Sequence to analyse | GTTTYGGTTA TTAGTTTTTY GTTTTTTAAA GTAAGAAATT AAATGTT |
| CD6 cg07380416 | Forward | GTGAGATATTTATAGGTTGGGTTTGA |
| | Reverse | CAAATCTACTCTACCCTTTACTATTCTT* |
| | Sequence | ATAGGTTGGGTTTGAT |
| | Sequence to analyse | YGTATGYGTG TYGGAGAGGA GAGAGTA |
| TMEM149 cg16927606 | Forward | GTTAGGGTTTTGGGTTTTAAGTT |
| | Reverse | TTCCCTCCACCCACACATAT* |
| | Sequence | TTTGGGTTTTAAGTTGG |
| | Sequence to analyse | TYGTGTTAAG TTTTTTAGTT GAAGTTGT |
| KLHL6 cg20425130 | Forward | TGGTAGAGATTTTAAATGGGGAAAAGGT |
| | Reverse | ATATCCACACACAAAATAACATCTATC* |
| | Sequence | TTTAGAATGGTTTGGAAATT |
| | Sequence to analyse | TTGYGAATGG AAAAYGTTTT GATAGATGTT |
| ZNF-154 cg21790626 | Forward | AGGTGTGTTTATAGGATTAGAGATAGTAGA* |
| | Reverse | ACAACCATCCCTATCCCAAACCTA |
| | Sequence | AAAAACCACCACTCCCCTCA |
| | Sequence to analyse | CRCCTTCRTA ACCCCAACTC |
| MGAT1 cg21898046 | Forward | GGTTTGTGGTTATTGATTGATTGTTTATTA* |
| | Reverse | ACCATACCTAATAACCCCTTAACAATTTT |
| | Sequence | CCTTTATTTCTCTAAAACCC |
| | Sequence to analyse | TACACRATTC RTAATTTCAT ATAAAAACAA A |
| HOXA9 cg27009703 | Forward | TGGTGATGGTGGTGGTATA* |
| | Reverse | ACTTCAACCCCTACAACTTCCAATCCAAA |
| | Sequence | CCCCTACAACTTCCAATCCAAAA |
| | Sequence to analyse | CRACRATATTTAA |
| IL22RA1 cg09152089 | Forward | ATGGGTATTTATTAGTTAGGGATTTTATAG |
| | Reverse | AACCCCAAAACTCCCAACCCT* |
| | Sequence | GGATTTTATAGTTAAGATGGTTAG |
| | Sequence to analyse | TAGYGTTTTT ATYGGGGTTG GTATAGAGTT |
| C13orf29 cg22131172 | Forward | AGAGGTATATTTAGTAGGTTGTATGAAAAG |
| | Reverse | TATCCATCAAAAAAAACTTACCCTAAACAT* |
| | Sequence | AAGGAAAGGTATTGGTATGT |
| | Sequence to analyse | TTYGTTTTTT GTGGGTAGTT TGGGGATGA |

Primers were reconstituted in nuclease free water (ThermoFisher Scientific) to a stock concentration of 100µM and further diluted to a working stock of 10µM. Temperature gradients for candidate loci were performed to optimize the thermal cycler annealing temperature conditions to increase yield of DNA during the PCR. For each reaction, the following mix was used: PCR Buffer 4.5µl, $MgCl_2$ 3.6µl, dNTP mix (1mM) 0.9µl, forward and reverse primers (10µmol/L) 2.25µl, FastStart Taq DNA polymerase (5U/µl), 1µl of bisulphite converted DNA was added to each reaction and 30.05µl nuclease free water was added to give a final volume of 45µl.

Thermal cycling conditions were as follows (annealing temperature dependent on specific primers and given below):

Step 1: Hot start (denaturation): 95°C for 10 minutes

Step 2: Touchdown (annealing): 60°C to 50°C for 10 cycles

Step 3: Denaturation: 95°C for 20s

Step 4: Annealing:

FAM113B 53°C for 20s

TMEM149 51°C for 20s

CD6, KLHL6, ZNF154, MGAT1, HOXA9, IL22RA1, c13orf29 57°C for 20s

Step 5: Elongation: 72°C for 20s

Repeat back to Step 2-5 x30

Step 6: Final elongation: 72°C for 5 minutes

Step 6: Final hold: 4°C for 10 minutes

Following PCR, 8µl of the products were run by electrophoresis on a 2% Agarose/1xTBE (Fisher Scientific) gel with 2µl orange loading dye (ThermoFisher Scientific) to confirm the presence of product, size and lack of contamination. 5µl Ethidium Bromide (Invitrogen) (10mg/ml) was added to the gel to visualize DNA when exposed to UV light.

The products were run on the gel at 100V for approximately 30 minutes and visualized under UV light. A 100bp DNA ladder (Invitrogen) was also run on the gel to ensure correct DNA fragment product size as expected with each individual assay.

#### 2.4.2.2 Pyrosequencing reaction and data analysis

Pyrosequencing was performed on a PyroMark Q96 MD (Qiagen) as per the manufacturer's protocol and in-house standard operating procedure (SOP). A brief summary of the protocol is as follows:

5-10µl of PCR product was combined with 2µl Streptavidin Sepharose HP beads (GE Healthcare), 38µl PyroMark binding buffer (Qiagen) and made up to a total volume of 80µl with $dH_20$ in a 96 well plate. This plate was then agitated at 1400rpm for 10 minutes allowing the biotinylated products to be captured onto the Streptavidin beads. 0.3µM sequencing primer containing PyroMark annealing buffer (Qiagen) was added to a 96 well sequencing plate. Following agitation the DNA product is captured onto the filter probes of the Pyromark Q96 Vacuum workstation tool for 15s until all the beads have been captured. The beads are washed in 70% ethanol, denaturing solution (4% w/v Sodium hydroxide) (Sigma Aldrich) and PyroMark wash buffer (Qiagen). The beads are then released onto the sequencing plate by shaking in a circular motion. The sequencing plate is heated for 80$^oC$ for 2 minutes to allow annealing of the sequencing primer to the product. Enzyme, substrate and nucleotides are loaded into the PyroMark Q96 Tips and dispensation was tested on an empty plate. The pyrosequencing plate was then loaded into the PyroMark Q96 MD pyrosequencer (Qiagen) and the reaction started.

Quality control samples were run in parallel including a 100% methylated control (Universal methylated DNA (Millipore)) and 0% methylated control (Whole genome amplification of cell-line DNA) to assess the bisulphite conversion efficiency and a $dH_20$ control to assess contamination in the PCR and pyrosequencing reactions.

Pyrosequencing was performed in duplicate on these controls to assess reproducibility of all the assays. All assays demonstrated a coefficient of variation (CV) of <2%.

Pyrosequencing of the PCR product generates a Pyrogram. PyroMark Q-CpG software was used to calculate the percentage of DNA methylation at that CpG site. This was calculated from the peak height of the methylated cytosines (C) /peak height of C + peak height of the unmethylated cytosines (T) – which are converted to thymine from uracil in the PCR.

Subsequent analysis was performed in Microsoft Excel 2010 and R version 3.2.1 (http://www.r-project.org).

### *2.4.3 Real time-PCR*

Total RNA was converted to complementary DNA (cDNA) using High capacity cDNA Reverse Transcription Kit (Applied Biosystems). Each reaction used 1µg of RNA in a total volume of 50µl giving a final concentration of 20ng/µl. qPCR was performed using Fast SYBR® Green Master Mix

(Applied Biosystems) on the 7900HT Fast Real-Time PCR System (Applied Biosystems) using specific primers to ZNF154, HOXA9 and endogenous control Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) using the following thermocycling conditions:

Step 1: 95$^{o}$C for 20 seconds

Step 2: 95$^{o}$C for 1 second

Step 3: 60$^{o}$C for 20 seconds

Repeat Steps 2-3 for 40 cycles

Step 4: 95$^{o}$C for 15 seconds

Step 5: 60$^{o}$C for 15 seconds

Step 6: 95$^{o}$C for 15 seconds

Primers were designed using the Roche Life Sciences Assay Design Center software (https://lifescience.roche.com/en_gb/brands/universal-probe-library.html.) The primer sequences are listed in Table 2.2.

Table 2.2. Primer sequences of real time PCR.

| Gene | Primer | Sequence |
|---|---|---|
| HOXA9 | Forward | CCCCATCGATCCCAATAA |
| | Reverse | CACCGCTTTTTCCGAGTG |
| ZNF154 | Forward | CAAAGATGCCTGTACCGTGA |
| | Reverse | CAAGGTGCTGCTTCTGATGA |
| GAPDH | Forward | AGCCACATCGCTCAGACAC |
| | Reverse | GCCCAATACGACCAAATCC |

Each reaction was performed following the manufacturers protocol using 10ng of cDNA. Target gene and endogenous controls were performed in triplicate for each sample as well as a no template control to ensure there was no contamination. The ΔCT was calculated using the relative standard curve method and normalised to the endogenous control.

## 2.5 Metabonomics

Untargeted metabonomic profiling using $^{1}$H NMR of plasma and urine samples from patients with epithelial ovarian cancer and an all-female control population was performed. This was coupled with a commercial MS kit based targeted platform from Biocrates Life Sciences AG; the Biocrates Absolute*IDQ*® p150 Kit to quantify over 150 endogenous metabolites in plasma. A list of the metabolites in the Biocrates Absolute *IDQ*® p150 Kit can be found in Appendix 2. Analysis was performed in the MRC-NIHR National Phenome Centre (NPC) as part of a successful application for a pilot study.

### *2.5.1 $^{1}$H Nuclear Magnetic Resonance (NMR) Spectroscopy*

#### 2.5.1.1 Sample preparation

In order to obtain global metabolic profiles, these plasma and urine samples were analysed using $^{1}$H NMR spectroscopy (NMR). Samples were prepared and analysed according to previously published protocols (Dona *et al.*, 2014).

Samples were randomized to avoid batch-effects and aliquots of samples were pooled to create quality control material. There is a study reference (SR) for each matrix which is a composite of all the samples included in the study and helps to elucidate intra-study variation in preparation and analysis. The long term reference (LTR) refers to an NPC-generated control and is a composite of samples that have previously been analysed in the centre and are independent from the study, this helps to elucidate inter-study variation in preparation, sample deterioration or analytical performance (Dona *et al.*, 2014).

##### 2.5.1.1.1 Urine

Urine samples were thawed and centrifuged at 12,000 x g at 4°C for 5 minutes. Urine samples were prepared into 96-well plates for NMR spectroscopy using a Bruker Sample Track system and a Liquid Handler 215 preparation robot. To a 96-well plate, 600 µl of urine was added and centrifuged at 1800 x g for 5 minutes. The NMR flow-injection Liquid Handler tool combines 540µl of sample with 60µl urine buffer (see Section 2.7) into the NMR tube. Sodium azide (0.2mM) was added to samples to avoid bacterial contamination.

##### 2.5.1.1.2 Plasma

Plasma samples were thawed and at least 400µl of plasma was centrifuged at 12,000 x g at 4°C for 5 minutes. Plasma samples were prepared into 96-well plates for NMR spectroscopy using a Liquid Handler 215 preparation robot. Again, samples were prepared into 96-well plates by adding 350µl of sample to each well combined with 300µl of plasma buffer (see Section 2.7) using the automated

liquid handler and 600µl added to each NMR tube. Sodium azide (3.1mM) was added to samples to avoid bacterial contamination.

#### 2.5.1.2 Sample acquisition

$^{1}$H NMR spectra were acquired on a 600 MHz Bruker Avance III NMR spectrometer (Bruker Biospin, Karlsrhue, Germany) at 301K for urine and 310K for plasma using a standard one-dimensional water pre-saturation pulse sequence [relaxation delay-gz,1-90°-t-90°-tm-gz,2-90°-acquire free induction decay (FID]. The relaxation delay was 4 seconds, t represents a short delay of approximately 3 µs, 90° represents a 90° RF pulse, tm is the mixing time of 100ms, gz,1 and gz,2 are magnetic field z-gradients applied for 1 ms and the acquisition period was 2.7s. The probe was automatically tuned and matched and locked to deuterated water and automatically shimmed using a tuning routine after the regular shimming.

Two dimensional (2D) *J*-resolved spectra were also acquired for both urine and plasma samples. The Bruker *J*-resolved pulse sequence used was [relaxation delay-90°-t1-180°-t1-acquire FID], where t1 is an incremented time period, the relaxation delay was 2 s and 90° and 180° refer to RF pulses at the degree specified.

In addition 1D CPMG (Carr-Purcell-Meiboom-Gill) spectra were acquired for plasma samples. The CPMG experiment provides more detailed information on lower molecular weight metabolites as the resonances belonging to proteins and other macromolecules are attenuated. The CPMG pulse sequence was [relaxation delay-90°-(t-180°-t)n-acquire FID] where t is the spin-echo delay of 0.3 ms and n is the minimum number of loops, 128. The acquisition parameters were set up in the same way as the one-dimensional water pre-saturation pulse sequence.

#### 2.5.1.3 Pre-processing of $^{1}$H NMR spectra

All spectra were automatically phased and baseline corrected. Spectra were visually inspected and manually adjusted if necessary using TopSpin (v3.1, Bruker BioSpin, Germany). Urine spectra were calibrated to the standard, TSP (3-trimethyl-silyl-[2,2,3,3-$^{2}H_4$] propionic acid, sodium salt (Sigma-Aldrich)) at δ0. Plasma spectra were calibrated to the α-glucose doublet at δ5.23.

The spectra are digitised creating a large data matrix in Matlab (R2014a, MathWorks, USA) for subsequent multivariate analysis. The spectral resonances containing TSP (δ0.0) and water (δ4.7) were removed to avoid confounding the analysis. Data was normalised to account for differences brought about by sample or instrumental factors using the probabilistic quotient normalization (PQN).

#### 2.5.1.4 LipoFIT

$^{1}$H NMR spectra were sent to Bruker BioSpin, Germany to obtain quantitative information regarding the main lipoprotein subfractions and their major lipid constituents in blood plasma using technology patented by LipoFIT Analytic GmbH (now termed numaresHEALTH). Total plasma concentrations of cholesterol, triglycerides and some apolipoproteins were quantified in the $^{1}$H NMR spectra. In addition, lipoprotein subfractions and their lipid consistuent for e.g. triglyceride content of VLDL particles was also quantified. The full list of subfractions of lipoprotein analysed are detailed in Appendix 2.

### *2.5.2 Targeted MS (Biocrates AbsoluteIDQ® p150 Kit)*

The Biocrates Absolute*IDQ*® p150 Kit is a targeted metabolic profiling approach to allow simultaneous (semi)-quantification of 158 metabolites from four classes including acylcarnitines, glycerophospholipids, sphingolipids and the sum of hexoses. The method combines the PITC-derivatisation and extraction of metabolites with isocratic flow injection analysis (FIA) with mass spectrometrical detection using MRM pairs. Isotope labelled and chemical homologous internal standards are integrated in the Absolute*IDQ*® plate filter for quantification of metabolites. The workflow for analysis is summarised below:

1. Register assay in laboratory information system.
    a. 96 well plate overview
2. Assay preparation
    a. Derivatisation with PITC
3. Perform FIA-MS
    a. ESI-MS
4. Convert Mass Spectrometer Data
    a. Files are converted and concentrations of metabolites automatically calculated
5. Validate the Kit Plate
    a. Automated quality assessment of standard, kit quality controls and internal standards
6. Evaluate and Export Data
    a. Results evaluated and exported into Microsoft Excel. Conditional formatting with quality assessment criteria.
7. Statistical analysis
    a. Multivariable – PCA, PLS-DA
    b. Univariate analysis – Mean, SD, %CV, Wilcoxon-rank sum test,
    c. Survival analysis – cox proportional hazard models

#### 2.5.2.1 Sample preparation, assay and data processing

Metabolites were extracted from plasma using a specific 96-well plate system for protein-removal, internal standard normalisation and derivatisation (Biocrates Absolute*IDQ*® p150 Kit) and performed according to the kit User Manual (Biocrates Life Sciences AG, Innsbruck, Austria). In brief, 10 µl of 80 samples were added across 4 plates to the centre of the filter on each of the upper 96-well plate kit and dried for 30 minutes under nitrogen flow. Samples were incubated for 20 minutes with 5% PITC (phenylisothiocynate) for derivatisation of the metabolites. After incubation, the filter spots were dried again under nitrogen flow for 60 minutes to remove excess reagent *in vacuo*. The metabolites were extracted using 300µl of a 5mM ammonium acetate solution in methanol for 30 minutes, shaken at 450 rpm and transferred by centrifugation for 2 minutes at 100 x g into the lower 96-deep well plate. The extracts were diluted with 600 µl of the FIA running solvent for further MS analysis using Waters tandem quadrupole mass spectrometers.

One blank sample (no internal standards and no sample added), three water-based zero samples (phosphate buffered saline), and three kit-based quality control samples as well as 3 LTR and 3 SR quality control samples were also added to each of the 4 Kit plates. The kit quality controls were comprised of human plasma samples containing metabolites, at several concentration levels and the SR and LTR are pooled plasma samples from the study and other studies within the NPC respectively and were used to verify the performance of the assay and mass spectrometer. The kit included a mixture of internal standards for the quantification of the endogenous metabolites. Chemical homologous internal standards were used for the quantification of the glycerophospholipid and sphingomyelin species and were therefore only semi-quantified whereas stable isotopes-labelled internal standards were used to fully quantify the other compound classes. The amount of internal standards in each well was identical and therefore the internal standard intensities of the zero samples were compared to the wells containing samples to allow for conclusions on effects of ion suppression.

Metabolites were analysed by isocratic flow-injection analysis (FIA) using the Waters tandem quadrupole mass spectrometers (Xevo TQ MS) in the NPC. Identification and quantification of the metabolites was achieved using internal standards and multiple reaction monitoring (MRM) detection. Data analysis including calculation of the metabolite concentrations and quality assurance was automated using the Met*IDQ*® software package (Biocrates Life Sciences AG, Innsbruck, Austria). After evaluation, the results were exported into Microsoft Excel and R version 3.0.2 (http://www.r-project.org) for statistical analysis.

## 2.6 Statistical analysis

### *2.6.1 Multivariate statistical analysis*

Multivariate statistical analysis was employed to understand the key sources of variation in the data and identify which metabolites are able to discriminate between classes in the form of the unsupervised method; principal component analysis (PCA) and the supervised method; partial least-squares discriminant-analysis (PLS-DA) and the orthogonal partial least-squares discriminant-analysis (OPLS-DA) using SIMCA-P (v14.0, Umetrics, Sweden).

Statistical total correlation spectroscopy (STOCSY) analysis was used to aid identification of correlation between peaks in the $^{1}$H NMR spectra and thus identify peaks arising from the same molecule and covariance of that molecule with other metabolites in the sample. The identification of these metabolites was then confirmed using in-house databases and the Human metabolome database (HMDB).

The robustness of the models produced were assessed using seven-fold cross validation and permutation analysis comparing the goodness of fit of the original model with the goodness of fit of several models based on data where the order of the Y-observations has been randomly permutated, while the X-matrix has been kept intact.

### *2.6.2 Univariate statistical analysis*

The NMR and targeted MS data were then analysed using univariate statistics in R v3.2.1 including Mann-Whitney U non parametric statistics to test for significant differences in concentration between classes where the threshold for significance was $p<0.05$ adjusting for multiple testing $q<0.05$ using the false discovery rate.

Correlation analysis was performed between individual metabolites that were significantly different using Pearson correlation and Kendall rank correlation for association of metabolites with stage and grade of disease. Spearman correlation was used for correlation analysis of differentially methylated CpG sites. Stepwise logistic regression was used to identify the combination of metabolites which most accurately differentiated malignant ovarian disease from benign according to Akaike information criterion (AIC). Each (thresholded) biomarker is added to the model in order of greatest improvement in model fit until convergence of the model-fitting procedure occurs and no additional biomarker improves the model. These predictive models were internally validated using the leave one out cross validation (LOOCV) method where one sample (the test set) is removed from the data (the training set) and then the test set is predicted based on the training set. This is repeated for all samples.

Finally, diagnostic performance of biomarkers was assessed using sensitivity, specificity and negative and positive predictive values and calculated as shown in Table 2.3. Receiver operating characteristic (ROC) curve analysis was performed using the 'pROC' package in R version 3.2.1.

Table 2.3. Table for calculation of the performance of a classification model, also known as a confusion matrix.

| | | State of disease | | |
|---|---|---|---|---|
| | | Disease + | Disease - | Statistic |
| Results of screening test | Test + | True positive (TP) | False positive (FP) | Positive predictive value (PPV) = TP/TP+FP |
| | Test - | False negative (FN) | True negative (TN) | Negative predictive value (NPV) = TN/TN+FN |
| | Statistic | Sensitivity or true positive rate (TPR) = TP/TP+FN | Specificity or true negative rate (TNR) = TN/TN+FP | Accuracy = TP+TN/TP+FP+FN+TN |

### *2.6.3 Survival analysis*

Univariate survival analysis was performed using the logrank test and multivariable survival analysis by Cox proportional hazard models adjusting for age, stage, grade and presence of residual disease.

# Materials

## 2.7 Chemicals, kits and buffers

### *DNA extraction*

QIAamp Circulating Nucleic Acid Kit (Qiagen)

QIAamp DNA Blood Mini Kit (Qiagen)

### *DNA quantification*

Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen)

### *Electrophoresis*

Agarose genetic analysis grade (Fisher Scientific)

TBE 10x (Imperial College laboratories)

Ethidium Bromide 10mg/ml (Invitrogen)

100bp DNA ladder (Invitrogen)

Orange DNA loading dye (ThermoScientific)

### *Bisulphite conversion*

EZ-DNA Methylation-Gold™ Kit (Zymo research)

### *PCR*

PCR FastStart Taq DNA Polymerase Kit (Roche)

Nuclease free water (ThermoFisher Scientific)

Universal methylated DNA (Millipore)

Illustra GenomiPhi V2 DNA Amplification kit (GE Healthcare)

### *Pyrosequencing*

PyroMark Gold Q96 reagents (Qiagen)

PyroMark Binding buffer (Qiagen)

PyroMark Annealing buffer (Qiagen)

PyroMark Wash buffer (Qiagen)

Sodium hydroxide pellets (Sigma-Aldrich)

Streptavidin Sepharose High performance beads (GE Healthcare)

### *Real-Time PCR*

High capacity cDNA Reverse Transcription Kit (Applied Biosystems)

Fast SYBR® Green Master Mix (Applied Biosystems)

### *$^{1}$H NMR*

Urine buffer

$KH_2PO_4$, Potassium dihydrogen phosphate, 99.99% ACS, anhydrous (Sigma-Aldrich)

TSP, 3-trimethyl-silyl-[2,2,3,3- $^{2}H_4$]propionic acid, sodium salt 98 atom % D (Sigma-Aldrich)

$NaN_3$, Sodium azide, 99.5% (Sigma-Aldrich)

KOH, Potassium hydroxide, ≥85% KOH (Sigma-Aldrich)

$D_2O$, Deuterium oxide, 99 atom % D (Sigma-Aldrich)

Plasma buffer

$NaH_2PO_4$, Disodium hydrogen phosphate, 99% anhydrous (Sigma-Aldrich)

TSP, 3-trimethyl-silyl-[2,2,3,3- $^{2}H_4$]propionic acid, sodium salt 98 atom % D (Sigma-Aldrich)

$NaN_3$, Sodium azide, 99.5% (Sigma-Aldrich)

NaOH, Sodium hydroxide, ≥97% NaOH (Sigma-Aldrich)

$D_2O$, Deuterium oxide, 99 atom % D (Sigma-Aldrich)

### *MS*

Biocrates Absolute*IDQ*® p150 Kit (Biocrates Life Sciences AG)

## 2.8 Equipment & plasticware

### *DNA quantification*

DNA quantification Nanodrop ND1000 (ThermoScientific)

LUMIstar optima microplate reader (BMG labtech)

### *Electrophoresis*

PCR Electrophoresis tank and power supply (Fisher Scientific)

UV light reader (uvitec)

### *Pyrosequencing*

Pyrosequencing PyroMark Q96 MD (Qiagen)

PyroMark Q96 Vacuum Workstation (Qiagen)

PyroMark Q96 Dispensing Tip Holder (Qiagen)

PyroMark Q96 Capillary & Reagent Tips (Qiagen)

PyroMark Q96 HS Plate (Qiagen, Hilden, Germany)

### *Real-Time PCR*

7900HT Fast Real-Time PCR System (Applied Biosystems)

### *$^{1}$H NMR*

600MHz Bruker Avance III NMR spectrometer (Bruker Biospin)

Bruker Sample Track System (Bruker Biospin)

Liquid Handler 215 preparation robot (Gilson Inc)

### *MS*

Waters Xevo tandem quadrupole MS (Xevo TQ MS) (Waters)

### *General Equipment & Plasticware*

Thermal cycler (G-Storm)

Heraeus multifuge X3R centrifuge (ThermoScientific)

Timix Plate Shaker (Edmund Bühler GmbH)

Dri-block (Techne)

Heraeus Pico 17 microfuge (ThermoScientific)

Sodium heparin tubes (BD Vacutainer)

DNA loBind 1.5ml tube (Eppendorf)

0.2ml PCR tubes (Starlab UK)

96 well PCR plate and adhesive covers (Starlab UK)

Graduated filter tips of various sizes (Starlab UK)

Stripettes (various sizes) (Corning B.V. Life Sciences)

Falcon tubes (15ml and 50ml) (VWR)

# CHAPTER 3. Cell-free circulating DNA methylation as a blood based diagnostic biomarker of epithelial ovarian cancer

## 3.1 Introduction

This chapter explores the differential DNA methylation between high grade serous ovarian cancer (HGSOC) and peripheral blood mononuclear cells (PBMCs) using publically available genome-wide methylation datasets. Evaluation of these candidate differentially methylated loci is performed in independent datasets using bisulphite pyrosequencing and their utility as diagnostic blood based biomarkers is explored.

## 3.2 Background

Ovarian cancer is often diagnosed in the late stage with poor prognosis and poor overall survival. In contrast diagnosis in the early stage has a considerably better prognosis. More sensitive biomarkers are required for early detection of ovarian cancer to reduce mortality from the disease. Identification and validation of biomarkers for diagnosis are required to improve on the performance of the currently available blood based biomarker, CA 125 and the recently developed Risk of Ovarian Cancer Algorithm (ROCA) which measures an increase in CA 125 from a women's baseline level. A combination of multiple independent biomarkers contributing to a robust risk algorithm would increase the overall sensitivity and provide a screening tool that could reduce disease-specific mortality.

Analysis of cell-free circulating DNA in patient body fluid such as plasma, urine or saliva provides an opportunity for non-invasive sampling of the tumour through detection of circulating tumour DNA (ctDNA). This 'liquid biopsy' allows for detection of tumour-specific alterations in the DNA such as; chromosomal alterations, structural rearrangements, sequence mutations and epigenetic changes and can be applied to disease diagnosis, prognosis and treatment monitoring.

Epigenetic alterations such as DNA methylation are a common feature of cancer. These stable, covalent modifications consistently occur at specific CpG sites in tumour DNA and make DNA methylation detection in non-invasive body fluid an ideal candidate biomarker.

The disadvantage of DNA methylation analysis in blood is the presence of DNA from peripheral blood mononuclear cells (PBMCs) such as leukocytes which can lyse during sample collection and processing. While other studies have investigated differential methylation between tumour and normal tissue as potential biomarkers, the majority of DNA present in blood will be from PBMCs rather than 'normal' ovarian tissue therefore in this chapter, data from TCGA and others are explored to identify differentially methylated CpG sites between serous ovarian tumours and PBMCs. The methylation of these candidate biomarkers are then verified and validated in independent datasets of solid frozen tumour tissue and PBMCs and finally their performance as

diagnostic biomarkers of HGSOC in blood is assessed. A flowchart of the study design and associated samples is shown in Figure 3.1.

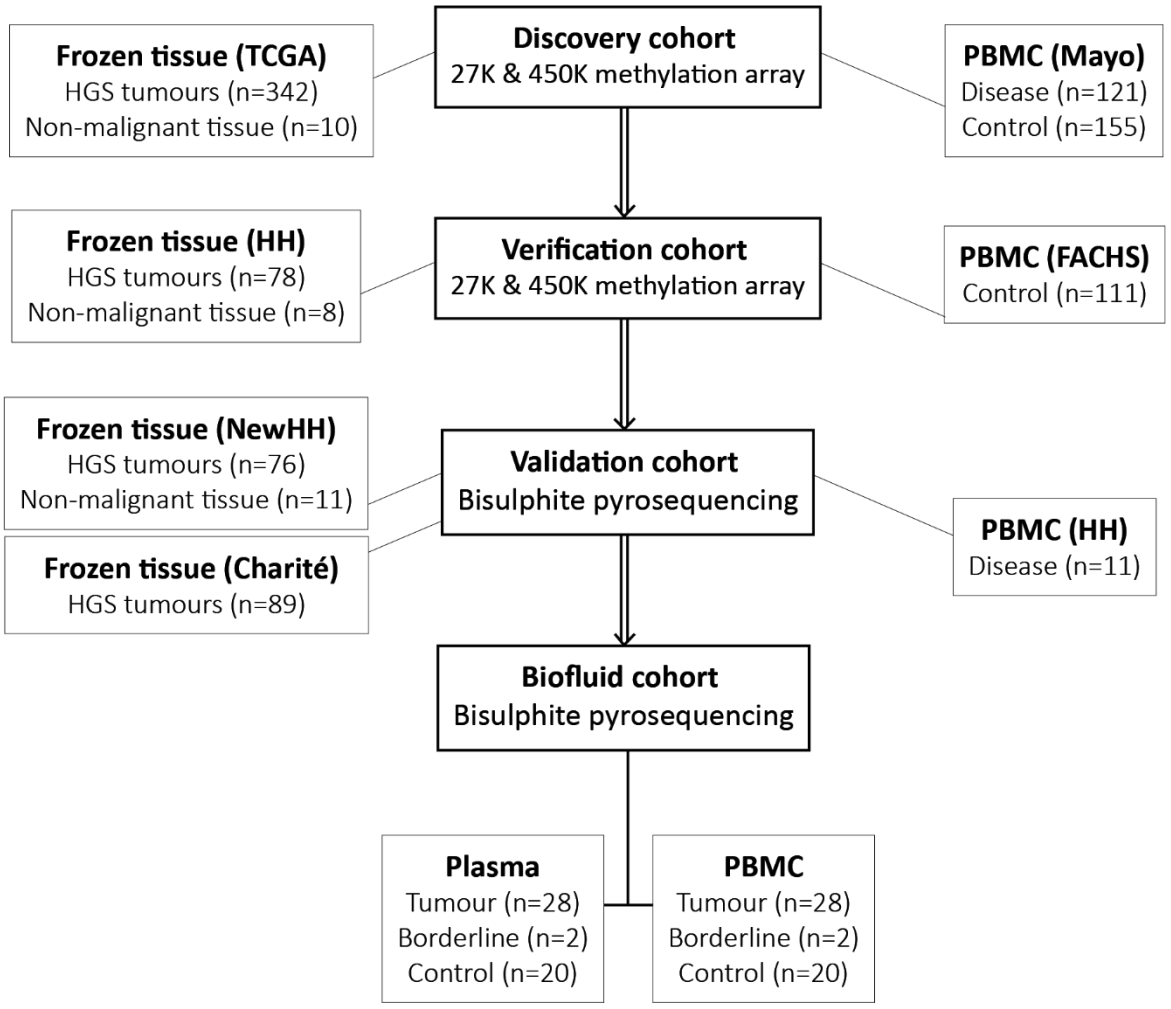



**Figure 3.1** Flowchart of study design for cfDNA methylation biomarker identification

## 3.3 Results

### 3.3.1 Discovery of differentially methylated loci between HGSOC and PBMCs

In order to identify locus-specific DNA methylation that can be used to differentiate HGSOC ctDNA from PBMC DNA in blood, data from Infinium HumanMethylation arrays of 276 PBMC samples were compared to 342 high-grade serous ovarian tumours, the majority of which were Stage 3C plus 10 'normal'/non-cancerous ovarian tissue samples from TCGA termed the 'discovery cohort' (see Figure 3.1). Table 3.1 shows the clinical characteristics for the patients included in the discovery cohort. Candidate loci were selected based on their differential methylation (DM) between tumours and PBMCs with hypermethylation defined by a median methylation beta-value >0.8 and hypomethylation loci defined by a median methylation beta-value <0.1. Candidate loci were selected if they were hypermethylated in one cohort and hypomethylated in the other.

The filtering of the probes in the discovery cohort for selection of the candidate loci is shown in Figure 3.2. Seven loci corresponding to seven different genes were shown to have a median beta-value >0.8 in the HGSOC cohort and a median beta-value <0.1 in the PBMC cohort. Conversely, only two loci from two genes were shown to have a median beta-value <0.1 in the HGSOC cohort and a median beta-value >0.8 in the PBMC cohort. A total of nine candidate cg sites were therefore identified that were differentially methylated between tumours and PBMCs. The box and whisker plot with overlaid scatter plot of the methylation values of the nine candidate cg sites is shown in Figure 3.3. The median, range and standard deviation of these candidate sites in the discovery cohort are shown in Table 3.2.

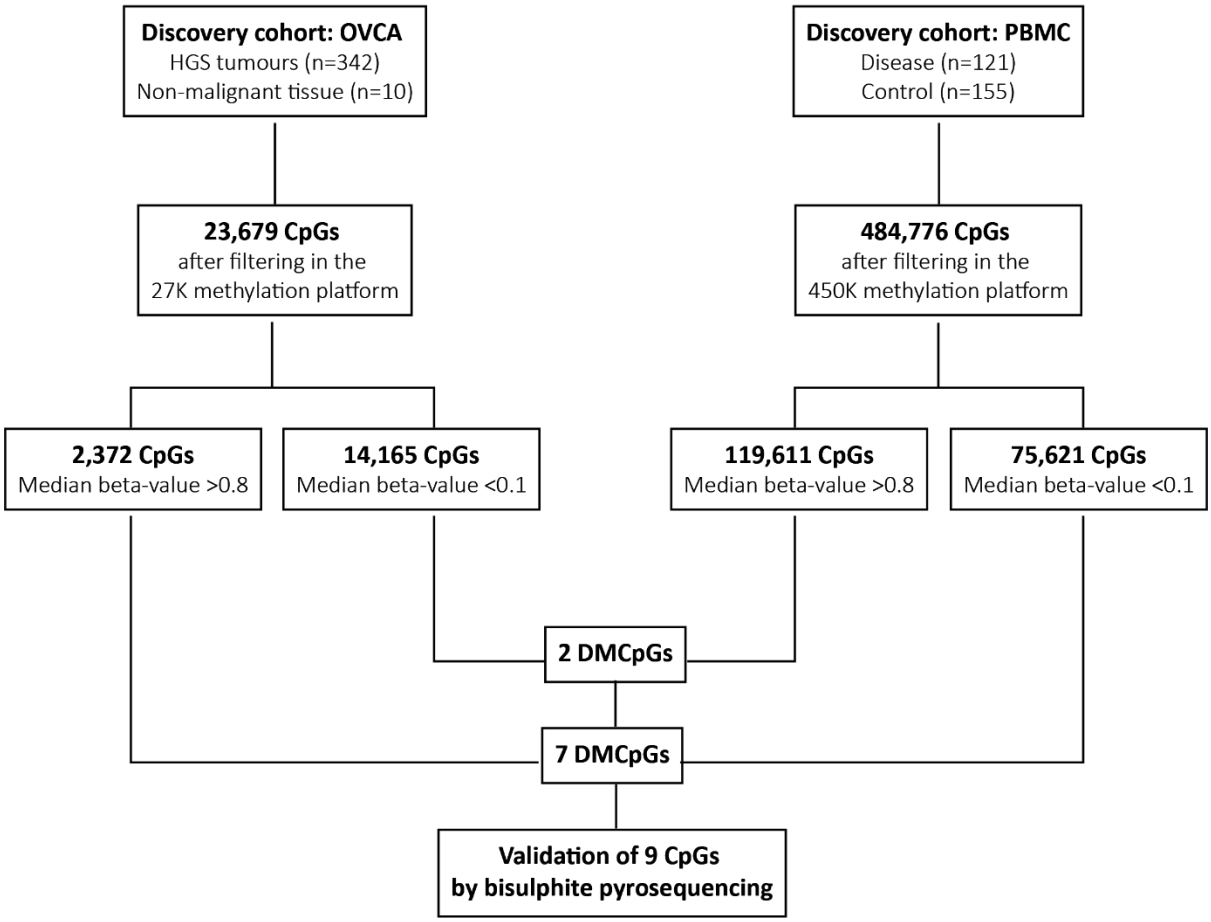



**Figure 3.2** Flowchart of candidate probe selection in the discovery cohort.

**Table 3.1** Clinical characteristics of study participants used in the discovery cohort.

| | | **TCGA** | | **Mayo clinic PBMC** | |
|---|---|---|---|---|---|
| | | Tumour (n=342) | 'Normal' (n=10) | Tumour (n=121) | Normal (n=155) |
| **Age** | Mean (range) | 60 (35-89) | NA | 62 (33-86) | 62 (33-88) |
| **Grade** | 1 | 2 | | 7 | |
| | 2 | 38 | | 9 | |
| | 3 | 295 | | 103 | |
| | Unknown | 7 | | 2 | |
| **Stage** | 1A,1B,1C | 7 | | 15 | |
| | 2A,2B,2C | 14 | | 6 | |
| | 3A | 3 | | | |
| | 3B | 18 | | 77 | |
| | 3C | 243 | | | |
| | 4 | 55 | | 23 | |
| | Unknown | 2 | | 0 | |
| **Histological type** | Serous | 342 | | 100 | |
| | Mucinous | 0 | | 3 | |
| | Endometrioid | 0 | | 15 | |
| | Clear cell | 0 | | 3 | |

**Table 3.2** Details of the nine differentially methylated candidate cg sites in tumour (TCGA) and PBMCs (Mayo clinic) datasets.

| Probe ID | Gene name | TCGA tumour median methylation (beta-value) | SD | TCGA normal median methylation (beta-value) | SD | Mayo median methylation Ca (beta-value) | SD | Mayo median methylation normal (beta-value) | SD |
|---|---|---|---|---|---|---|---|---|---|
| **cg05596756** | FAM113B | 0.834 (0.411-0.936) | 0.080 | 0.830 (0.644-0.873) | 0.079 | 0.090 (0.066-0.133) | 0.011 | 0.088 (0.065-0.142) | 0.012 |
| **cg07380416** | CD6 | 0.886 (0.605-0.955) | 0.055 | 0.882 (0.747-0.899) | 0.054 | 0.087 (0.058-0.118) | 0.011 | 0.085 (0.060-0.120) | 0.011 |
| **cg16927606** | TMEM149 | 0.878 (0.639-0.951) | 0.053 | 0.850 (0.722-0.886) | 0.060 | 0.100 (0.074-0.127) | 0.010 | 0.097 (0.076-0.132) | 0.010 |
| **cg20425130** | KLHL6 | 0.849 (0.123-0.935) | 0.076 | 0.804 (0.631-0.844) | 0.067 | 0.085 (0.068-0.112) | 0.009 | 0.086 (0.065-0.114) | 0.009 |
| **cg21790626** | ZNF154 | 0.847 (0.007-0.984) | 0.218 | 0.075 (0.044-0.675) | 0.230 | 0.058 (0.025-0.120) | 0.018 | 0.059 (0.023-0.138) | 0.025 |
| **cg21898046** | MGAT1 | 0.802 (0.017-0.970) | 0.251 | 0.743 (0.518-0.809) | 0.095 | 0.073 (0.058-0.103) | 0.008 | 0.070(0.054-0.092) | 0.008 |
| **cg27009703** | HOXA9 | 0.818 (0.012-0.967) | 0.226 | 0.349 (0.203-0.719) | 0.151 | 0.083 (0.037-0.242) | 0.038 | 0.096 (0.034-0.320) | 0.055 |
| | | | | | | | | | |
| **cg09152089** | IL22RA1 | 0.097 (0.016-0.671) | 0.083 | 0.133 (0.085-0.324) | 0.077 | 0.801 (0.615-0.861) | 0.040 | 0.802 (0.691-0.866) | 0.032 |
| **cg22131172** | C13orf29 | 0.093 (0.024-0.362) | 0.063 | 0.182 (0.114-0.395) | 0.093 | 0.857 (0.760-0.894) | 0.024 | 0.857 (0.803-0.895) | 0.016 |

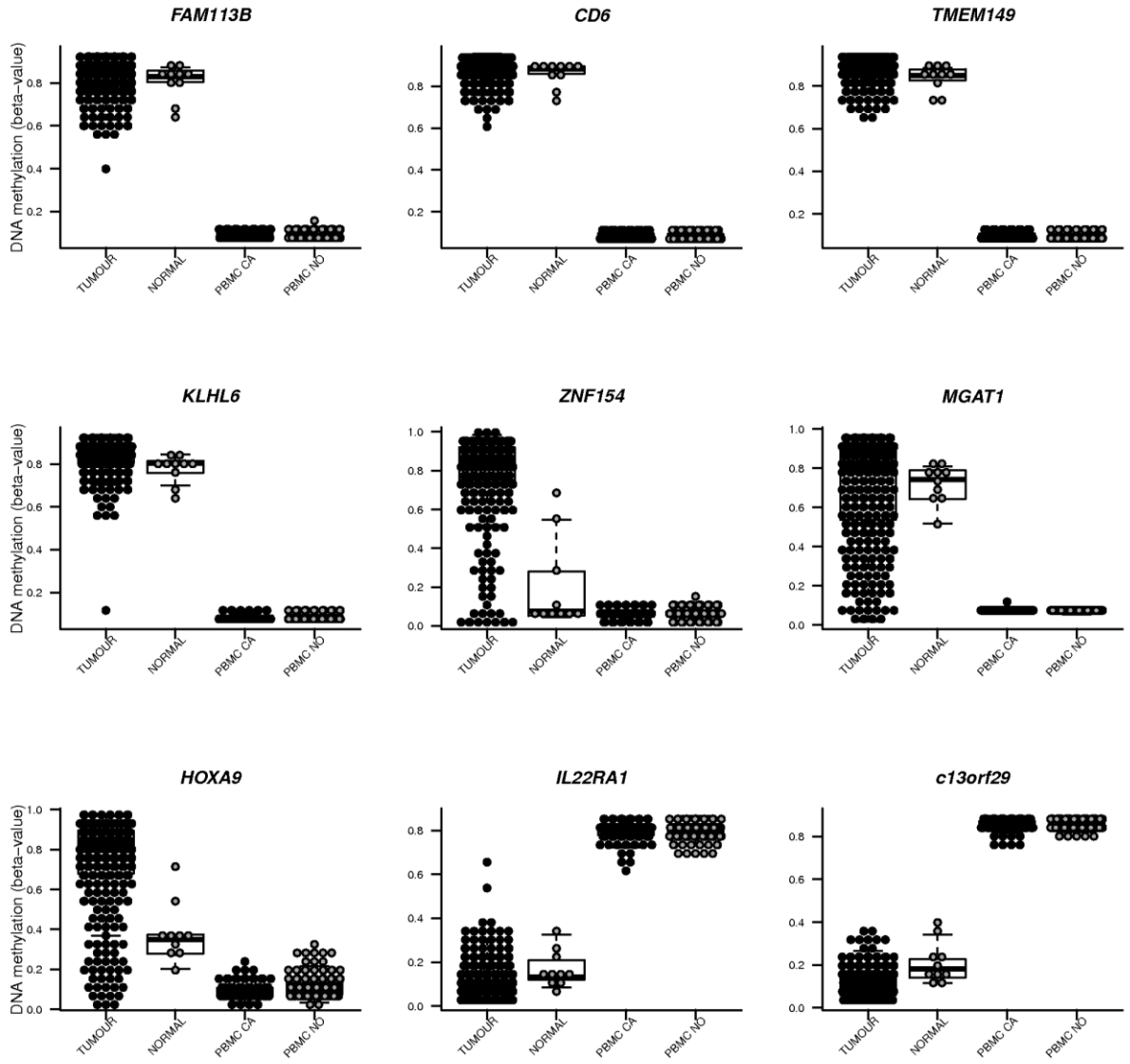



**Figure 3.3** Box and whisker plots of the candidate cg sites in the discovery cohort. (Serous ovarian tumours from TCGA and PBMCs from the Mayo clinic).

### 3.3.2 Verification of differentially methylated candidate loci

Verification of the selected methylation biomarkers was conducted using Infinium HumanMethylation 27 BeadChip data from a cohort of patients collected at Hammersmith Hospital, London termed the HH cohort and from Infinium HumanMethylation 450 BeadChip data published by Dogan *et al*., from the FACHS study looking at the effect of smoking on DNA methylation of PBMCs in African American women (Dogan *et al.*, 2014). Together these datasets form the verification cohort. In addition to the 78 tumour samples, 8 'normal' samples were also analysed. Two samples of ovary and fallopian tube were taken from 4 patients with benign pathologies including; myometrial fibroids, fibrothecoma, adenomyosis and benign leiomyomas. The clinical characteristics of the samples in this cohort are shown in Table 3.3.

For the cancer-related hypermethylated genes that were previously selected for methylation beta-value >0.8 in HGSOC and <0.1 in PBMCs, all loci had a beta-value <0.1 in the PBMC dataset of the verification cohort except for *HOXA9* which had a median value of 0.143. Cg sites in genes; *FAM113B*, *CD6* and *KLHL6* met the criteria set out in the initial discovery set (median methylation beta-value >0.8 in tumours and <0.1 in PBMCs), however the remaining four did not meet these strict criteria with lower median methylation levels in the tumour. The cg site with the lowest median methylation of the 7 candidate hypermethylated loci was *MGAT1* with a median methylation of 0.646 (Figure 3.4). Although these four genes did not verify the results seen in the discovery cohort, one, *TMEM149* showed complete separation between HGSOC and PBMCs, supporting it's suitability as a candidate biomarker and the other three genes, *ZNF154*, *MGAT1* and *HOXA9* showed some degree of differential methylation between HGSOC and benign tissue (Figure 3.4).

For PBMC-specific hypermethylated genes that were previously selected with a beta-value <0.1 in HGSOC and beta-value >0.8 in PBMCs, the two genes had median beta-values of 0.236 and 0.123 in the HGSOC cohort and 0.737 and 0.774 in the PBMC cohort (Figure 3.4). Although these two genes did not meet the criteria, there is still a complete separation in methylation values between HGSOC and PBMCs (Figure 3.4), supporting their suitability as candidate biomarkers. The median, range and standard deviation of these candidate sites in the verification cohort are shown in Table 3.4.

Of the nine candidate loci, two cg sites are located within CpG islands. In the discovery and verification cohorts both of these cg sites, cg21790626 in the *ZNF154* gene and cg27009703 in the *HOXA9* gene showed differential methylation between HGSOC and 'normal/benign' ovarian tissue (Figures 3.3 and 3.4). The median methylation of cg21790626 in the non-malignant group of the discovery cohort was 0.075 and 0.068 in the verification cohort. The median methylation of cg27009703 was 0.349 in the discovery cohort and 0.304 in the verification cohort. In the discovery

cohort, using a beta-value cut-off of >0.2 to diagnose tissue as ovarian cancer, methylation of cg21790626 was able to differentiate primary tumour and non-malignant tissue with a sensitivity of 94% and a specificity of 91%.

The results from the verification study further suggest the suitability of the 9 candidate loci as potential biomarkers for detection of HGSOC in plasma and therefore all of the markers were selected for validation by bisulphite pyrosequencing and assays were successfully designed for all 9 candidate loci.

**Table 3.3** Clinical characteristics of study participants in the verification cohort.

| | | **HH cohort** | |
|---|---|---|---|
| | | Tumour (n=78) | Benign (n=8†) |
| **Age** | Mean (range) | 61 (30-87) | 57 (44-69) |
| **Grade** | 1 | 2 | |
| | 2 | 15 | |
| | 3 | 61 | |
| | Unknown | 0 | |
| **Stage** | 1A,1B,1C | 0 | |
| | 2A,2B,2C | 0 | |
| | 3A | 0 | |
| | 3B | 5 | |
| | 3C | 53 | |
| | 4 | 20 | |
| | Unknown | 0 | |
| **Histological type** | Serous | 78 | |
| | Mucinous | 0 | |
| | Endometrioid | 0 | |
| | Clear cell | 0 | |

†4 patients with two samples; one from ovary and one from fallopian tube

**Table 3.4** Details of the nine differentially methylated candidate cg sites in the verification cohort. (HGSOC tumours and normal tissue from the HH cohort and PBMCs from the FACHS study).

| Probe ID | Gene name | HH tumour median methylation (beta-value) | SD | HH normal median methylation (beta-value) | SD | PBMC median methylation normal (beta-value) | SD |
|---|---|---|---|---|---|---|---|
| **cg05596756** | FAM113B | 0.816 (0.368-0.918) | 0.099 | 0.886 (0.836-0.928) | 0.031 | 0.036 (0.021-0.059) | 0.008 |
| **cg07380416** | CD6 | 0.866 (0.514-0.937) | 0.075 | 0.908 (0.851-0.932) | 0.028 | 0.035 (0.023-0.087) | 0.009 |
| **cg16927606** | TMEM149 | 0.769 (0.370-0.883) | 0.086 | 0.820 (0.767-0.833) | 0.027 | 0.053 (0.036-0.081) | 0.008 |
| **cg20425130** | KLHL6 | 0.814 (0.435-0.906) | 0.081 | 0.838 (0.754-0.908) | 0.057 | 0.051 (0.038-0.085) | 0.008 |
| **cg21790626** | ZNF154 | 0.688 (0.071-0.925) | 0.216 | 0.068 (0.056-0.108) | 0.016 | 0.045 (0.012-0.229) | 0.043 |
| **cg21898046** | MGAT1 | 0.646 (0.105-0.909) | 0.203 | 0.503 (0.256-0.598) | 0.128 | 0.030 (0.014-0.181) | 0.015 |
| **cg27009703** | HOXA9 | 0.661 (0.058-0.884) | 0.186 | 0.304 (0.180-0.516) | 0.116 | 0.143 (0.050-0.368) | 0.063 |
| | | | | | | | |
| **cg09152089** | IL22RA1 | 0.236 (0.058-0.607) | 0.110 | 0.178 (0.086-0.274) | 0.079 | 0.737 (0.657-0.806) | 0.026 |
| **cg22131172** | C13orf29 | 0.123 (0.035-0.571) | 0.087 | 0.164 (0.109-0.230) | 0.048 | 0.774 (0.683-0.833) | 0.028 |

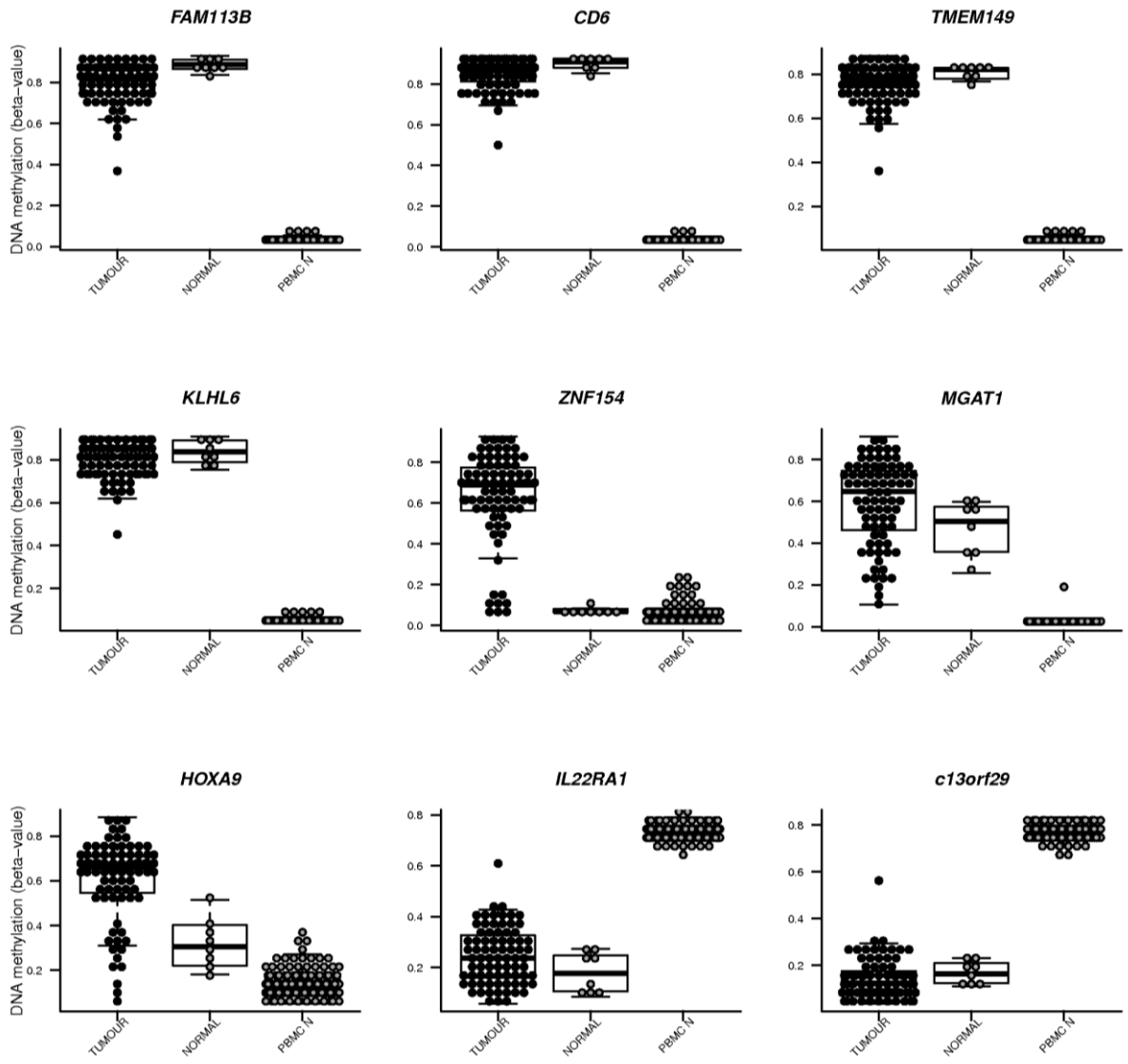



**Figure 3.4.** Box and whisker plots of the candidate cg sites in the verification cohort. (HGSOC tumours and normal tissue from the HH cohort and PBMCs from the FACHS study).

### 3.3.3 Development of pyrosequencing assays for the differentially methylated candidates

As the array based technology is inappropriate for a routine diagnostic test, technical validation of the candidate biomarkers was performed using bisulphite-pyrosequencing. Pyrosequencing assays were designed and optimised for measurement of DNA methylation in cell-free circulating DNA from plasma samples.

In the development and optimisation of the pyrosequencing assays DNA was extracted from 5 different cell-lines including ovarian adenocarcinoma cell-lines; PEA1 and SK-OV3 as well as the normal ovarian surface epithelial cell-line; OSEC and two cell-lines derived from fallopian tubes from healthy controls, FT190 and FT194. Methylation levels were assessed in these 5 cell-lines using the developed bisulphite pyrosequencing assays. Another control, commercially available female DNA, was also assessed using the pyrosequencing assays. Methylation of these cell-lines and the normal female control for the 9 candidate cg sites is shown in Figure 3.5.

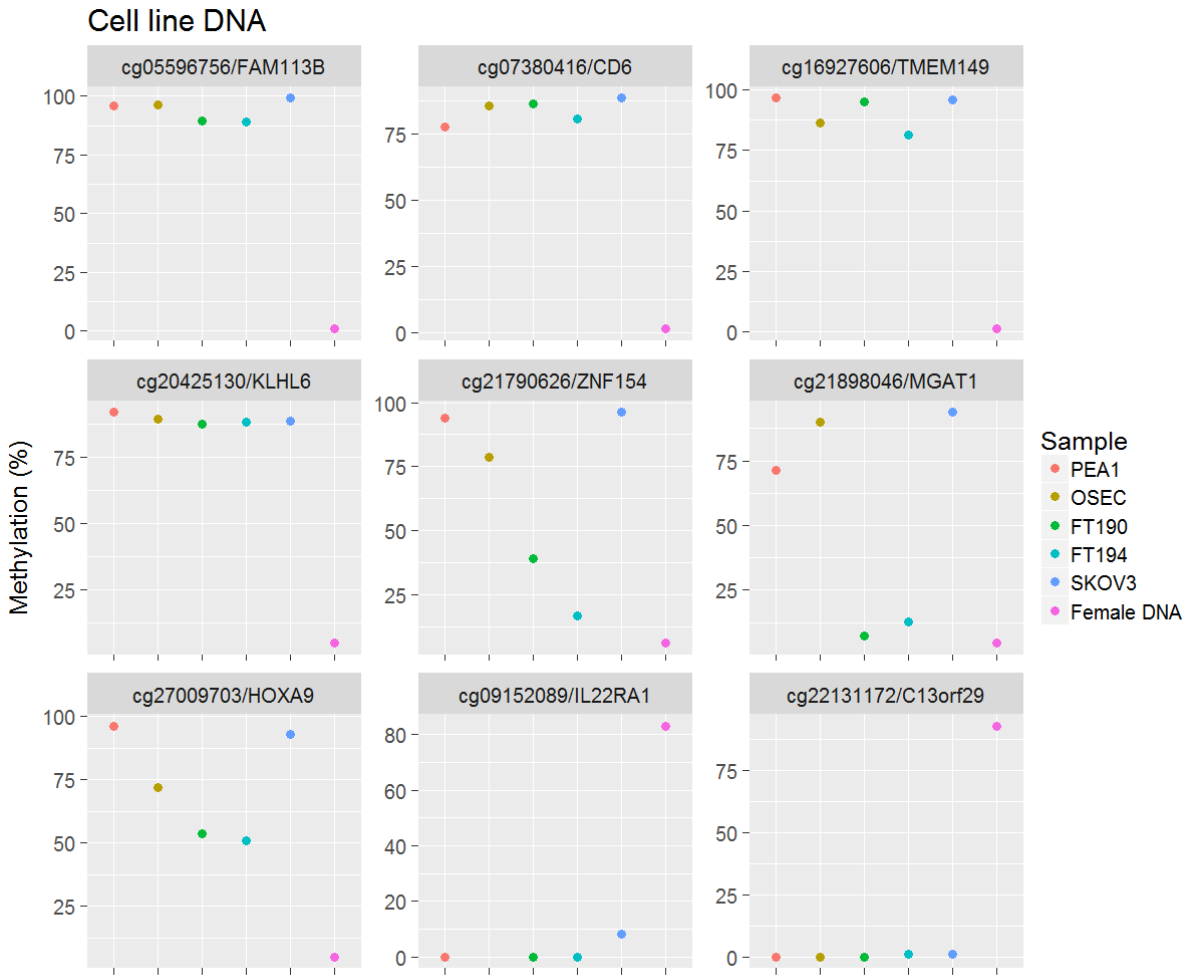
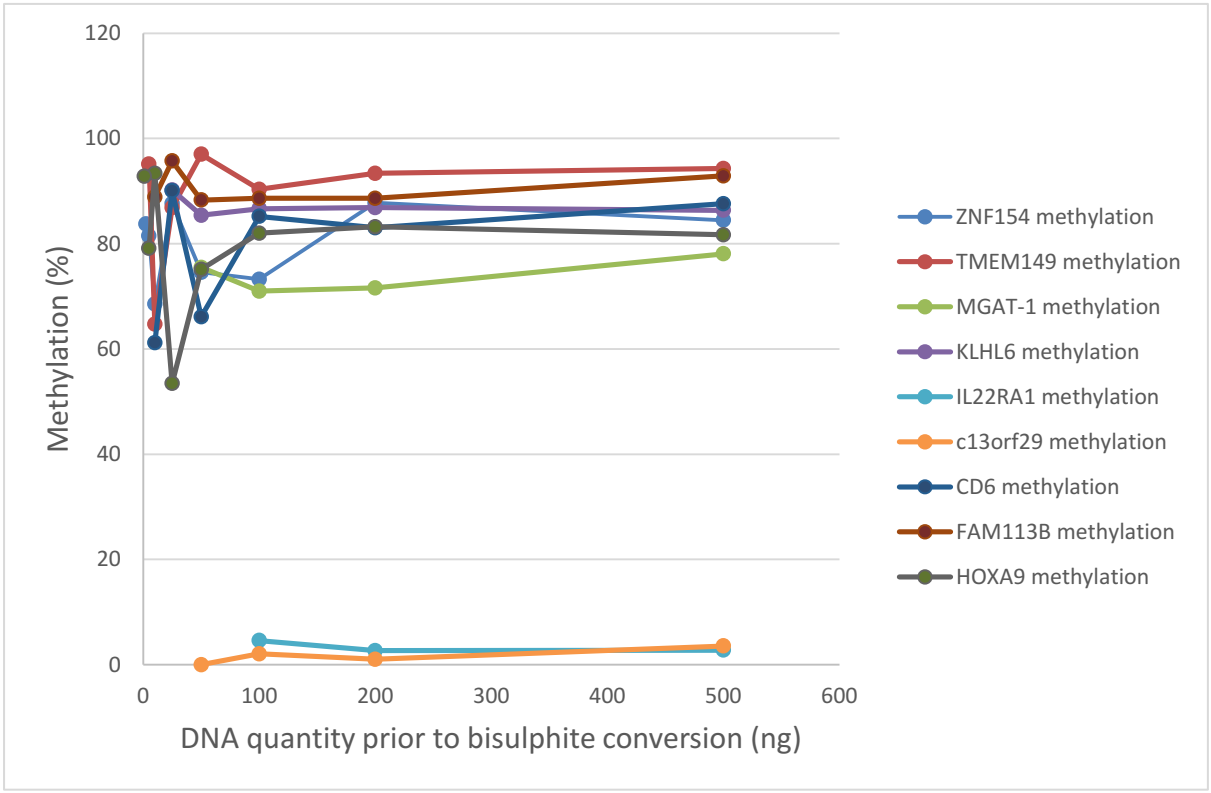



**Figure 3.5.** Scatterplot of the candidate cg sites in DNA extracted from cell-lines using bisulphite pyrosequencing.

As the amount of ctDNA present in plasma samples is considerably less than the tumour itself, the sensitivity and performance of individual assays in the candidate panel were assessed. A tumour DNA sample from the HH cohort was serially diluted from 500ng, the recommended DNA amount for bisulphite conversion as per the manufacturer's protocol. The methylation value for each dilution was then compared to the methylation value determined at 500ng by pyrosequencing. Figure 3.6A shows the % methylation of each locus in the candidate panel at decreasing amounts of DNA. Bisulphite conversion of 200ng DNA demonstrated consistent results compared to 500ng, with less than 10% change in methylation. A delta change in methylation of 15% or less was shown for 100ng of DNA (Figure 3.6B), demonstrating that bisulphite conversion of this amount of DNA may be used to produce reproducible results by pyrosequencing.

A




B

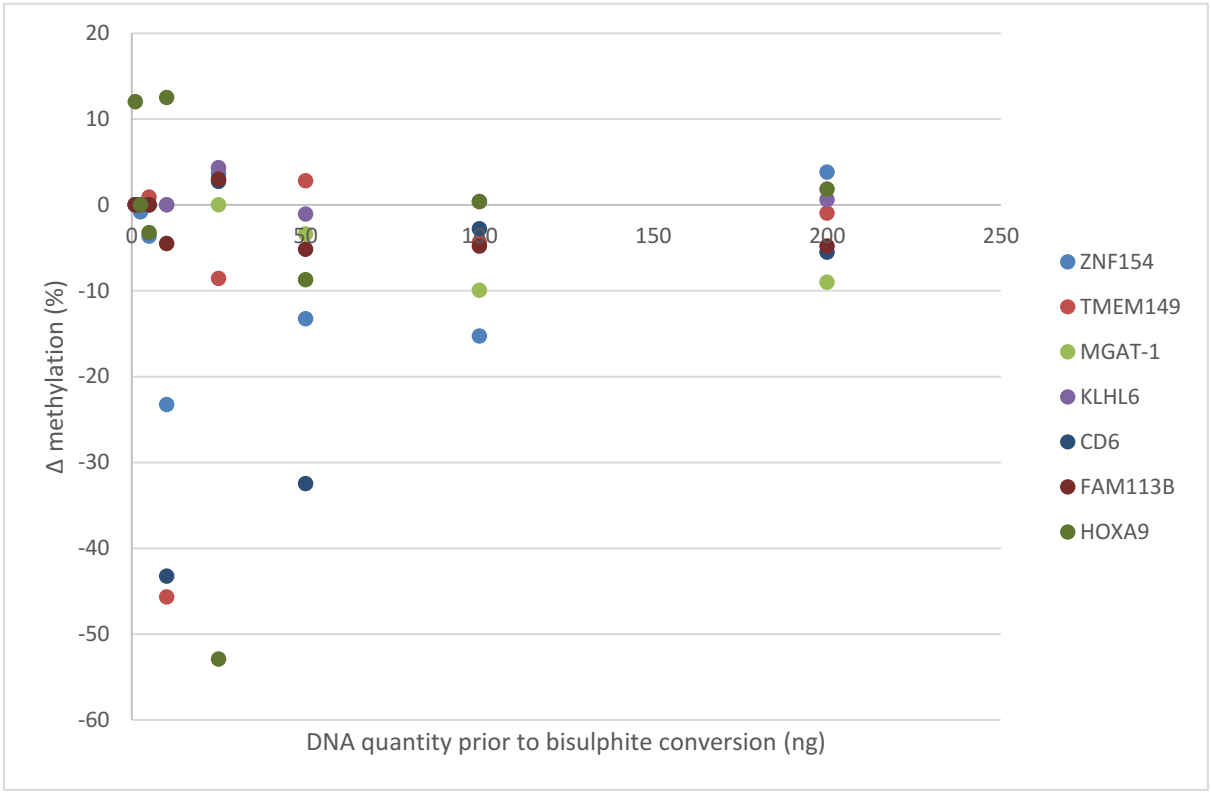



**Figure 3.6** Graphical representation of methylation of serially diluted DNA; 200ng, 100ng, 50ng, 25ng, 10ng, 5ng and 2.5ng using bisulphite pyrosequencing. **A**, methylation values **B**, change in methylation (Δ) compared to 500ng DNA.

High grade serous tumours from the HH cohort were selected to assess the correlation between measurement of cg21790626 methylation by bisulphite pyrosequencing and by Illumina Infinium 27K methylation array. As shown in Figure 3.7 the two methods show very good correlation (rho 0.871, p<0.001) and highlight that bisulphite pyrosequencing is an appropriate method to detect methylation and accurately reproducing equivalent methylation values from the methylation array.

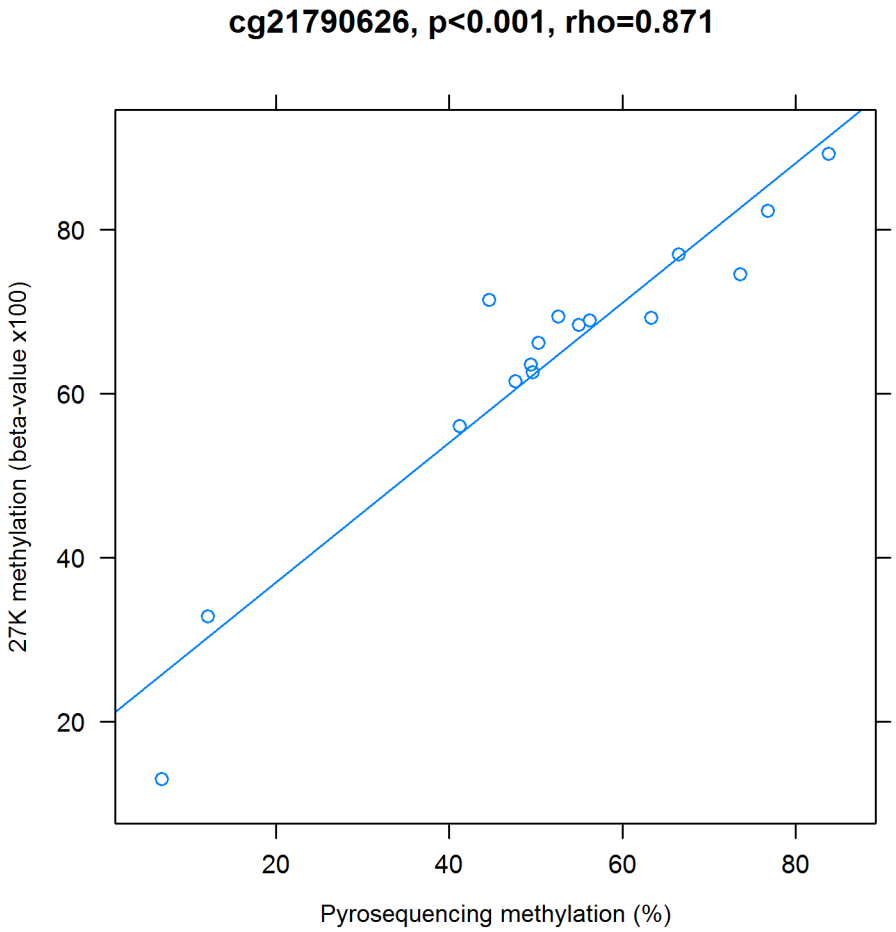



Figure 3.7. Correlation of cg21790626 methylation detected by bisulphite pyrosequencing and Illumina Infinium methylation array in HH cohort.

### 3.3.4 Pyrosequencing of differentially methylated candidates in an independent cohort of tumour and PBMC samples

To validate the 9 candidate loci and determine if bisulphite pyrosequencing is a suitable detection method that can be used in routine diagnostic testing in the clinic, 89 tumour DNA samples were received from Charité Universitätsmedizin, Berlin and 76 high grade serous ovarian cancer samples were collected at Hammersmith Hospital, London (termed the New HH cohort). In addition, 11 DNA samples from benign ovarian tissue, collected as part of the HH cohort but not analysed on the methylation array, were used as part of the validation cohort and 11 PBMC samples from patients

with ovarian cancer were also collected as part of the New HH cohort for purposes of validating the 9 candidate loci using bisulphite pyrosequencing. Together with the tumour samples from Charité, Berlin these form the validation cohort. Characteristics of the participants in this validation study are shown in Table 3.5.

Median methylation, SD and range of methylation detected using bisulphite pyrosequencing in each of the groups in the validation study are shown in Table 3.6. Box and whisker plots of the candidate cg methylation in these groups are also shown in Figure 3.8. Methylation values from pyrosequencing of samples in the candidate validation cohort are consistent with those seen in the discovery and verification cohorts; specifically the differential methylation of *ZNF154*, *HOXA9* and *MGAT1* in tumour DNA compared to benign and PBMC DNA.

**Table 3.5** Clinical characteristics of study participants used in the candidate validation study.

| | | **Charité** | **HH cohort** | **New HH cohort** | |
|---|---|---|---|---|---|
| | | Tumour (n=89) | Normal (n=11) | Tumour (n=76) | Tumour PBMC (n=11) |
| **Age** | Mean (range) | 59 (35-92) | 45 (29-61) | 63 (38-90) | 59 (36-77) |
| **Grade** | Borderline | 0 | | | 2 |
| | 1 | 0 | | | 1 |
| | 2 | 19 | | | 1 |
| | 3 | 70 | | 76 | 5 |
| | Unknown | 0 | | | 2 |
| **Stage** | 1A,1B,1C | 1 | | | 3 |
| | 2A,2B,2C | 1 | | | |
| | 3A | 1 | | | 1 |
| | 3B | 2 | | 9 | |
| | 3C | 63 | | 50 | 4 |
| | 4 | 21 | | 12 | 3 |
| | Unknown | 0 | | 5 | |
| **Histological type** | Serous | 89 | | 76 | 7 |
| | Mucinous | 0 | | | 2 |
| | Endometrioid | 0 | | | |
| | Clear cell | 0 | | | 1 |
| | Granulosa cell | 0 | | | 1 |

**Table 3.6.** Details of the nine differentially methylated candidate cg sites in the Charité cohort, New HH cohort, HH cohort and PBMCs using bisulphite pyrosequencing.

| Probe ID | Gene name | Charité tumour median methylation (%) | SD | New HH cohort median methylation (%) | SD | HH benign median methylation (%) | SD | PBMC median methylation normal (%) | SD |
|---|---|---|---|---|---|---|---|---|---|
| **cg05596756** | FAM113B | 77.07 (40.25-98.51) | 14.567 | 81 (24-96.54) | 15.104 | 92.21 (83.92-95.44) | 3.487 | 2.03 (0-5.41) | 1.261 |
| **cg07380416** | CD6 | 71.1 (39.87-94.35) | 13.773 | 77.81 (22.63-94) | 15.313 | 89.17 (79.94-92.45) | 4.344 | 0.54 (0-1.95) | 0.817 |
| **cg16927606** | TMEM149 | 73.12 (41.02-95.79) | 12.557 | 81.42 (24.72-94.17) | 15.057 | 85.05 (77.76-88.66) | 3.655 | 3.21 (1.66-4.84) | 1.042 |
| **cg20425130** | KLHL6 | 77.75 (29.02-90.29) | 11.352 | 82.83 (27.9-92.69) | 13.817 | 85.69 (73.92-88.72) | 4.583 | 2.285 (0-3.28) | 1.094 |
| **cg21790626** | ZNF154 | 41.89 (3.11-89.57) | 20.816 | 52.38 (0-86.89) | 24.717 | 5.14 (2.27-9.45) | 2.104 | 5.32 (1.99-12.75) | 3.092 |
| **cg21898046** | MGAT1 | 37.93 (6.76-70.11) | 15.801 | 40.28 (4.72-87.91) | 22.051 | 20.53 (11.42-37.16) | 7.764 | 5.23 (4.12-7.35) | 1.178 |
| **cg27009703** | HOXA9 | 40.82 (3.94-92.27) | 19.588 | 49.94 (2.78-88.52) | 24.598 | 11.2 (7.64-14.04) | 1.464 | 8.35 (4.66-12.05) | 2.014 |
| | | | | | | | | | |
| **cg09152089** | IL22RA1 | 20.43 (0-60.41) | 13.731 | 13.01 (0-64.65) | 13.814 | 8 (2.28-15.89) | 3.941 | 85.72 (58.13-92.13) | 9.914 |
| **cg22131172** | C13orf29 | 12.52 (1.55-46.31) | 11.236 | 8.93 (1.4-68.34) | 13.253 | 15.19 (7.47-30.6) | 7.673 | 95.55 (88.30-97.68) | 3.434 |

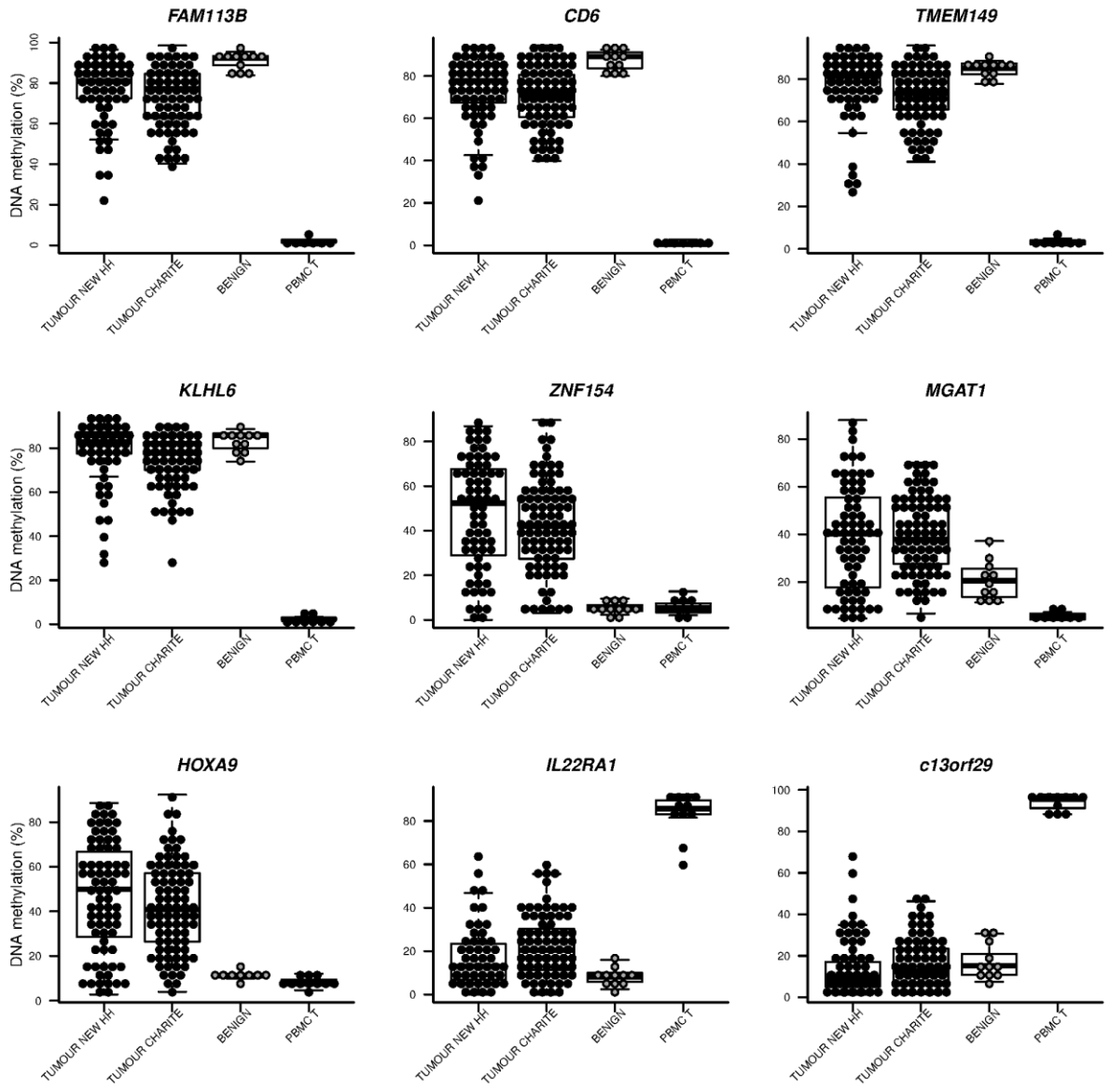



**Figure 3.8.** Box and whisker plots of the candidate cg site methylation in the validation cohort of HGSOC tumours and PBMCs from patients with ovarian tumours using bisulphite pyrosequencing.

### 3.3.5 Pyrosequencing of differentially methylated candidates in prospectively collected blood samples

Blood was drawn from 46 patients undergoing gynaecological surgery at Hammersmith Hospital into sodium heparin vacutainers under strict processing conditions (see Materials and Methods: Section 2.3.4.2). Samples were also collected from 11 healthy controls and processed in the same way. Both plasma and PBMC samples were extracted to assess the level of contamination of the plasma with PBMC DNA. The diagnosis, grade and stage of tumours were confirmed histologically at Imperial College Healthcare NHS Foundation Trust's Histopathology Department at Hammersmith Hospital. For the purposes of this study, tumours of non-ovarian origin (n=4), duplicate samples from the same patient (n=2) and any samples taken from patients who had previously or were undergoing

chemotherapy treatment (n=1) were excluded from the analysis. In total, 30 blood samples from patients with tumours including two borderline cases, 9 patients undergoing surgery for benign ovarian pathologies and 11 healthy controls were included. Characteristics of the participants in this validation study are shown in Table 3.7.

**Table 3.7.** Clinical characteristics of study participants used in the candidate validation study of prospectively collected blood samples from patients undergoing gynaecological surgery at Hammersmith Hospital and non-affected volunteers.

| | | **Biofluid cohort** | | |
|---|---|---|---|---|
| | | Tumour (n=30*) | Benign (n=9) | Normal (n=11) |
| **Age** | Mean (range) | 60 (34-81) | 50 (37-74) | 33 (24-41) |
| **Grade** | Borderline | 2 | | |
| | 1 | 1 | | |
| | 2 | 1 | | |
| | 3 | 23 | | |
| | Unknown | 3 | | |
| **Stage** | 1A,1B,1C | 3 | | |
| | 2A,2B,2C | | | |
| | 3A | | | |
| | 3B | 4 | | |
| | 3C | 11 | | |
| | 4 | 11 | | |
| | Unknown | 1 | | |
| **Histological type** | Serous | 21 | | |
| | Mucinous | 1 | | |
| | Endometrioid | | | |
| | Clear cell | | | |
| | Granulosa cell | 1 | | |
| | Mixed | 2 | | |
| | Carcinosarcoma | 4 | | |
| | Small cell | 1 | | |

*Tumours of non-ovarian origin were excluded from analysis

Methylation of the 9 candidate cg sites in plasma and PBMC fractions of the prospectively collected blood samples are shown in Figure 3.9. To validate the findings from the discovery cohort, only samples with a diagnosis of high grade serous ovarian cancer were used in the cancer arm of the analysis. ROC curves were generated to assess the performance of each marker independently. Of the nine candidate loci, five of them showed a significant difference, using the student's t test in methylation of cell-free circulating DNA between the HGSOC group and the non-tumour controls (Figure 3.9). ROC analysis of the 5 significant loci showed an AUC of 0.911 for *TMEM149*, AUC of 0.753 for *CD6*, AUC of 0.738 for *ZNF154*, AUC of 0.758 for *HOXA9* and 0.811 for *IL22RA1* (Figure 3.10) suggesting a promising accuracy of these biomarkers for HGSOC diagnosis.

As no marker predicted HGSOC with 100% accuracy, there is always a trade-off between sensitivity and specificity when defining the optimum cut-off for methylation value that distinguishes cases from controls and this becomes easier with an increased number of patients. With the samples from this study, the following sensitivity and specificity with associated methylation cut-off values can be achieved for each of the 5 significant biomarkers; *CD6* sensitivity = 67%, specificity=75% with a methylation cut-off at 13.15%, *TMEM149* sensitivity =79% and specificity = 100% with a methylation cut-off at 10.59%, *ZNF154* sensitivity = 64% and specificity = 67% with a methylation cut-off at 7.03%, *HOXA9* sensitivity = 83% and specificity =60% with a methylation cut-off at 4.22% and *IL22RA1* sensitivity = 73% and specificity = 84% with a methylation cut-off at 78.59%. Methylation values above all of the cut-offs predicts patients with EOC except for *IL22RA1* where methylation values below the cut-off predict EOC.

To improve on the diagnostic accuracy of the markers and produce an individual probability of having HGSOC, a combination of the markers could be performed in a logistic regression model to yield an even higher AUC. Unfortunately, due to the small numbers of samples it is not possible to establish a model of this kind on the current data-set.

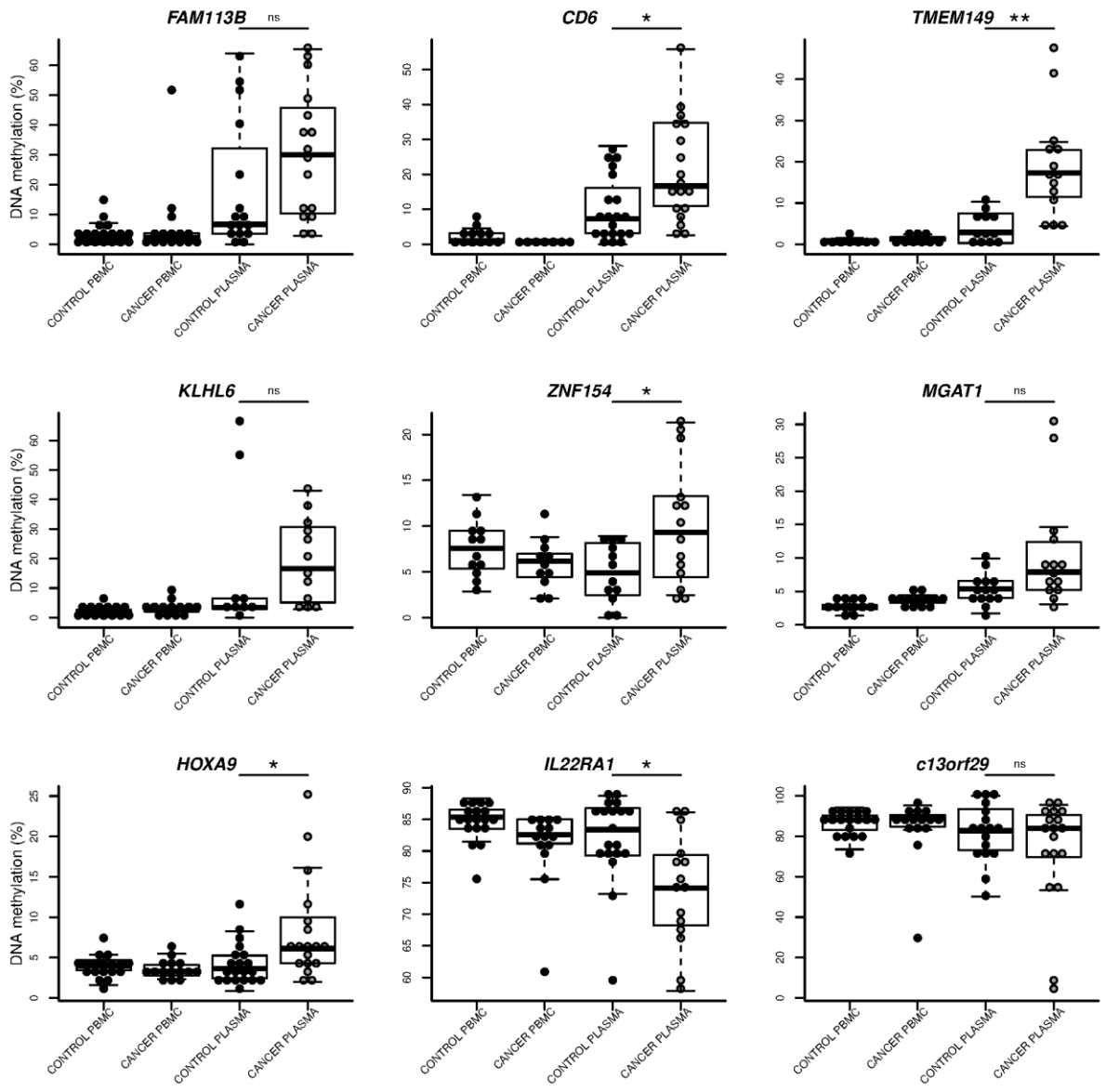



**Figure 3.9.** Box and whisker plots of the differentially methylated candidate cg sites in prospectively collected blood samples from patients with HGSOC (cancer) and benign pathologies or non-affected volunteers (controls). *p<0.05, **p<0.01, ns- not significant.

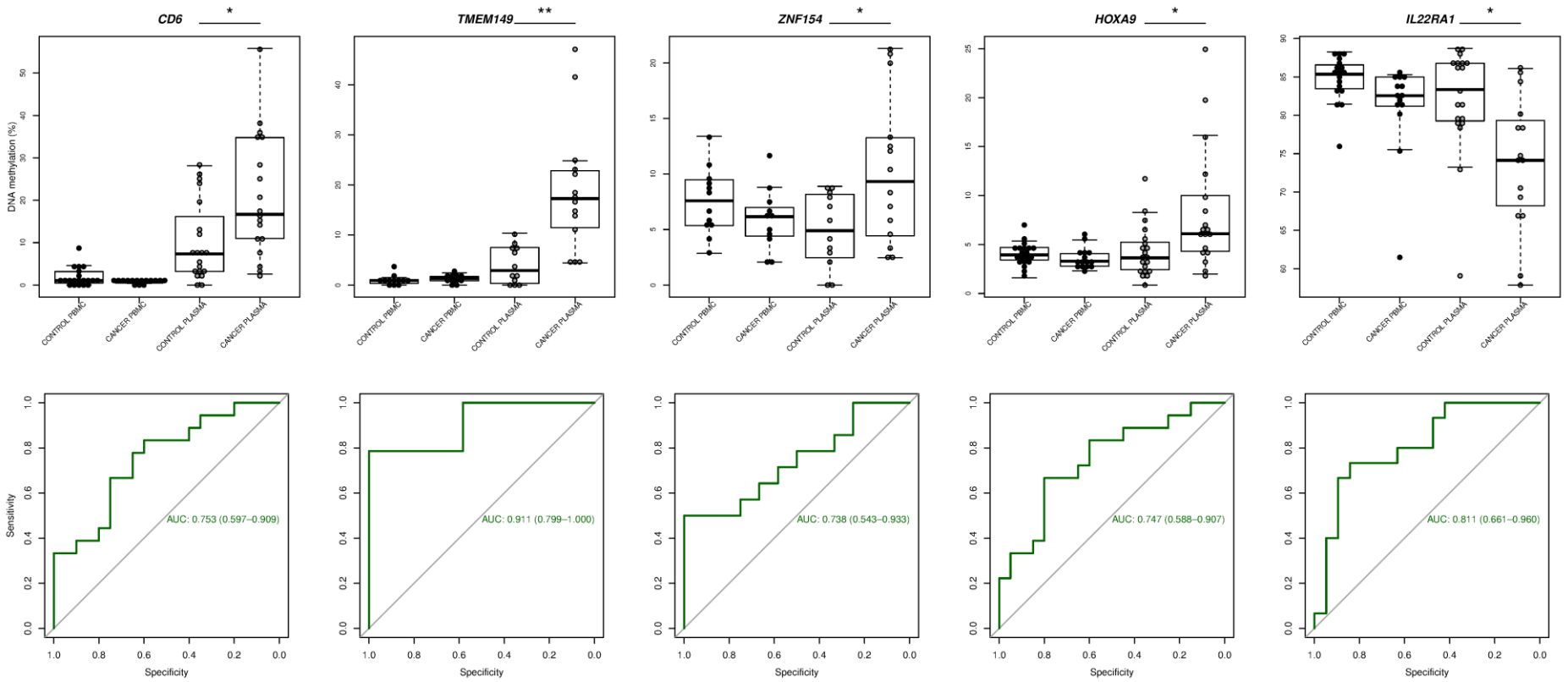



**Figure 3.10.** Box and whisker plots of the significantly differentially methylated candidate cg sites in prospectively collected blood samples from patients with HGSOC (cancer) and benign pathologies or non-affected volunteers (controls) and their associated ROC curves. *p<0.05, **p<0.01.

### 3.3.6 Exploring the relationship between the 9 candidate CpG sites

Given the stringent nature of the candidate selection of hyper and hypo-methylated CpG sites it would be expected that methylation of the 7 serous ovarian carcinoma hypermethylated genes would be positively correlated with each other and the 2 PBMC hypermethylated genes would be positively correlated with each other. Spearman correlation of the 9 candidate CpG sites in the discovery, verification and validation cohorts of DNA from tumour tissue was assessed. Interestingly, the same pattern of correlation was seen for all 4 tumour cohorts: TCGA, HH cohort, Charité and the New HH cohort, Figure 3.11. A positive correlation was seen for methylation of *FAM113B*, *CD6*, *TMEM149* and *KLHL6* however the two CpG-island associated loci, *ZNF154* and *HOXA9* as well as *MGAT1* showed no correlation with any of the other loci.

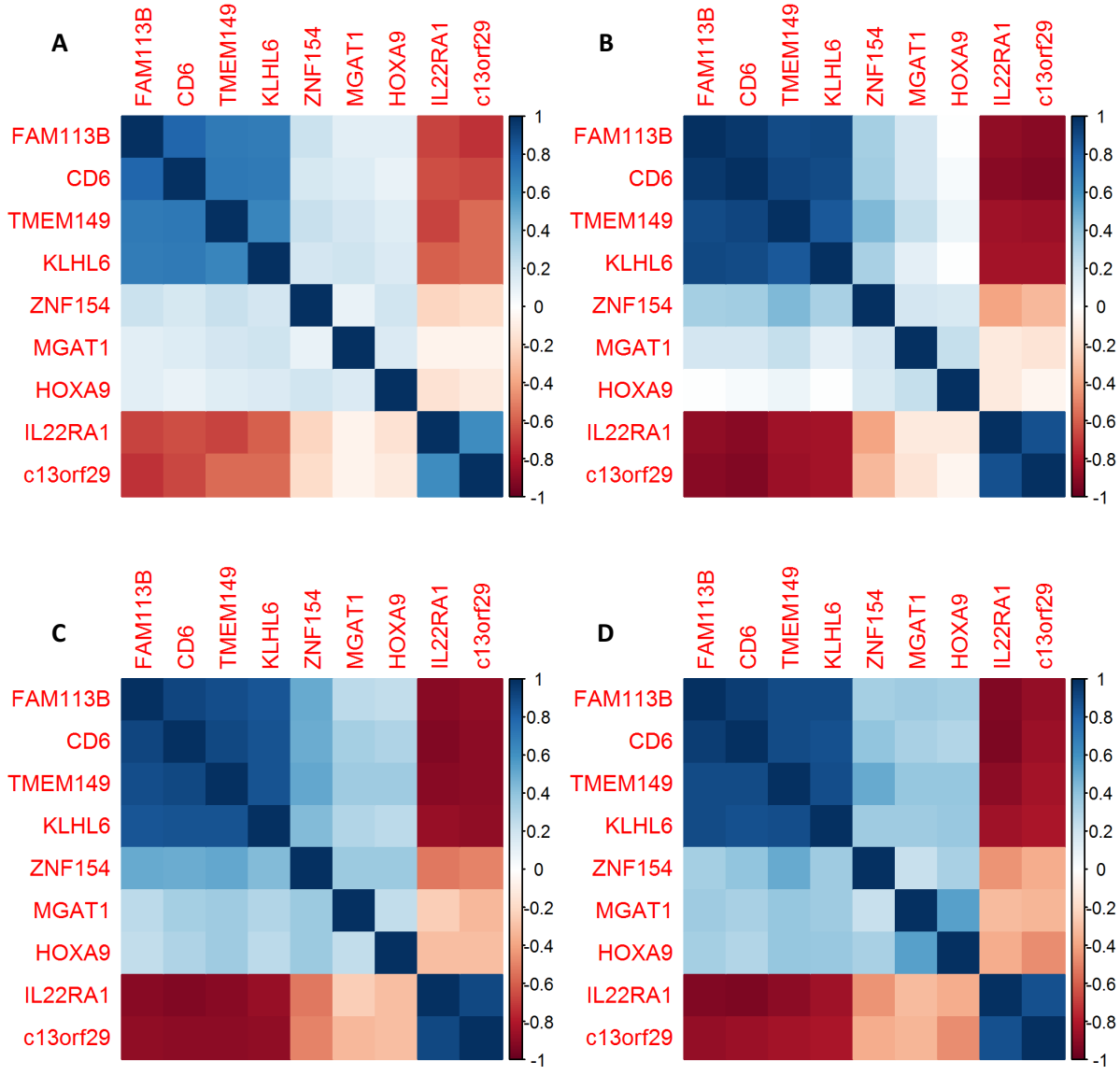



**Figure 3.11.** Correlation of 9 candidate hits in **A**, TCGA (discovery cohort), **B**, HH cohort (verification cohort), **C**, Charité (validation cohort), **D**, New HH cohort (validation cohort).

### 3.3.7 Correlation of methylation and expression of CpG Island associated *ZNF154* and *HOXA9*

Promoter methylation of multiple consecutive CpGs is recognised as an important mechanism by which genes may be transcriptionally silenced in both physiological and pathological conditions including the progression of many common human tumours. RT-qPCR was used to evaluate the association between methylation of the CpG-island associated hypermethylated genes, *HOXA9* and *ZNF154* and gene transcription. Relative mRNA levels of ZNF154 was markedly lower in HGSOC tumour samples compared to benign ovarian samples (Figure 3.12) and a negative correlation exists between methylation of cg21790626 and transcription of ZNF154 (rho=-0.723) $p<0.001$ suggesting

methylation associated transcriptional regulation of this gene has a role in the development and/or progression of high grade serous ovarian cancer. The same trend was not seen for *HOXA9*, rho=0.095, p=0.6179, shown in Figure 3.13 with similar levels of mRNA in both the HGS tumours and the benign ovarian samples in this study.

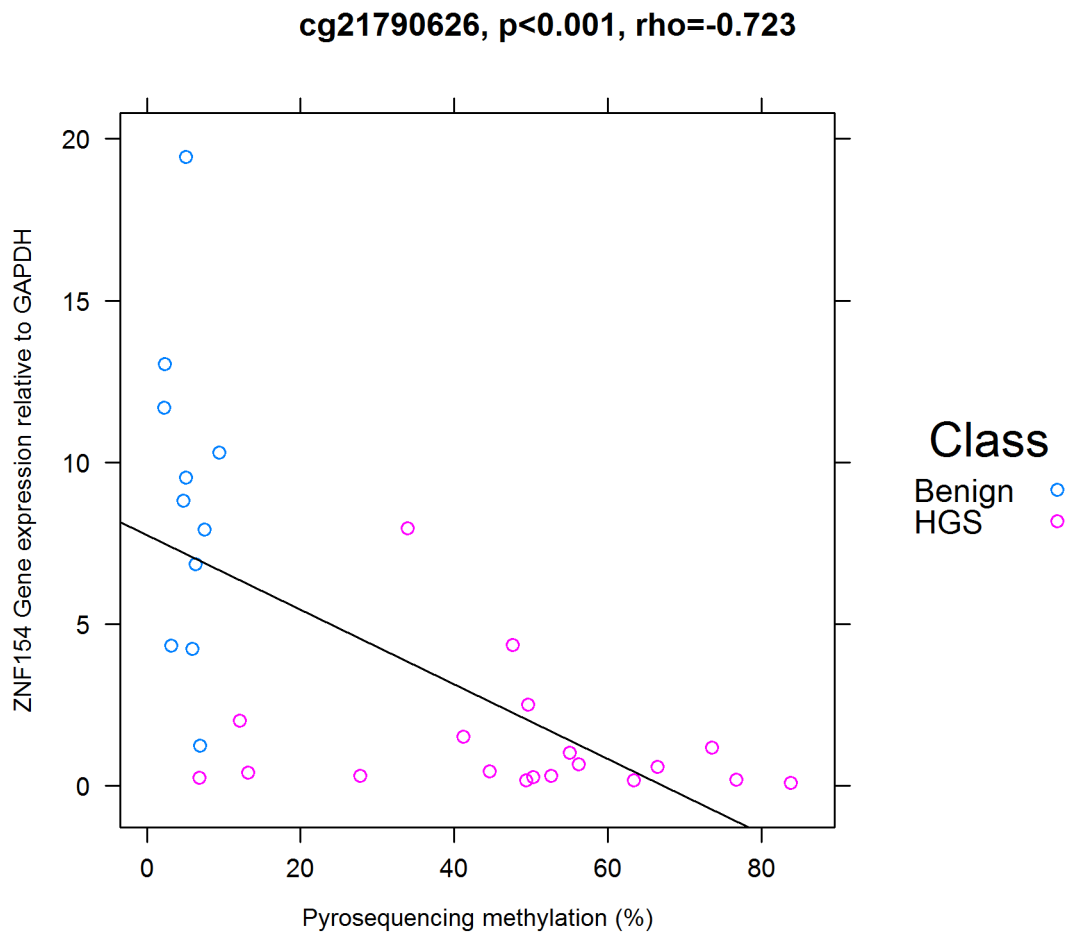



**Figure 3.12.** Scatterplot showing the correlation of DNA methylation of cg21790626 in the *ZNF154* gene determined by bisulphite pyrosequencing and expression of *ZNF154* relative to *GAPDH* determined by RT-qPCR. (Benign samples are shown in blue and high grade serous ovarian carcinoma samples are shown in pink).

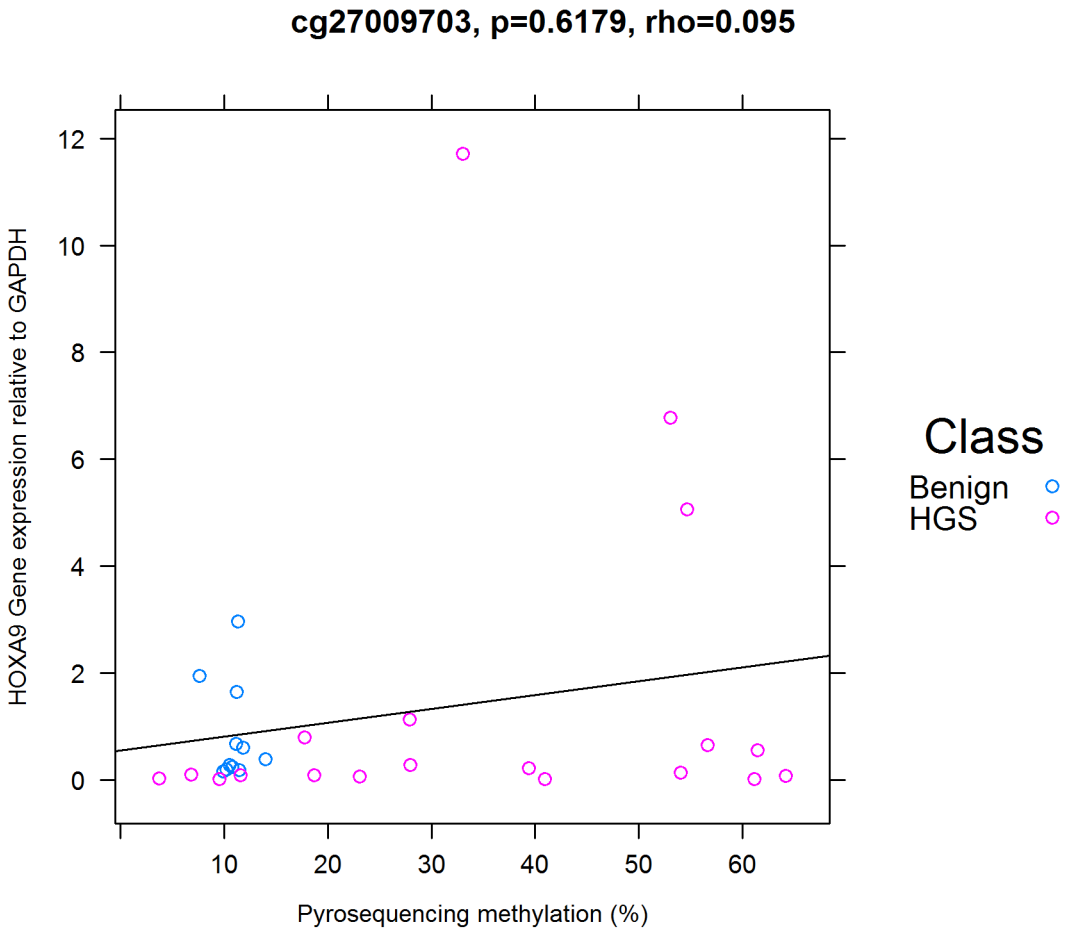



**Figure 3.13.** Scatterplot showing the correlation of DNA methylation of cg27009703 in the *HOXA9* gene determined by bisulphite pyrosequencing and expression of *HOXA9* relative to *GAPDH* determined by RT-qPCR. (Benign samples are shown in blue and high grade serous ovarian carcinoma samples are shown in pink).

### 3.3.8 Methylation of *ZNF154* (cg21790626) as a pan-cancer biomarker

Of the nine differentially methylated candidate cg sites identified in the discovery phase of the study, two cg sites; *ZNF154* (cg21790626) and *HOXA9* (cg27009703) showed differential methylation between tumour and non-tumour DNA, including PBMC DNA and DNA from benign tissue, suggesting that methylation at these loci might be tumour-specific. Data from the TCGA online data portal was further evaluated to assess whether this difference in methylation of *ZNF154* between tumour and 'normal' tissue was also observed in other cancer datasets. Box and whisker plots of the 16 other methylation array data sets of various cancer types is shown in Figure 3.14.

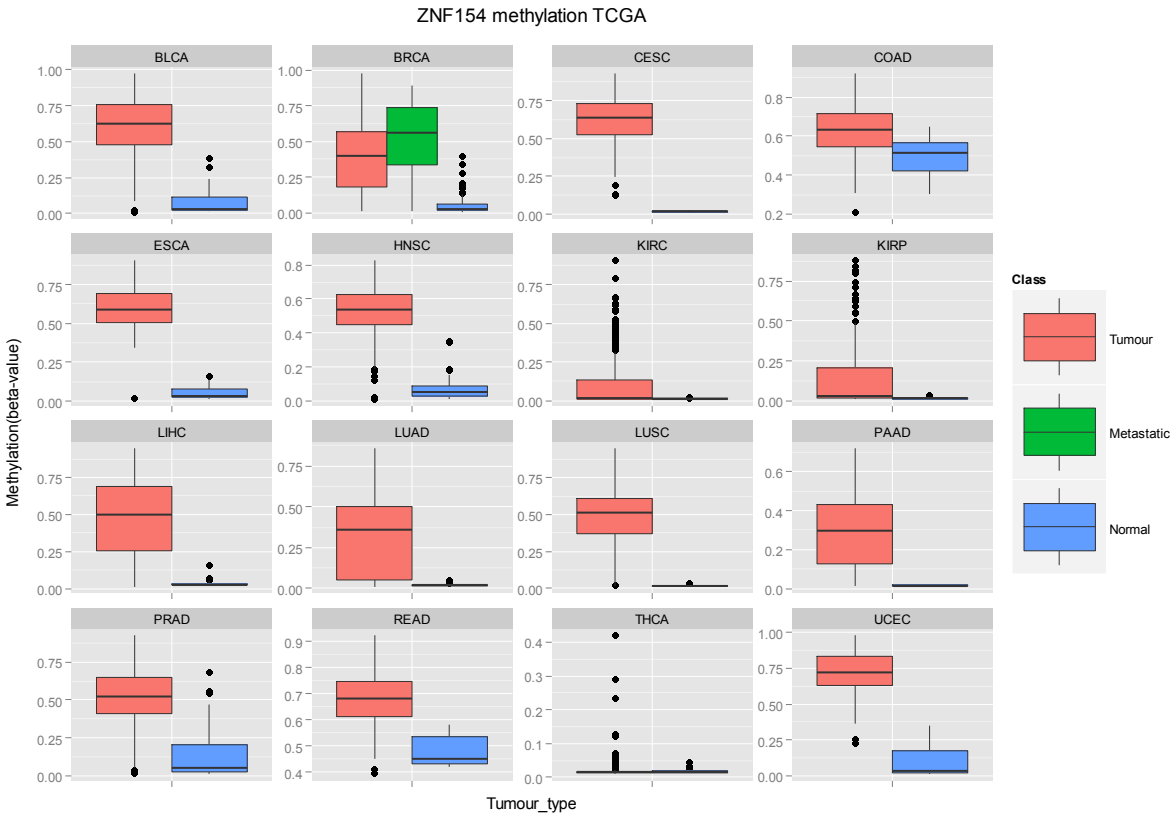



**Figure 3.14.** Box and whisker plots of methylation of *ZNF154* (cg21790626) across all tissue types from TCGA. BLCA (Bladder Urothelial Carcinoma), BRCA (Breast invasive carcinoma), CESC (Cervical squamous cell carcinoma and endocervical adenocarcinoma), COAD (Colon adenocarcinoma),ESCA (Esophageal carcinoma), HNSC (Head and Neck squamous cell carcinoma), KIRC (Kidney renal cell carcinoma),KIRP (Kidney renal papillary cell carcinoma), LIHC (Liver hepatocellular carcinoma), LUAD (Lung adenocarcinoma),LUSC (Lung squamous cell carcinoma), PAAD (Pancreatic adenocarcinoma),PRAD (Prostate adenocarcinoma),READ (Rectum adenocarcinoma),THCA (Thyroid carcinoma), UCEC (Uterine Corpus Endometrial Carcinoma

The differential methylation between tumour and normal tissue seen in the ovarian serous adenocarcinoma cohort of TCGA is also observed in nearly all the other tumour tissue types, with the exceptions of tumours of the kidney; Kidney renal cell carcinoma and Kidney renal papillary cell carcinoma as well as thyroid carcinoma. Combined with the significant difference between cell free circulating DNA methylation of *ZNF154* between patients with high grade serous ovarian cancer and controls shown in this chapter, this suggests that DNA methylation in cell-free circulating DNA methylation may be used as a pan-cancer diagnostic biomarker in blood.

### 3.3.9 Methylation of *ZNF154* (cg21790626) and the association with survival

An abstract from the proceedings of the Annual AACR meeting in 2012 from Okamoto *et al*. showed that increased methylation of *ZNF154* is associated with poor overall survival for advanced stage ovarian serous carcinomas (Okamoto *et al.*, 2012) and this suggests it may also function as a prognostic biomarker.

To investigate whether methylation of *ZNF154* is associated with overall and progression-free survival, data from TCGA, the HH, Charité and New HH cohorts were all analysed in an attempt to verify the findings of Okamoto *et al*. Univariate logrank statistical tests and multivariate cox proportional hazard models adjusting for age, stage, grade and volume of residual disease were performed. Progression free survival was defined as the time between treatment initiation and tumour progression or death from any cause, with censoring of patients who were lost to follow-up.

#### 3.3.9.1 Survival analysis in HH cohort

Overall survival (OS) and progression free survival (PFS) data was available on 72 patients in the HH cohort. There was no significant difference in OS dependent on methylation dichotomised at the median in all case subjects (n=72), log-rank p-value= 0.153. The Kaplan-Meier survival curve is shown in Figure 3.15A. There was a significant difference in PFS between patients with *ZNF154* methylation above and below the median, Figure 3.15B. Progression free median survival was 346 days (95% CI 309-469) in the group with methylation above the median and 487 days (95% CI 411-712) in the group with methylation below the median. However this significance was lost in multivariate cox proportional hazard analysis adjusting for age, stage, grade, chip and residual disease, p=0.0738, hazard ratio (HR) 1.63 (0.95-2.79). There was a significant difference in PFS for grade (p=0.024) and residual disease (p=0.001).

As Kaplan Meier survival curves for both OS and PFS suggest a difference between the groups dichotomised at the median, methylation values were split into tertiles to identify if there was a significant difference between methylation above the top tertile (beta-value >0.728) and below the bottom tertile (beta-value <0.62). There was a significant difference in OS between the groups dependent on *ZNF154* methylation, log-rank p-value=0.0365, Figure 3.15C. Overall survival in the group with higher methylation was 785 days (95% CI 685-NA) and 1594 days (95% CI 1181-NA) in the group with lower methylation. However, once again this significance was lost in multivariate cox proportional hazard analysis adjusting for age, stage, grade, chip and residual disease, p=0.0869, HR 2.21 (0.89-5.47). There was a significant difference in OS for residual disease (p=0.024). There was also a significant difference in PFS between the groups dependent on *ZNF154* methylation, log-rank p=value=0.009, Figure 3.15D. Progression free median survival was 549 days (95% CI 433-1597) in

the lower methylation group and 359 days (95% CI 309-644) in the higher methylation group. This difference remained significant in multivariate cox proportional hazard analysis adjusting for age, stage, grade, residual disease and chip, p=0.015, HR 2.42 (1.18-4.94). There was a significant difference in PFS for grade (p=0.049) and residual disease (p=0.004).

This data suggests that there may be an association between methylation of *ZNF154* and survival, with a higher methylation associated with a shorter time to progression. Analysis of other data sets however, namely TCGA, Charité and the New HH cohort did not validate these findings, data shown in Appendix 1. An association between methylation of *ZNF154* and overall or progression free survival can therefore not be concluded from this analysis and it is unlikely that methylation of *ZNF154* may be useful as a prognostic marker.

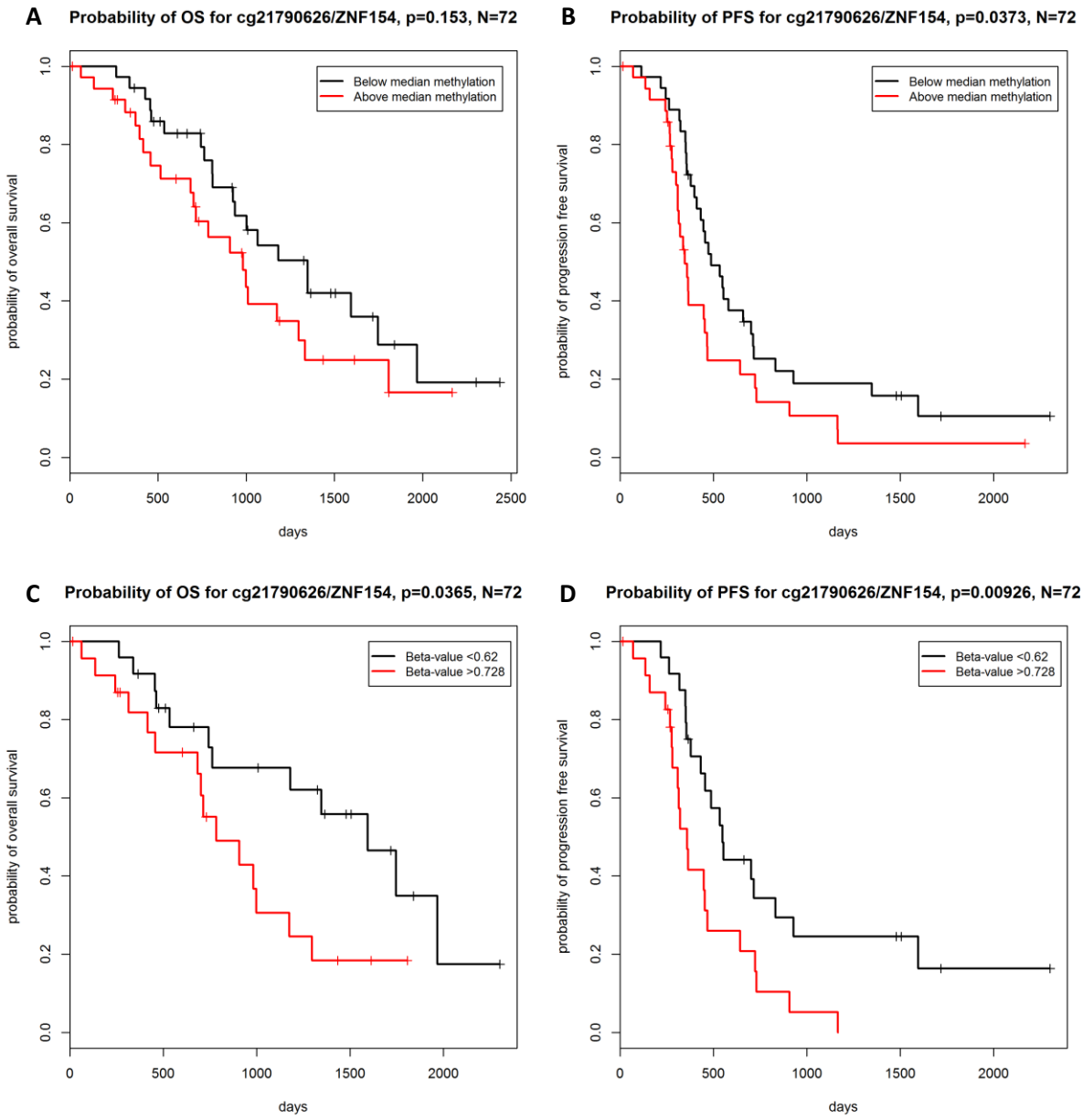



**Figure 3.15.** Kaplan-Meier survival plots for OS and PFS in HH dataset of all primary tumours. **A**, OS and **B**, PFS dependent on *ZNF154* methylation dichotomised at the median. **C**, OS and **D**, PFS dependent on *ZNF154* methylation dichotomised at the upper and lower tertiles.

## 3.4 Discussion

### 3.4.1 Differentially methylated loci as diagnostic biomarkers of serous ovarian cancer

Analysis of methylation array datasets of serous ovarian tumours from TCGA and PBMCs from the Mayo clinic has identified nine candidate loci that are differentially methylated between the two cohorts. Selection criteria for differential methylation was defined by median methylation beta-value >0.8 in one cohort and median methylation beta-value <0.1 in the other. The pattern of methylation between serous ovarian cancer and PBMCS was subsequently verified in other Infinium methylation array datasets. Only three out of the nine candidate loci met the same stringent criteria in the verification cohort but the difference in median methylation between tumour and PBMC was sufficiently high enough to progress all nine loci to the validation stage of pyrosequencing assay design. The majority of DNA present in blood arises from PBMCs, therefore it was prudent to compare methylation of tumour DNA to PBMC DNA rather than the conventional tumour compared to normal or benign ovarian tissue.

Seven of the nine candidate loci were hypermethylated (median beta-value >0.8) and two were hypomethylated (median beta-value <0.1) in serous ovarian tumours. Of the seven, four showed consistent methylation in tumour and normal ovarian tissue suggesting the methylation of these loci may be specific to cells of the ovary, independent of the tumour, such as epithelial cells. The other three loci: *ZNF154*, *MGAT1* and *HOXA9* had varying degrees of methylation in the tumours, represented by the larger standard deviation for these loci and showed a difference in methylation between the serous tumours and the normal ovarian tissue suggesting the methylation of these loci could be specific to the tumour in the ovary. The varying degree of methylation seen for these loci in the tumour samples could therefore indicate a heterogeneous mix of cell type within each tumour sample.

### 3.4.2 Differentially methylated loci between serous tumours and benign ovarian tissue and their association with gene expression.

Two of the candidate loci showed differential median methylation between serous tumours and normal ovarian tissue. These two CpG sites; cg21790626 in the *ZNF154* gene and cg27009703 in the *HOXA9* gene are also the only two sites to be located within CpG islands. Methylation of cg21790626 in the *ZNF154* gene inversely correlated with mRNA expression and is therefore associated with transcriptional silencing of this gene suggesting a potential role of DNA methylation in the functional regulation of *ZNF154* in the development and evolution of serous ovarian cancer. Although differential DNA methylation of the *HOXA9* gene was shown between serous ovarian tumours and

benign ovarian and fallopian tube samples, there was no correlation between mRNA expression and methylation in the HGS tumours and benign ovarian samples in this study.

Methylation of the Homeobox A9 (*HOXA9*) gene has previously been reported to occur in HGSOC with one study reporting methylation in >95% of cases and the ability to discriminate HGSOC from benign ovarian surface epithelium with 95% sensitivity and 92% specificity (Montavon *et al.*, 2012).

HOX genes encode transcription factors involved in cell differentiation during normal embryonic development (Du and Taylor, 2004) with HOXA9 regulating serous differentiation of the müllerian ducts to fallopian tubes (Taylor *et al.*, 1997). As expression of HOXA9 specifies regional identity in the reproductive tract and is therefore expressed in areas destined to become the fallopian tube this may point to why the benign samples in this study showed low levels of mRNA expression, as they were all of ovarian origin. This is further supported in another study by Kelly *et al.,* as normal ovarian tissue in this study also showed very little expression of HOXA9 amongst other HOXA genes (Kelly *et al.*, 2016). Hypermethylation and low levels of mRNA expression of this gene in ovarian cancer may reflect a shift towards epithelial cell dedifferentiation with a loss in transcriptional plasticity which maintains the cells in an uncommitted state susceptible to malignant transformation (Widschwendter *et al.*, 2009).

### 3.4.3 Bisulphite pyrosequencing is an appropriate analytical technique for methylation detection in solid tissue

Bisulphite pyrosequencing was chosen as a validation technique as it is an affordable and quantitative method suitable for routine use. It overcomes some weaknesses of previous and extensively used methods, such as MSP, due to its easy standardisation and lower false positive rate (Mikeska *et al.*, 2012). Moreover, it is a robust method able to detect multiple CpGs not only in tumour samples but also in minimally and non-invasive bio fluid samples with potential use in clinical settings (Diaz-Lagares *et al.*, 2016).

Methylation of various normal and cancerous cell-lines was assessed in the optimisation of the pyrosequencing assays. These results further confirmed the findings in primary tumour and benign tissue seen in the discovery and verification cohorts. All seven of the hypermethylated candidate loci in serous ovarian cancer were hypermethylated in the ovarian serous carcinoma cell-lines and hypomethylated in the control female DNA. In addition, normal fallopian tube cell-lines showed a decrease in methylation of *ZNF154* and *HOXA9* but not to the same degree as in the normal primary tissue samples. Methylation of *MGAT1* in these cell-lines was also considerably lower supporting the hypothesis that these three loci are tumour specific. Pyrosequencing of primary tumour, benign and PBMC samples replicated the pattern seen in the discovery and verification cohorts with *MGAT1*,

*ZNF154* and *HOXA9* showing the most dispersion of values around the median for the HGS tumours and most notably *ZNF154* hypomethylation in the benign samples from the HH cohort.

### 3.4.4 *ZNF154* as a pan-cancer diagnostic and prognostic biomarker

The *ZNF154* gene encodes a protein that is a member of the zinc finger krüppel family of transcriptional regulators. There is evidence for the implication of several zinc-finger transcription factors in cancer and they are thought to regulate downstream genes involved in proliferation, apoptosis, migration and invasion (Jen and Wang, 2016). This pattern of differential methylation makes this an exciting candidate as a diagnostic biomarker in blood and further investigation of the function of this gene and the impact of DNA methylation is therefore warranted as it may be a target for epigenetic therapy.

This is not the first study to identify differential methylation of this gene in cancer. In a study by Sanchez-Vega *et al*., *ZNF154* hypermethylation was identified in a number of solid epithelial tumours and cancer cell-lines. In keeping with our findings they also showed consistent ubiquitous hypermethylation of this gene across multiple cancer types from TCGA (15/16 assessed) and stressed its potential as a diagnostic biomarker (Sanchez-Vega *et al.*, 2013). Most recently, this same group, having identified the need for further exploration, have published a study using bisulphite amplicon sequencing to quantitate methylation levels of all the CpGs within the amplicon to reveal patterns of DNA methylation useful for distinguishing tumour from normal samples (Margolin *et al.*, 2016). They use computational simulation to demonstrate the utility of these features in correctly classifying samples as tumour or normal at a number of abundance levels and try to mimic the fractions at which ctDNA is recovered from blood further reinforcing the potential of this site as a biomarker for ctDNA. To date, however, no studies have been published that have assessed the performance of this biomarker in blood in vivo.

Other publications have suggested that *ZNF154* methylation may be associated with poor overall survival (Okamoto *et al.*, 2012) as well as detection of tumour recurrence (Reinert *et al.*, 2012). Analysis of TCGA survival and methylation data revealed no association between cg21790626 methylation and overall or progression free survival in HGSOC. Data from the HH cohort, a study performed within our research group, suggested that methylation of *ZNF154* may predict overall survival as well as disease progression but when the model was adjusted for confounding factors such as age, stage, grade and residual disease only the association between methylation at the top and bottom tertiles with progression free survival remained significant. These findings were not validated in other cohorts, namely the Charité and New HH cohort, another cohort performed in-house with survival annotation from the case notes and therefore it appears that after extensive

analysis in four independent cohorts from a number of different centres across the world that *ZNF154* methylation is not associated with survival and has little use as a prognostic biomarker however future studies could investigate the association between expression of the gene and survival. There was also no association between *HOXA9* methylation and overall or progression free survival in any of the above named cohorts (data not shown).

### 3.4.5 Biological function of candidate diagnostic biomarkers

Spearman correlation of the nine candidate DMCpG sites in the discovery, verification and validation cohorts of HGS tumours revealed a remarkably similar pattern between the nine candidate loci. As the cut-off of methylation in the selection of the candidate loci was very stringent it was predicted that the seven cancer-related hypermethylated loci (beta-value >0.8 or percentage methylation >80%) would positively correlate with each other and the PBMC-related hypermethyated loci (beta-value <0.1 or percentage methylation <10%) would positively correlate with each other and each of these groups would negatively correlate with the other. This was true for the PBMC-related hypermethylated loci but of the seven cancer-related hypermethylated loci; *FAM113B*, *CD6*, *TMEM149* and *KLHL6* all positively correlate with each other whereas the two CpG island-associated hypermethylated loci, *ZNF154* and *HOXA9* as well as *MGAT1* do not correlate with any of the other loci. This can be explained by the spread of methylation values for these three cg sites in all three cohorts. There appears to be more variability in the methylation of these genes in tumour tissue compared to the other genes and no clear correlation between methylation of these genes between different tumours which may link to the heterogenous nature of individual ovarian tumours. This combined with the differential methylation between tumours and normal tissue may suggest that these genes are tumour-specific whereas the other 4 higher correlated genes may be associated with tumours or a cell type associated with the tumour but not specific to them such as for example epithelial cells.

CD6 is a plasma membrane protein associated with T cell activation and differentiation. A study looking at cell-type specific DNA methylation signatures showed that *CD6* was consistently hypomethylated in various haematopoietic cell types and beta-values of the cg site cg07380416 correlated with the proportion of counted epithelial cells (Eipel *et al.*, 2016). Combining this with our data for the same cg site it suggests that this protein is transcriptionally regulated by methylation and expressed on T cells and other haematopoietic cell types but not in epithelial/tumour cells. Methylation of this cg site represents the presence of epithelial cells and it's methylation in plasma DNA reflects the presence of DNA from epithelial cell tumours such as HGSOC.

TMEM149 is a transmembrane protein with cell-cell adhesion function. A study by Andreopoulos and Anastassiou investigated lymphocyte-specific gene expression and methylation signatures across multiple cancer types. In all cancer types, including ovarian, they found a signature of co-expression of a gene set R and methylated sites M, which positively correlated (M+) or negatively correlated (M-) with the expression of hsa-miR-142 (Andreopoulos and Anastassiou, 2012). Interestingly, 5 of our candidate biomarkers *CD6*, *KLHL6*, *TMEM149*, *FAM113B* (that also positively correlate with each other) as well as *IL22RA1* all appear in M for both ovarian cancer and GBM. Furthermore *CD6*, *KLHL6*, *TMEM149* and *FAM113B* are M- i.e. negatively correlated with miR-142 and are also hypermethylated in ovarian cancer in our study and *IL22RA1* which is M+ or positively correlated with miR-142 is hypomethylated in ovarian cancer in our study. They highlight that recent work has shown miR-142 to be expressed in haematopoietic and leukocytic cells, thus supporting the association of this miRNA with the methylation pattern of these transmembrane proteins seen in ovarian cancer and lymphocytes in our study and suggest that this specific signature may represent haematopoietic-specific processes that may trigger a loss of cell-cell adhesion in ovarian cancer (Andreopoulos and Anastassiou, 2012). *CD6*, *KLHL6*, *FAM113B* and *TMEM149* are hypomethylated in leukocytes which is associated with an increase in expression in these cells and they show that methylation is also negatively correlated with expression of miR-142 which has also been shown to be expressed in leukocytes. *IL22RA1* is hypermethylated in leukocytes which is associated with a decrease in expression in these cells and they show that methylation of *IL22RA1* is positively correlated with expression of miR-142 and IL22RA1 is a membrane protein, interleukin receptor exclusively expressed on epithelial and tissue cells and not on immune cells (Lim and Savan, 2014). Therefore, this supports the hypothesis that these genes are transcriptionally regulated by methylation and that the differential methylation seen in our study represents differential gene expression between tumours and PBMCs and methylation of these genes in plasma reflects the presence of DNA from epithelial cells or epithelial cell tumours such as HGSOC.

### 3.4.6 Bisulphite pyrosequencing of ctDNA methylation in plasma can diagnose high grade serous ovarian cancer

A final validation step was conducted using minimally invasive prospectively collected double-spun plasma and PBMC samples. Evaluation of locus-specific methylation in plasma showed that five of the nine candidate loci were significantly different between patients with high grade serous ovarian cancer and benign/healthy controls with one marker, *TMEM149* achieving an AUC of 0.901 and sensitivity and specificity of 79% and 100% respectively. However sample sizes were limited and there were varying success rates for pyrosequencing assays in plasma possibly due to the low amount of DNA present within the sample. The pyrosequencing assay with the highest success rate

in the double-spun plasma samples was *HOXA9* (95% of affected and 100% of unaffected individual's samples gave a signal distinguishable from background) suggesting this assay has a greater analytical sensitivity. The assay with the lowest success rate was *KLHL6* with 63% of affected and only 45% of unaffected individual's samples achieving the appropriate signal distinguishable from background suggesting a poorer analytical sensitivity for this assay.

This may suggest that some bisulphite pyrosequencing assays are not sensitive enough to detect methylation from the small amount of DNA that may be present in the sample, which is necessary for a diagnostic biomarker. To address this issue, alternative, more sensitive techniques that could enrich the tumour DNA would need to be explored. One such method is that developed by Agena Bioscience for quantification of foetal DNA in maternal blood by use of methylation-based discrimination analysis. The assay uses methylation-sensitive restriction enzymes, HhaI and HpaI to digest non-methylated sequences (maternal fraction) of the DNA sample. The undigested foetal DNA fraction is then co-amplified in the presence of a synthetic oligonucleotide for competitive PCR and quantified by single-base extension and MALDI-TOF MS analysis (Nygren *et al.*, 2010).

During this study, a collaboration was set-up between our research group at Imperial College London, Agena Bioscience and the Institute of Cancer Research (ICR) to use this method to target tumour DNA in a background of PBMC DNA in plasma samples. The method would utilise the differentially methylated loci, such as the ones identified in this study, and digest the non-methylated PBMC fraction and the remaining fraction containing tumour DNA, amplified and sequenced. In addition, analysis is not restricted to one locus per assay; therefore a number of candidate loci can be assessed in parallel, further enhancing the performance of the assay. Markers shown in this study that are epithelial cell specific and tumour specific could be determined in parallel and if successful, inclusion of loci in the *ZNF154* gene may result in a method for pan-cancer diagnosis. Due to the increased sensitivity of the method, approximately 50 copies of target DNA per ml of plasma, it may be able to pick up relatively small amounts of DNA in the plasma and therefore be useful as an early detection marker, but validation is still in the very early stages.

One of the limitations of this study was the use of the Infinium 27K methylation array in the discovery cohort from TCGA. This study looked at 23,679 probes in ovarian cystadenocarcinoma samples whereas the Mayo clinic study in PBMCs assessed 484,776 probes. Analysis of data from an Infinium 450K or EPIC array of ovarian cystadenocarcinoma samples would provide more differentially methylated candidate loci. To this end, as part of the collaboration with Agena Bioscience, a 450K infinium methylation array dataset of 82 primary ovarian cystadenocarcinoma samples and 9 normal samples from another collaborating group was assessed in comparison to the

Mayo clinic 450K Infinium methylation array to identify differentially methylated cg sites. In addition, methylation data from breast cancer samples, normal breast samples and whole blood epigenome analysis from Dr James Flanagan's group was also assessed in an attempt to find loci that are differentially methylated in both ovarian and breast cancer samples compared to PBMC and whole blood samples.

Thirteen loci had a beta-value of >0.8 in both cancer cohorts and <0.1 in the PBMC and whole blood cohorts this included two cg sites within the *CD6* gene including the cg site identified in the discovery cohort; cg07380416, further supporting the value of this methylation marker in differentiating both ovarian and breast cancer from blood. This data has not been included in this thesis due to confidentiality and intellectual property reasons but currently validation of the candidate sites is being assessed by Agena Bioscience through identification of restriction enzyme digestion sites. Validation of these sites will take place in solid tumour and benign tissue from the HH cohort and PBMC samples from the New HH cohort. Retrospective single spun plasma samples and prospective double spun plasma samples will be collected using the double spin protocol to minimise contamination from PBMCs and will subsequently be assessed by Agena Bioscience. Comparisons will be drawn between single and double spun plasma samples but there is a need to standardise blood sampling within institutions to increase the number of samples available for analysis by various methods to help move the field towards routine clinical use.

### 3.5 Summary

In summary, this work has identified and independently validated differentially methylated biomarkers between HGSOC and PBMCs for diagnosis of high grade serous ovarian cancer in minimally invasive samples. It is suspected that subsets of these methylation markers are cell-type specific and represent transcriptional regulation in epithelial, tumour and leukocyte cells that could, when used in parallel, add discriminatory power to diagnosis in blood plasma. Initial results of double spun plasma samples showed 5 out of 9 candidate loci to be significantly differentially methylated between plasma from HGSOC patients and benign/healthy controls. Prospective double spun plasma sample collection is ongoing to further validate the diagnostic accuracy of these markers. Collaboration with Agena Bioscience will validate another method that will enrich for tumour DNA in plasma based on candidate loci from 450K Infinium methylation array analysis of ovarian, breast, PBMC and whole blood datasets.

# CHAPTER 4. $^{1}$H NMR-based metabonomic profiling of plasma and urine to identify metabolites associated with ovarian cancer

## 4.1 Introduction

This chapter focuses on the use of proton nuclear magnetic resonance ($^1$H NMR) to identify differences in the global metabolic profile of patients with epithelial ovarian cancer (EOC) compared to patients with benign gynaecological disease and healthy controls. NMR spectra from plasma and urine were acquired using a 600 MHz Bruker Avance II NMR spectrometer (Bruker Biospin, Karlsrhue, Germany) and lipoprotein subfractions quantified using the *lipo*FIT® test: a software based analysis of NMR spectra (numaresHEALTH, Regensburg, Germany).

Two different NMR pulse sequences were conducted in the plasma samples, the 1D nuclear Overhauser enhancement spectroscopy (NOESY) presat for water suppression and the Carr-Purcell-Meiboom-Gill (CPMG) presat to remove the signal from macromolecules. Only the 1D NOESY-presat pulse sequence was conducted in urine. Metabolites or metabolic features from the NMR spectra with the most discriminatory power between the groups were explored further and associated with disease stage, grade and prognosis to identify potential diagnostic and prognostic biomarkers.

## 4.2 Background

The study of small molecular weight compounds or metabolites and their relationship with biological pathways such as fatty acid oxidation and carbohydrate metabolism within a biological system is referred to as metabonomics. The metabonome is considered the end-point of the omics cascade and therefore can be thought of as closest to the phenotype (Dunn *et al.*, 2011).

Body fluids such as urine or plasma inherently carry detailed information on the metabonome and the metabolic response of the host to pathological stimuli or genetic modification (Nicholson and Lindon, 2008). Profiling of the metabonome provides a relatively new opportunity for biomarker development with respect to particular characteristics of diseases such as cancer and identification of alterations in metabolic pathways associated with cancer pathogenesis.

## 4.3 Results

### *4.3.1 Characteristics of the patients in the study*

As detailed in Chapter 2: Materials and Methods (Section 2.3.4.1), plasma and urine samples were collected from patients undergoing gynaecological surgery at Hammersmith Hospital, London. Although a large proportion of the patients diagnosed with a malignancy had an epithelial ovarian histology (62/106), a proportion of the patients were diagnosed with a malignancy not arising from the ovary such as the endometrium (44/106). This subset of patients is termed the 'Malignant' cohort and has been analysed separately to the EOC cohort, as the primary aim of the study was to identify biomarkers specific to EOC.

After a successful application to the MRC-NIHR National Phenome Centre (NPC), a dedicated centre for large-scale metabolic phenotyping at Imperial College London, samples were analysed using $^{1}$H-NMR.

A total of 305 samples including internal quality control samples were analysed by $^{1}$H-NMR over 4 racks. Any samples from patients who were not pre-operative, had neoadjuvant chemotherapy or were significantly delayed in processing (>24 hours) were removed from analysis, this totalled 69 samples. Fourteen patients had duplicate samples and these are discussed in Section 4.3.2.4. After initial processing of the spectra (see Chapter 2: Section 2.5.1.3) 20 samples were removed from further analysis due to large interfering peaks, most likely from exogenous substances such as drugs causing phasing issues of the spectra.

A total of 186 individual patients and control subjects were included in the study and the clinical characteristics of these subjects are detailed in Table 4.1.

Table 4.1 Clinical characteristics of participants used in the ovarian NMR study.

| | | Healthy (n=14) | Benign (n=50) | Borderline (n=16) | EOC (n=62) | Malignant (n=44) |
|---|---|---|---|---|---|---|
| Age | Mean (range) | 30 (22-43) | 51 (20-84) | 51 (26-72) | 61 (36-85) | 61 (32-83) |
| Grade | 1 | | | | 3 | 8 |
| | 2 | | | | 12 | 16 |
| | 3 | | | | 45 | 15 |
| | Undifferentiated | | | | 0 | 0 |
| | Unknown | | | | 2 | 5 |
| Stage | 1A,1B,1C | | | 6, 1, 4 | 4, 0, 7 | 13, 7, 2 |
| | 2A,2B,2C | | | 1 | 3, 2, 1 | 0, 2, 0 |
| | 3A | | | 1 | 1 | 1 |
| | 3B | | | | 5 | 1 |
| | 3C | | | | 23 | 5 |
| | 4 | | | | 15 | 2 |
| | Unknown | | | 3 | 1 | 11 |
| Histological subtype | Serous | | | | 40 | 3 |
| | Mucinous | | | | 2 | 2 |
| | Endometrioid | | | | 3 | 26 |
| | Clear cell | | | | 10 | 2 |
| | Mixed | | | | 6 | 1 |
| | Carcinosarcoma | | | | 1 | 0 |
| | Unknown | | | | 0 | 10 |
| Residual disease | 0mm | | | | 24 | 2 |
| | 1-10mm | | | | 17 | 0 |
| | >11mm | | | | 16 | 3 |
| | Unknown | | | | 5 | 39 |

### *4.3.2 $^{1}$H CPMG NMR analysis of plasma samples from EOC patients and controls*

Internal quality control samples included in the study consisted of pooled samples; the sample reference (SR) is a pool of all the samples in the study and the long term reference (LTR) is a pooled QC specific to the NPC. In total, 8 LTR and 8 SR samples were included; 2 sets per rack.

Examples of plasma spectra from the healthy, benign and EOC classes are shown in Figure 4.1. Figure 4.1A shows the 600 MHz $^{1}$H- CPMG NMR spectra of a plasma sample from a healthy volunteer with no known malignancy, Figure 4.1B shows the spectra from a patient with benign cystadenofibroma and Figure 4.1C shows the spectra from a patient with early stage EOC. Regions and chemical components have been assigned based on previously published data (Holmes *et al.*, 1997, Nicholls *et al.*, 2003, Wang *et al.*, 2004) as well as in-house identifications.

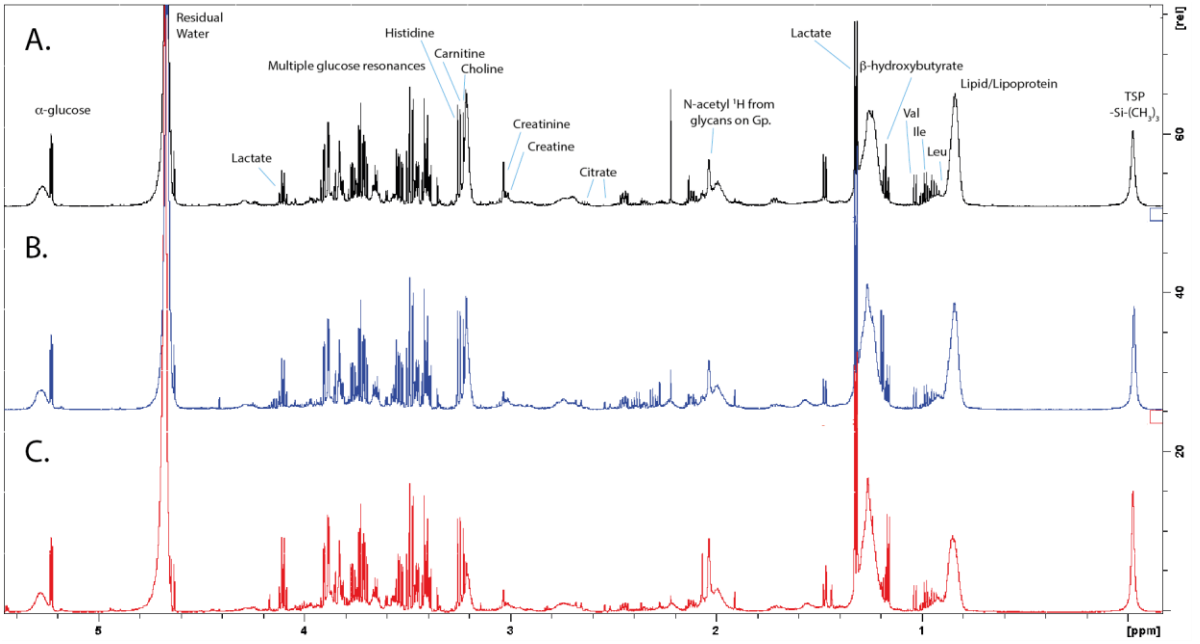



Figure 4.1 Comparison of $^{1}$H CPMG NMR plasma spectra from subjects with **A**, healthy control, **B**, benign cystadenofibroma and **C**, early stage high grade serous papillary EOC.

Plasma samples from patients with EOC (n=62), benign gynaecological disease (n=50) and healthy volunteers (n=14) were averaged and the mean spectra were visually inspected for differences in regions between the classes. Figure 4.2 shows regions between 0.8ppm and 1.2ppm that appear to show differences in lipid resonances at 0.84ppm and in beta-hydroxybutyrate at 1.2ppm. Figure 4.3 shows regions in the averaged spectra between 1.9ppm and 2.6ppm that appear to show differences in glycoprotein acetyls at 2.04ppm and in beta-hydroxybutyrate at 2.31ppm and 2.41ppm.

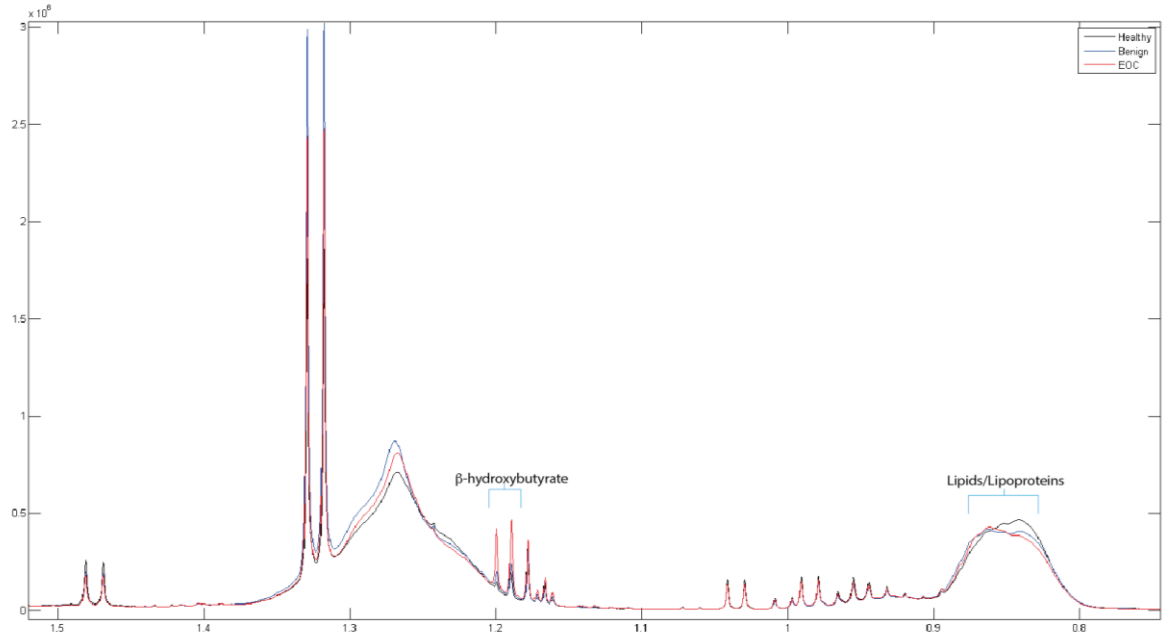



Figure 4.2 Averaged $^{1}$H NMR CPMG plasma spectra highlighting differences in lipid resonances (0.84ppm) and beta-hydroxybutyrate (1.2ppm).

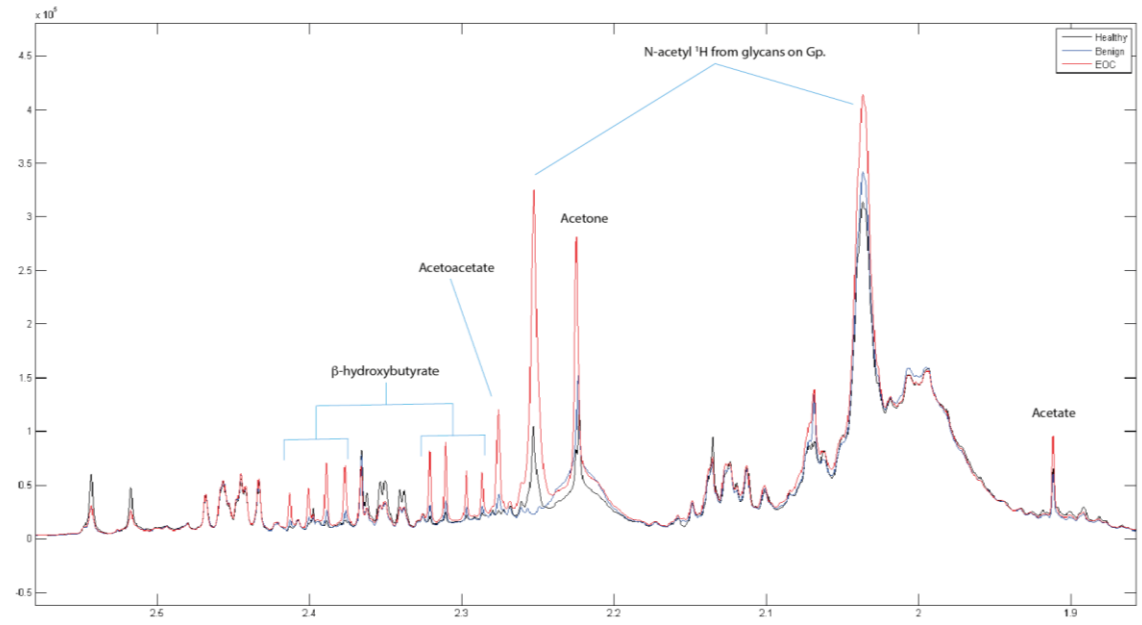



Figure 4.3 Averaged $^{1}$H NMR CPMG plasma spectra highlighting differences in glycoprotein acetyls at 2.04ppm and beta-hydroxybutyrate doublet of doublets at 2.31ppm and 2.41ppm.

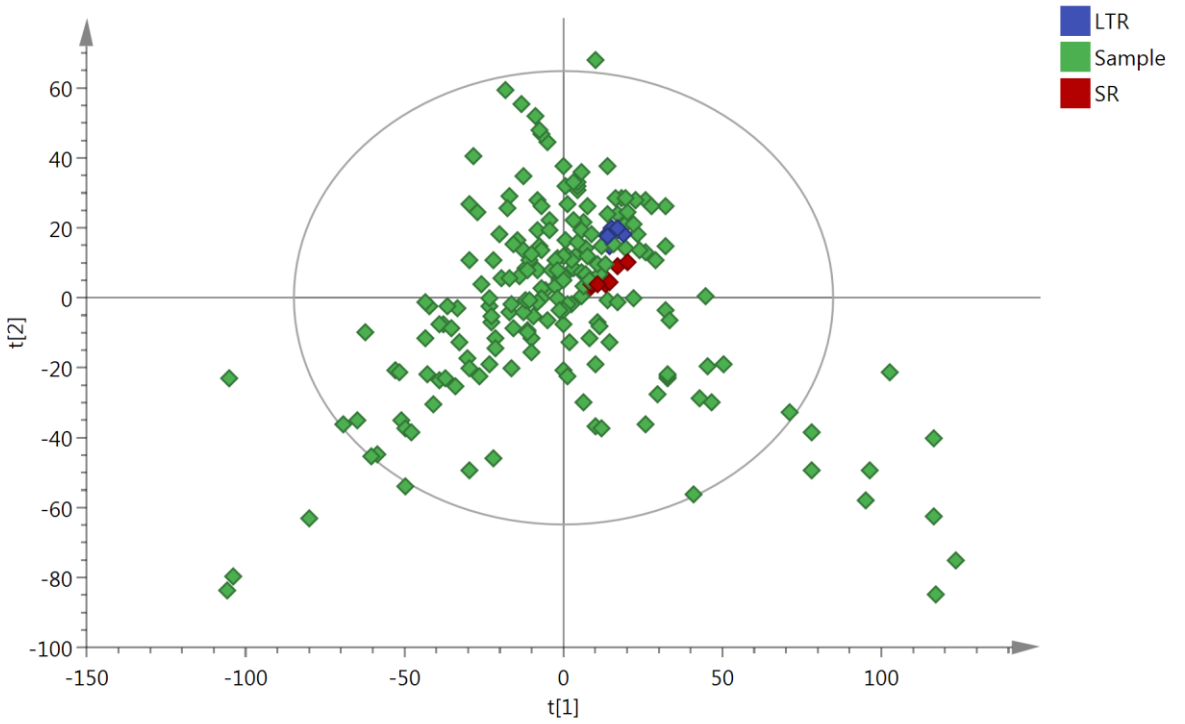
Visual inspection of the averaged $^{1}$H CPMG NMR plasma spectra highlighted differences in resonances corresponding to the ketone, beta-hydroxybutyrate, and broad singlets at 0.84ppm and 2.04ppm, which are likely to be from lipoprotein and glycoprotein acetyls. The integrals of these

peaks were calculated from the CPMG and NOESY spectra and are discussed in Section 4.3.2.2. The differences in lipoprotein subfractions and their lipid constituents were also calculated from the NMR spectra and this is discussed in Section 4.3.5.

After removal of spectral resonances containing $H_20$ and TSP followed by normalisation of the spectra, (see Chapter 2: Section 2.5.1.3) the data was exported to SIMCA-P (v14.0, Umetrics, Sweden) for multivariate statistical analysis.

### 4.3.2.1 Multivariate data analysis of $^1$H CPMG NMR spectra of plasma and quality control samples

All samples were visualised using principal component analysis (PCA) to assess assay performance including clustering and inter-rack reproducibility of IQC samples.

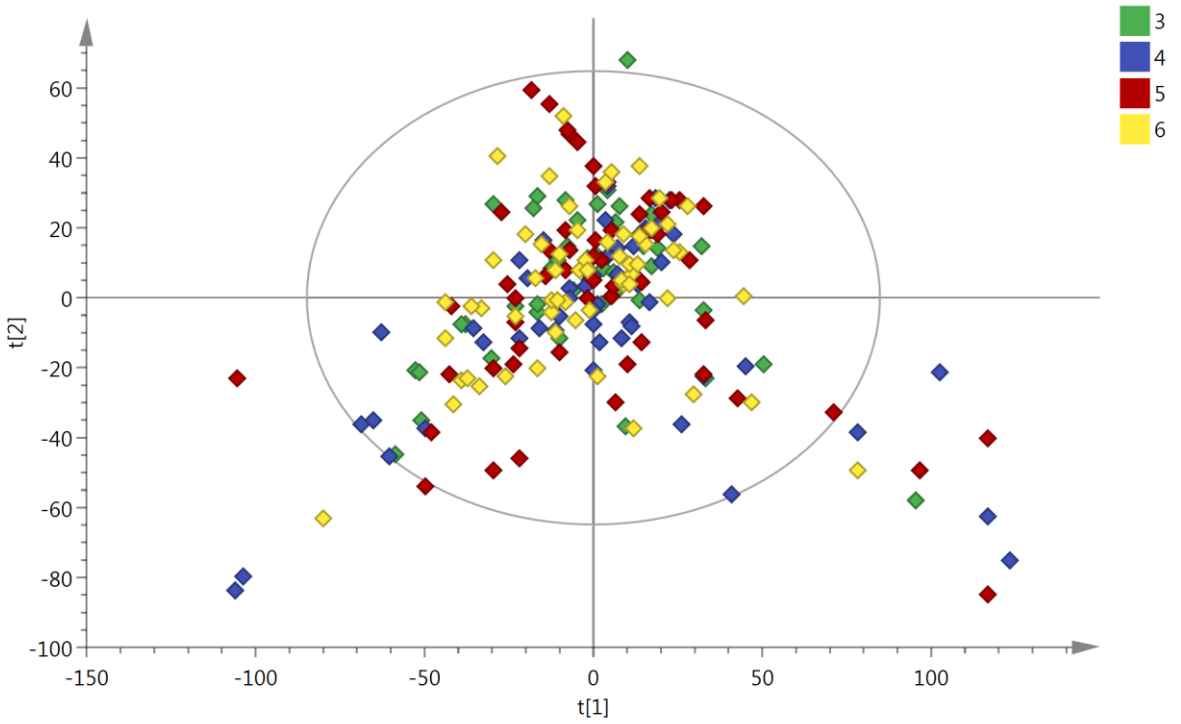



R2X[1] = 0.228 R2X[2] = 0.133 Ellipse: Hotelling's T2 (95%)

Figure 4.4 Unsupervised PCA scores plot of all plasma samples (n=200) and IQC samples: sample reference (SR) (n=8) and long term reference (LTR) (n=8) in the $^1$H NMR CPMG data.

The PCA scores plot in Figure 4.4 shows clustering of the LTR and SR IQC samples in the metabolic space and demonstrates a good repeatability of the IQC samples across the experiment.

To assess if any variability was introduced over time during the experiment, the PCA was coloured according to Rack number to assess the inter-rack variability, shown in Figure 4.5.

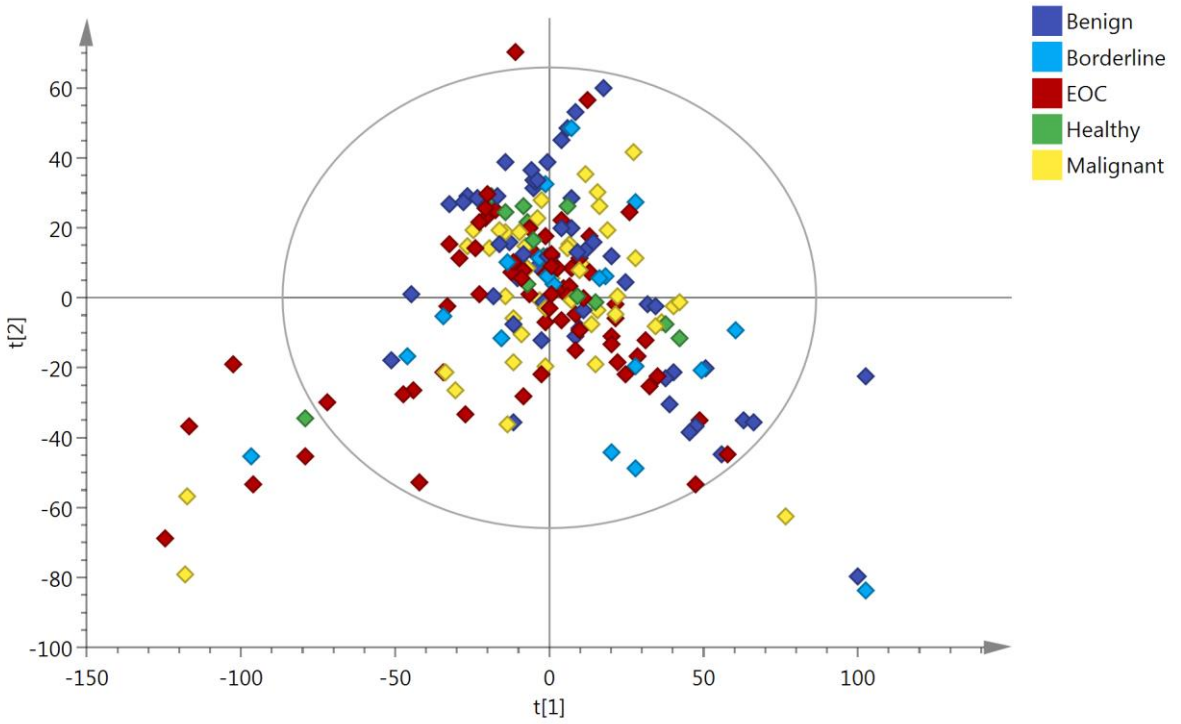



R2X[1] = 0.228 R2X[2] = 0.133 Ellipse: Hotelling's T2 (95%)

Figure 4.5 Unsupervised PCA scores plot of all samples (n=200) and IQC samples (n=16) coloured by rack number (3-6) in the $^{1}$H NMR CPMG data.

In Figure 4.5, PC1 on the x axis explains 22.8% of the variance, PC2 explains 13.3% of the variance. From visual assessment there may be some clustering in PC2 with the majority of samples from Rack 4 (blue) clustering in the lower half of the PCA scores plot. This potential variation in PC2 should be considered in subsequent analysis.

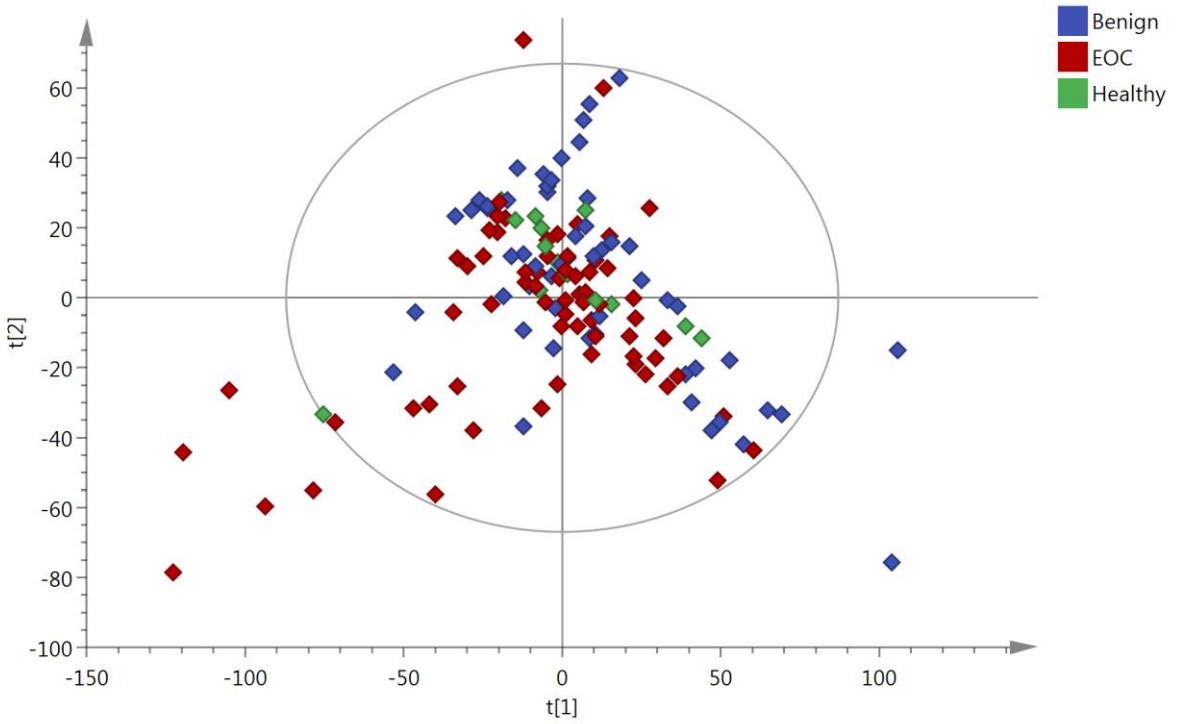
To identify differences in the metabolic signature generated from $^{1}$H NMR analysis of plasma samples, PCA analysis of all the samples in the study is shown in Figure 4.6.




R2X[1] = 0.239 R2X[2] = 0.138 Ellipse: Hotelling's T2 (95%)

Figure 4.6 Unsupervised PCA scores plot of plasma samples coloured according to class in the $^{1}$H NMR CPMG data. Benign – benign gynaecological pathologies (n=55), Borderline – Borderline ovarian mass (n=19), EOC – epithelial ovarian cancer (n=67), Healthy – No known cancer (n=14), Malignant – Cancer not arising from the ovary (n=45).

The PCA scores plot in Figure 4.6, shows clustering of the small number of healthy samples in the upper right hand quadrant. Principal component analysis showed that the largest variance in the model was not explained by age, ethnicity or histological sub-type alone. Further analysis on just benign, EOC and healthy subjects by PCA analysis is shown in Figure 4.7.

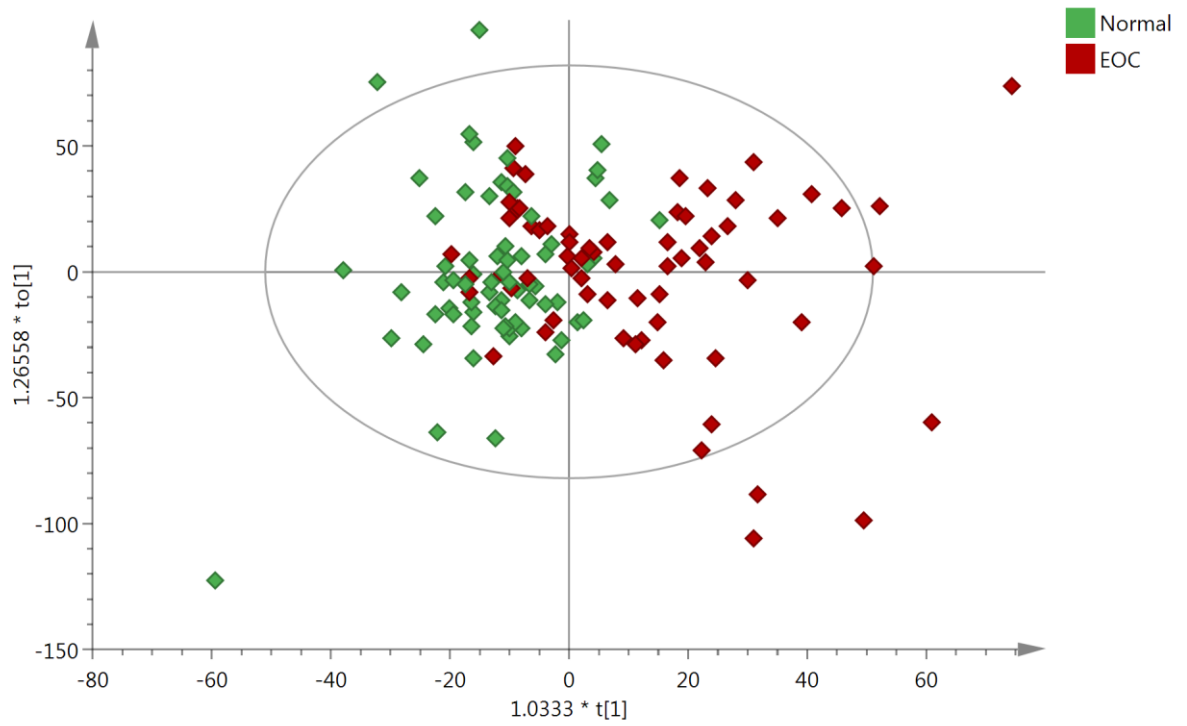



R2X[1] = 0.28 R2X[2] = 0.165 Ellipse: Hotelling's T2 (95%)

Figure 4.7 Unsupervised PCA scores plot of plasma samples from patients with EOC (n=67), benign ovarian disease (n=55) and healthy volunteers (n=14) in the $^{1}$H CPMG NMR data .

The PCA scores plot in Figure 4.7. appears to show some degree of clustering, specifically of the healthy class, with one outlier in the bottom left quadrant but a supervised model would be better suited to understand the features in the spectrum that are different between the groups.

Supervised analysis of the EOC and the 'Normal' group which combines the healthy and benign groups was performed to optimise separation between the groups and identify specific metabolic features/regions that may be different between the EOC group and the 'Normal' group.




R2X[1] = 0.0961 R2Xo[1] = 0.248 Ellipse: Hotelling's T2 (95%)

Figure 4.8 Supervised OPLS-DA scores plot of plasma samples from patients with EOC (n=67) and the 'Normal' group (n=69) combining patients with benign gynaecological disease (n=55) and healthy volunteers (n=14) in the $^{1}$H NMR CPMG data. R2Y=0.362, Q2Y=0.221, R2X=0.494 using one predictive component plus two orthogonal components.

To assess the validity of the supervised model, a permutation plot was produced, Figure 4.9.This plot shows, for a selected Y-variable (Class) the values of R2 and Q2 for the original model (far right) on the vertical axis and for the Y-permuted models further to the left. The horizontal axis shows the correlation between the permuted Y-vectors and the original Y-vector for the selected Y or Class. The original Y has the correlation of 1.0 with itself, defining the high point on the horizontal axis.

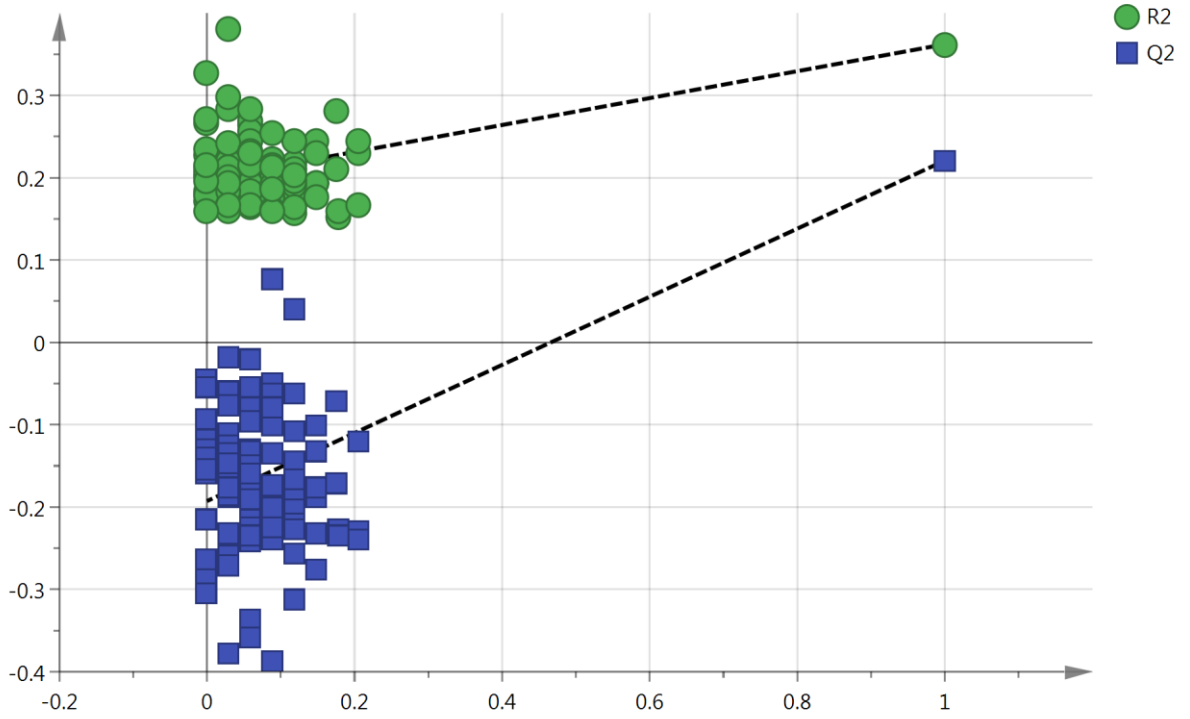



Figure 4.9 Permutation plot of 100 permutations of OPLS-DA model between EOC and 'Normal' groups in the $^{1}$H CPMG NMR data of plasma samples.

As the Y permuted Y-vectors are all lower than the original Y-vector for the selected Y variable (Class) this suggests that the model is valid.

To further elucidate the differences in metabolites of case vs. control the influential loadings responsible for spectral differences from the supervised OPLS-DA was produced, Figure 4.10.

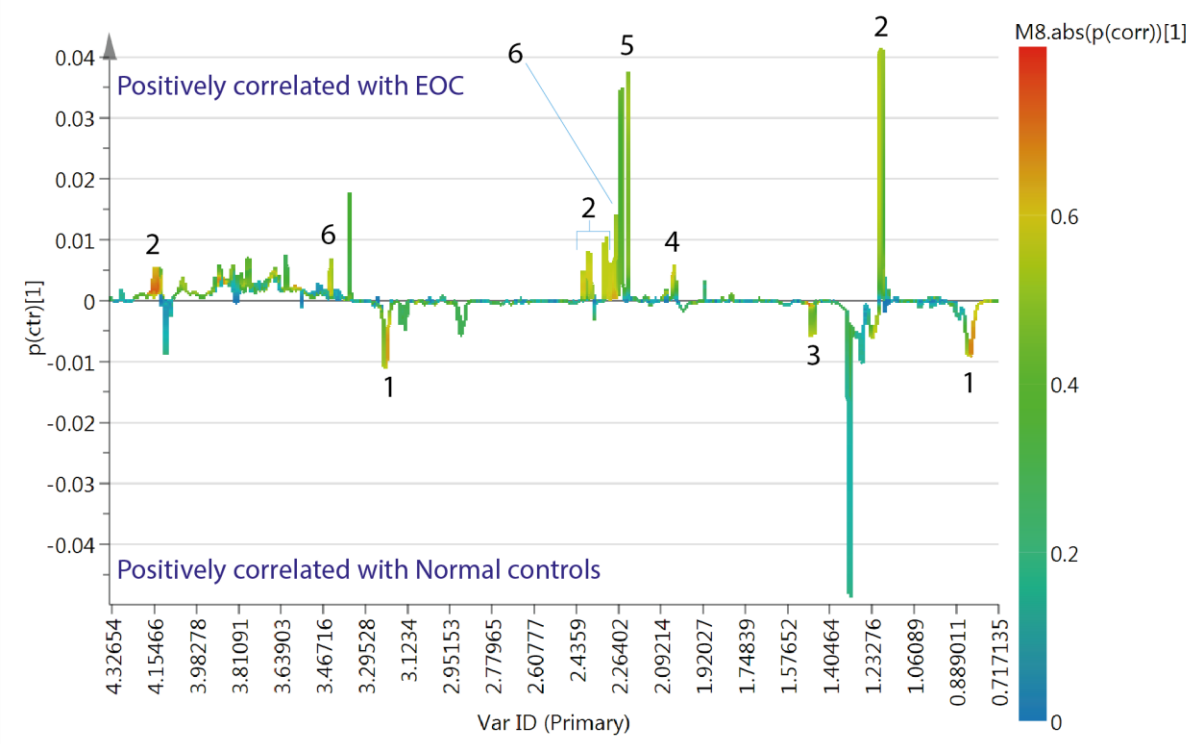



Figure 4.10 Correlation coefficient plot corresponding to $^{1}$H CPMG NMR OPLS-DA model of EOC vs. Normal controls of plasma samples. OPLS-DA coefficient plot labels: **1**, lipoprotein, **2**, beta-hydroxybutyrate, **3,** alanine **4**, N-acetyl glycoprotein **5**, acetone, **6**, acetoacetate, (one predictive component plus two orthogonal components; R2Y=0.362, Q2Y=0.221, R2X=0.494). Figure highlights spectral region between 0.7ppm to 4.33ppm only.

The OPLS-DA correlation coefficient plot allows identification of metabolites that correlate with a particular class. The colour coding refers to the magnitude of correlation and the direction of the peak indicates to which class the variable is positively correlated. In Figure 4.10 the peaks pointing upwards are positively correlated in EOC and the peaks pointing downwards are positively correlated in plasma of control patients.

As predicted from the averaged spectra, the region at 0.84ppm were shown to be correlated with the control group and resonances corresponding to beta-hydroxybutyrate at 1.2ppm, 2.31ppm and 2.41ppm were shown to be associated with the EOC group.

To further confirm the identity of these peaks, a STOCSY analysis of the peak at 1.2ppm was performed. A pseudo NMR spectrum is created showing the correlation of peaks in the spectrum with the peak of interest or driver peak. A STOCSY analysis on a driver peak at δ1.20 is shown in Figure 4.11.

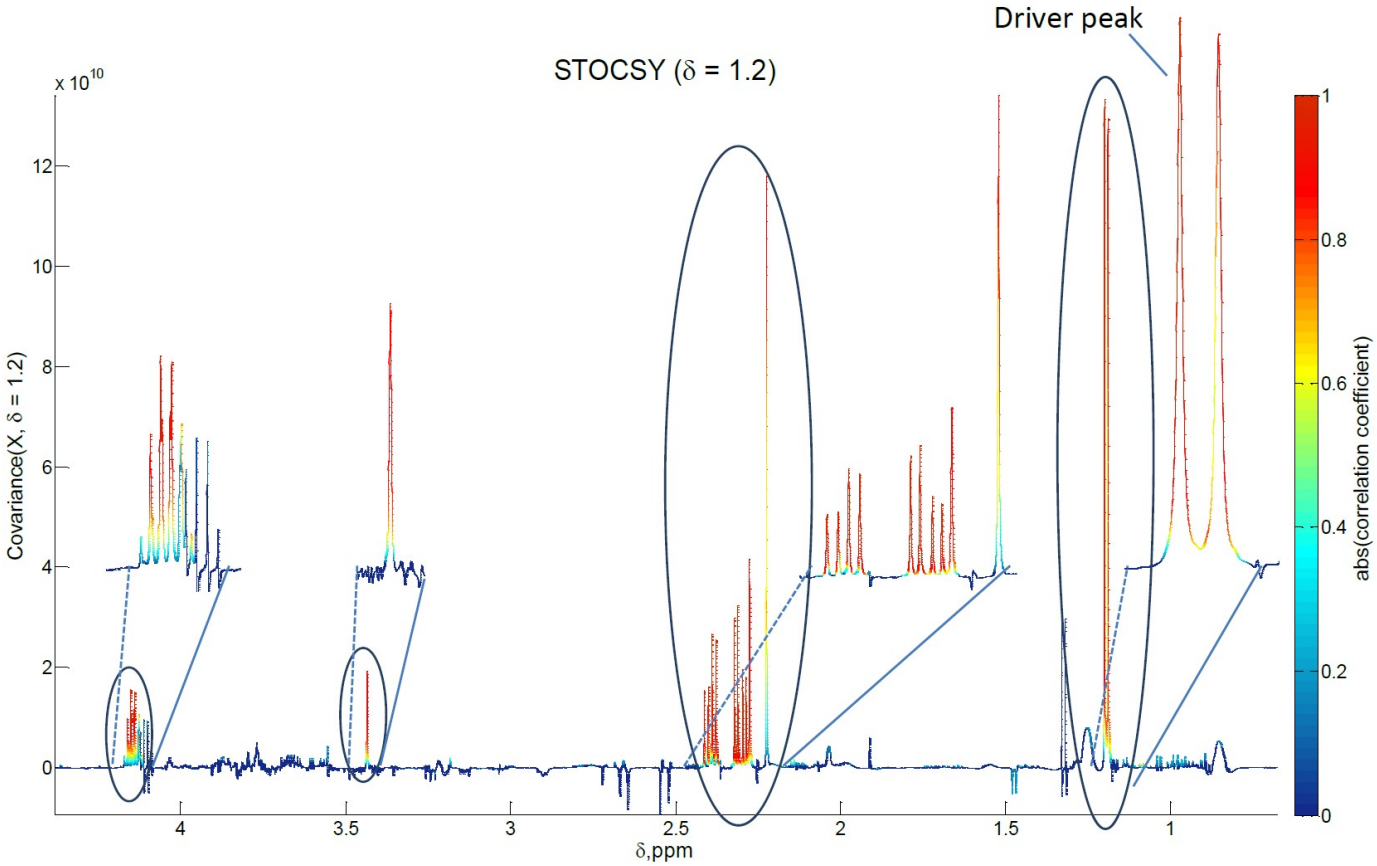



Figure 4.11 STOCSY analysis of a driver peak at δ1.20; part of a double of peaks in this region strongly correlates with a doublet of doublets at δ2.31 and δ2.40 as well as a doublet of triplets at δ4.15 (as indicated by the high correlation coefficient). This strongly suggests that these chemical groups arise from the same molecule, 3-hydroxybutyric acid or also known as beta-hydroxybutyrate. In addition strong correlations exist with singlet peaks at δ2.22, δ2.27 and δ3.43 indicating that these peaks arise from molecules with a biological relationship to the driver peak and can be tentatively assigned as acetone (δ2.22) and acetoacetate (δ2.27 & δ3.43).

To assess the performance of the global NMR profile in the form of the supervised OPLS-DA model in distinguishing patients with EOC from those without, ROC analysis was performed, Figure 4.12.

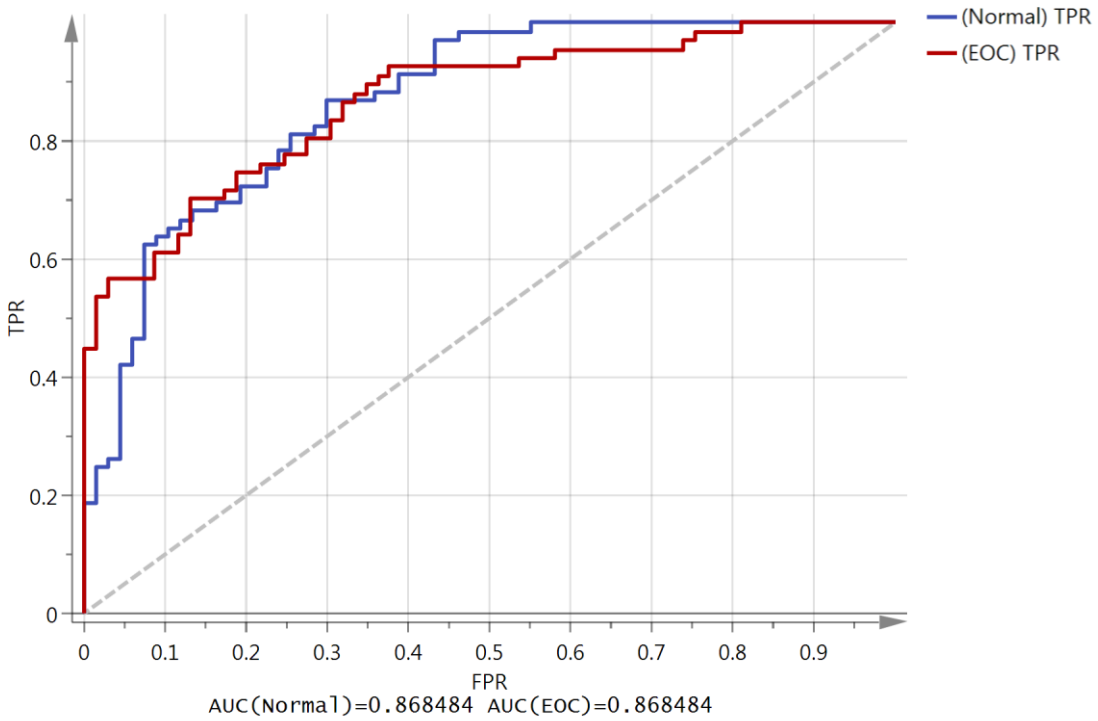



Figure 4.12 ROC curve for 'Normal' versus EOC by metabolic signature based on $^{1}$H NMR CPMG of plasma samples. Note: each curve is a mirror image of the other and therefore AUC represents the ability of the metabolic signature to discriminate between the two groups.

### 4.3.2.2 Univariate data analysis of peak integrals of $^{1}$H CPMG NMR spectra of plasma samples

Visual inspection of the averaged spectra (Figures 4.2 and 4.3) and STOCSY analysis (Figure 4.11) suggested differences in ketone bodies including beta-hydroxybutyrate as well as resonances corresponding to lipoproteins.

Thirty individual peaks including the beta-hydroxybutyrate peak at 1.2ppm were integrated from the CPMG $^{1}$H-NMR spectra for each plasma sample to investigate the phenotype-associated changes in these metabolites. Integration of peaks was performed by Dr ChungHo Lau from Dr Hector Keun's group. The chemical shift, upper and lower boundaries of the peaks and the tentative metabolite assignment are shown in Table 4.2. Any peaks with unassigned identifications are termed unknown.xx.ppm where xx is the region of the peak from the origin in the spectra in parts per million. The %CV in the IQC samples for the 30 metabolites is shown in Table 4.3.

Table 4.2 Table of peaks from CPMG $^{1}$H NMR spectra of plasma samples that were integrated, the chemical shift and upper and lower boundaries for integration.

| **Metabolite** | **Resonance chemical shift (ppm)** | **Peak lower bound (ppm)** | **Peak upper bound (ppm)** |
|---|---|---|---|
| 3-hydroxybutyrate(1.2ppm) | 1.20 | 1.2 | 1.208 |
| Acetate | 1.92 | 1.91 | 1.922 |
| Acetoacetate | 2.26 | 2.247 | 2.272 |
| Acetone | 2.23 | 2.224 | 2.233 |
| Alanine | 1.49 | 1.48 | 1.492 |
| choline/carnitine | 3.20 | 3.193 | 3.202 |
| Citrate | 2.55 | 2.538 | 2.562 |
| creatine/creatine phosphate | 3.91 | 3.907 | 3.918 |
| Creatinine | 4.09 | 4.091 | 4.098 |
| DMSO | 2.73 | 2.713 | 2.743 |
| Ethanol | 1.18 | 1.178 | 1.187 |
| Formate | 8.46 | 8.455 | 8.462 |
| Glucose | 5.24 | 5.227 | 5.25 |
| Glutamate | 2.36 | 2.352 | 2.361 |
| Glutamine | 2.46 | 2.455 | 2.467 |
| Glycine | 3.56 | 3.56 | 3.564 |
| Histidine | 7.06 | 7.046 | 7.07 |
| Isoleucine | 1.02 | 1.009 | 1.022 |
| Lactate | 1.33 | 1.328 | 1.34 |
| Leucine | 0.97 | 0.9557 | 0.9766 |
| Methanol | 3.36 | 3.357 | 3.37 |
| Pyruvate | 2.37 | 2.367 | 2.376 |
| Succinate | 2.40 | 2.398 | 2.403 |
| Tyrosine | 6.90 | 6.887 | 6.915 |
| Valine | 1.03 | 1.028 | 1.041 |
| unknown (1.43ppm) | 1.43 | 1.424 | 1.43 |
| unknown(1.81ppm) | 1.81 | 1.8 | 1.82 |
| unknown(1.84ppm) | 1.84 | 1.834 | 1.855 |
| unknown(2.39ppm) | 2.39 | 2.383 | 2.391 |
| unknown(2.9ppm) | 2.90 | 2.86 | 2.94 |

Table 4.3 Table of the %CV values for the peak integration in the CPMG $^{1}$H NMR spectra of the QC samples.

| Metabolite | %CV in SR (n=8) | %CV in LTR (n=8) |
|---|---|---|
| 3-hydroxybutyrate(1.2ppm) | 3.69 | 3.98 |
| acetate | 3.53 | 4.19 |
| acetoacetate | 2.37 | 3.16 |
| acetone | 7.09 | 2.80 |
| alanine | 2.79 | 3.06 |
| choline/carnitine | 3.64 | 3.48 |
| citrate | 42.47 | 21.28 |
| creatine/creatine phosphate | 2.94 | 2.95 |
| creatinine | 3.70 | 4.00 |
| DMSO | 2.32 | 2.36 |
| ethanol | 2.20 | 2.63 |
| formate | 8.55 | 24.71 |
| glucose | 2.61 | 3.07 |
| glutamate | 8.75 | 3.56 |
| glutamine | 4.74 | 4.88 |
| glycine | 7.09 | 7.24 |
| histidine | 3.19 | 3.17 |
| Isoleucine | 4.19 | 3.87 |
| lactate | 4.50 | 3.64 |
| leucine | 2.76 | 3.07 |
| methanol | 2.64 | 4.75 |
| pyruvate | 2.76 | 4.49 |
| succinate | 3.07 | 2.72 |
| tyrosine | 2.24 | 4.09 |
| valine | 3.00 | 2.41 |
| unknown (1.43ppm) | 5.56 | 3.57 |
| unknown(1.81ppm) | 3.81 | 4.63 |
| unknown(1.84ppm) | 4.06 | 3.71 |
| unknown(2.39ppm) | 4.24 | 4.60 |
| unknown(2.9ppm) | 5.42 | 2.82 |

The Wilcoxon rank sum test was used to identify peak integrals that were significantly different between the EOC and the Normal group. A total of 30 peak integrals were investigated in EOC (n=62), benign (n=50) and healthy groups (n=14). Duplicate samples were removed to ensure each sample was from a unique patient. The p-values from the Wilcoxon-rank sum test were FDR adjusted using the Benjamini-Hochberg method and an adjusted p-value <0.05 was considered significant.

A total of 10 metabolites were significantly different between the EOC and Normal groups including acetoacetate and acetone. Beta-hydroxybutyrate was not significantly different between groups when adjusted for multiple testing however; there was a significant difference before FDR-

correction. Box and whisker plots of the FDR-corrected and not-FDR corrected metabolites are shown in Figures 4.13 and 4.14 respectively.

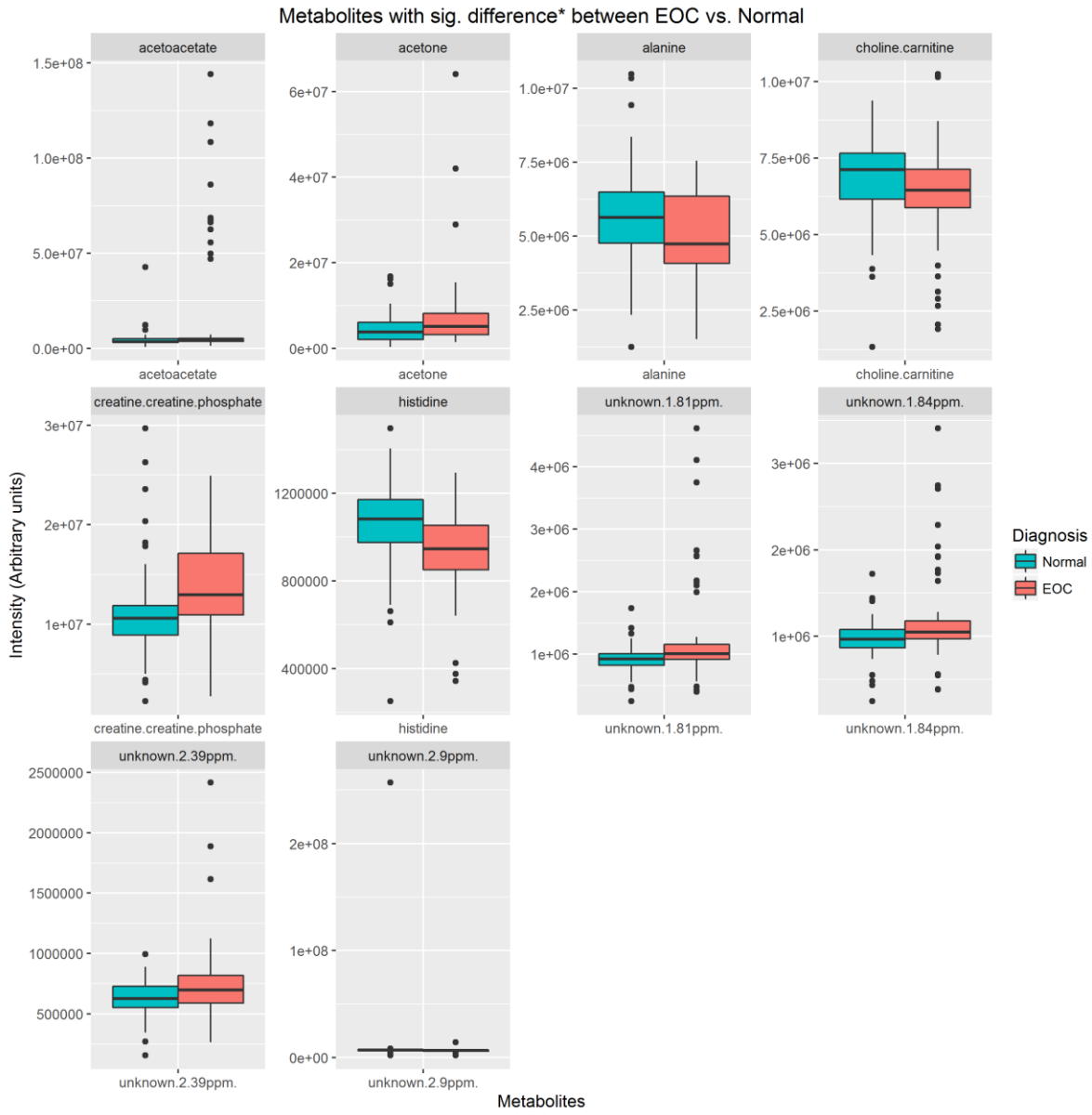



Figure 4.13. Metabolite integrals from CPMG-$^{1}$H NMR data that are significantly different between the EOC and 'Normal' groups. *FDR adjusted p-value <0.05.

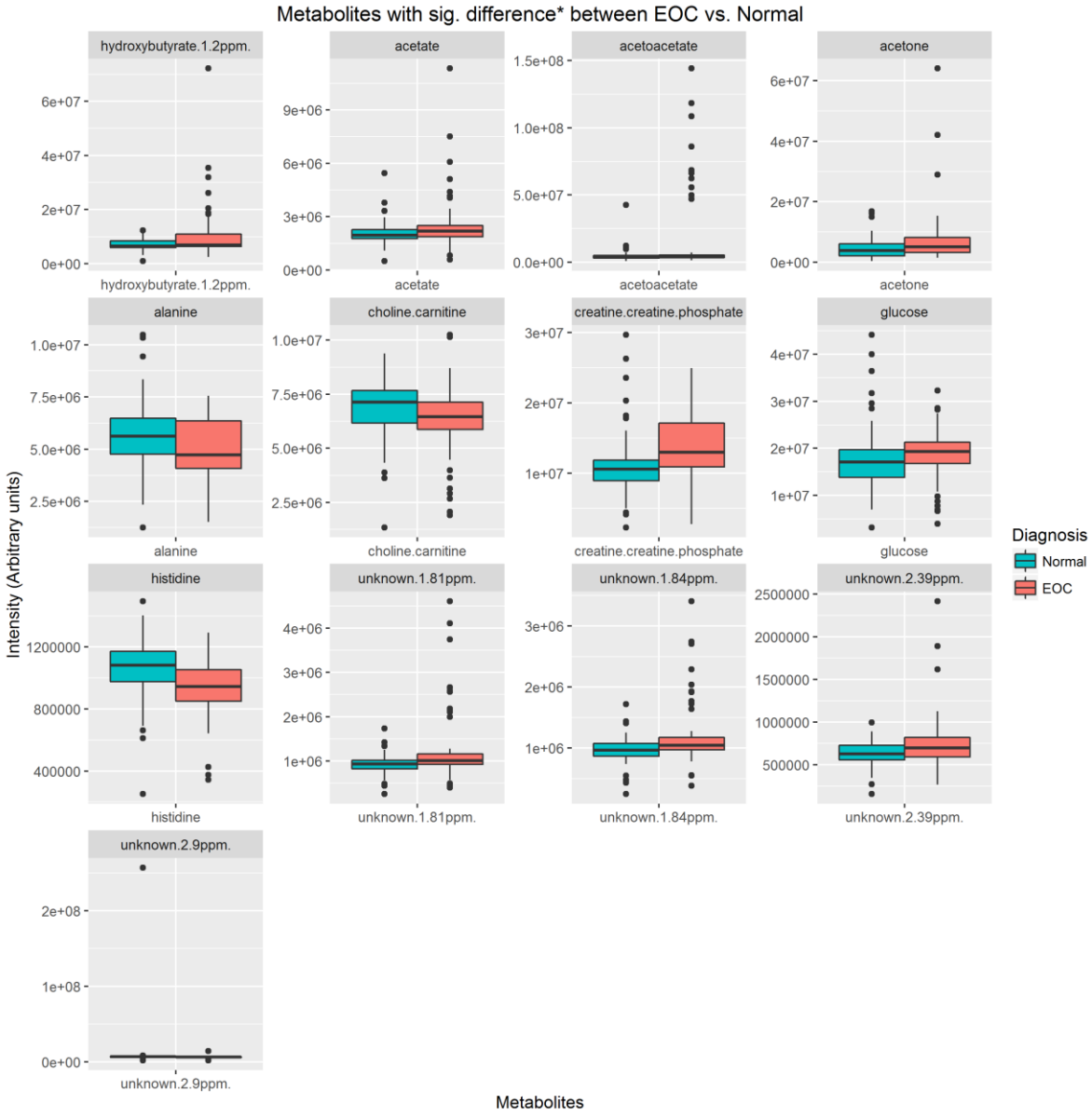



Figure 4.14. Metabolite integrals from CPMG-$^{1}$H NMR data that are significantly different between the EOC and 'Normal' groups. *p-value <0.05.

#### 4.3.2.2.1 Association of metabolites with grade and stage of disease

To further elucidate the potential role of these metabolites, rank correlation using kendall's tau was used to determine whether these metabolites were also significantly associated with stage or grade, i.e. for the metabolites that are higher in EOC were they also higher in patients with more advanced stage disease.

Rank correlation of the significantly different metabolites between the EOC and 'Normal' group showed that histidine and unknown.2.39ppm were associated with stage. Following the pattern seen between the EOC and normal group, histidine showed a decreasing concentration based on

stage while unknown.2.39ppm showed an increasing concentration with increasing stage. In addition logistic regression analysis of stage (dichotomised into high (Stage III/IV) and low (Stage I/II) also showed a significant difference in peak integrals of alanine and beta-hydroxybutyrate.1.2ppm where higher stage disease is associated with higher values of beta-hydroxybutyrate and lower values of alanine. Figure 4.15 shows the box and whisker plot with the scatter of samples dichotomised to high stage (stage III and IV) and low stage (stage I and II). Values for the benign/healthy samples are shown for comparison.

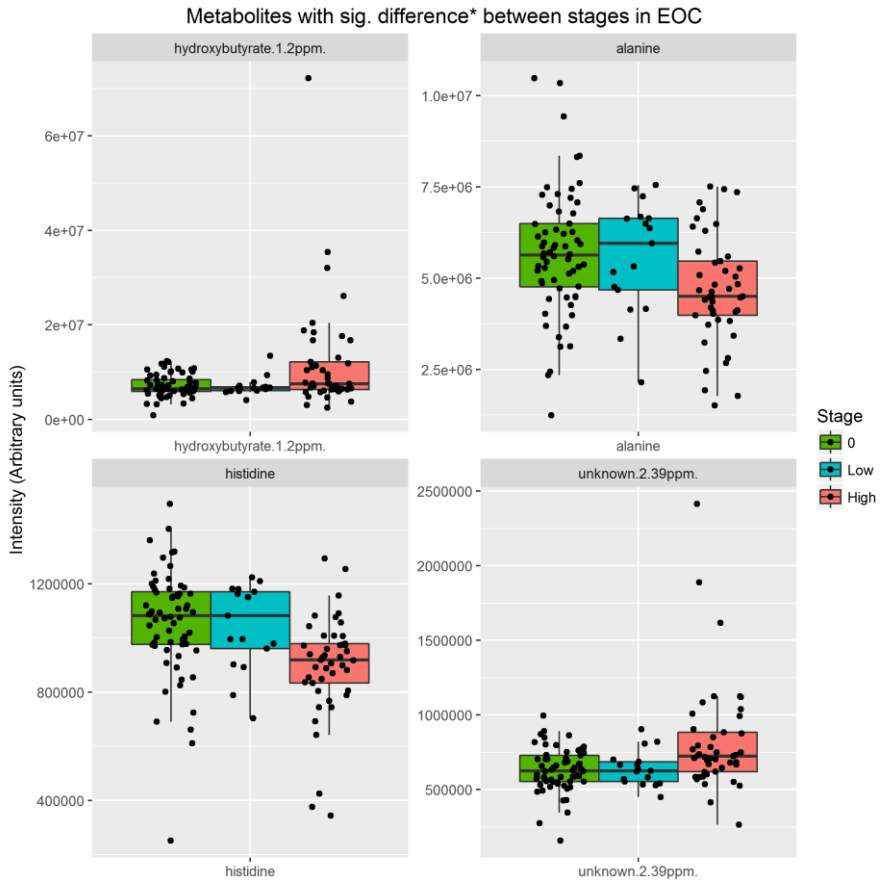



Figure 4.15 Box and whisker plots of metabolite integrals from CPMG-$^{1}$H NMR data that show an association between stage (Low = Stage I,II and High = Stage III,IV) and metabolite integral, in the EOC group only. Values are also shown for samples with stage 0 (benign/healthy) for comparison.

There was no association between metabolite peak integral value and grade of disease in any of the significantly different metabolites between EOC and the normal group.

### 4.3.2.3 Predictive capacity of metabolite integrals for diagnosis of EOC in plasma

To assess the performance of each of the significantly different metabolites in diagnosing EOC, ROC curves were produced and the AUC determined for each metabolite peak integral. The highest AUC

was based on metabolite peak integrals of histidine (AUC=0.697), Creatine/Creatine phosphate (AUC=0.686) and unknown.1.81ppm (AUC=0.681). A combination of markers in a predictive model may further increase the predictive power of the metabolites and therefore a stepwise additive regression was used from known predictors to find the best model fit according to Akaike information criterion (AIC).The model requires all metabolites to be independent of one another and therefore Pearson correlation between each metabolite was determined, see Figure 4.16.

Pearson's correlation coefficient was calculated for the peak integral values of each of the thirteen metabolites that were significantly different between samples from patients with EOC compared to healthy and benign patients against the others. Only metabolites with an $R^2$ <0.7 compared to the other metabolites were included. The concentration of histidine, creatine/creatine phosphate, unknown.1.81ppm and unknown.2.39ppm were combined in a step-wise logistic regression model to produce a final model consisting of 3 of these metabolites excluding creatine/creatine phosphate.

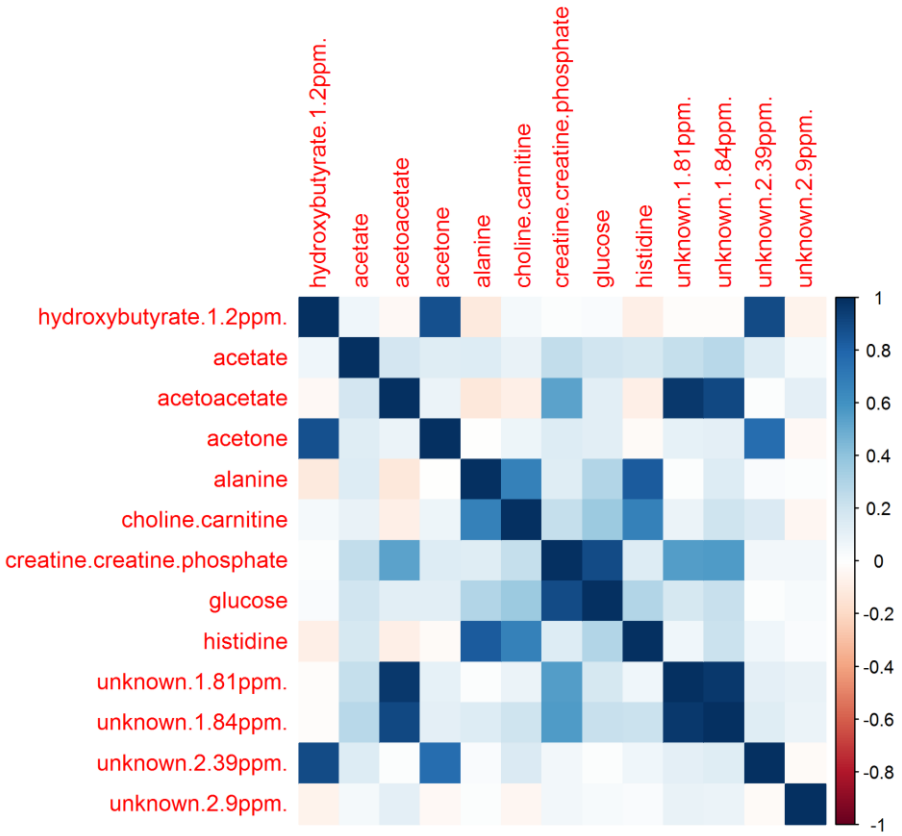



Figure 4.16 Correlation plot of the 13 metabolite integrals from CPMG-$^1$H NMR data between the EOC and 'Normal' groups.

As there was no prospectively collected patient cohort to test the model, the combined predictive model was internally validated using the Leave one out cross validation (LOOCV) method.

Using the Leave one out Cross Validation (LOOCV) method based on metabolite integral of the 3 metabolites, 100 out of 126 (79%) samples were correctly predicted and all samples were classified, as there was no missing data for the peak integrals from $^{1}$H CPMG NMR.

A confusion matrix of the predictive ability of the model is shown in Table 4.4. The positive predictive value (PPV) = 82% and negative predictive value (NPV) =77 % and the sensitivity = 74% and specificity = 84%.

Table 4.4 Confusion matrix for the performance of the cross-validated step-wise logistic regression model for the 3 metabolite integrals from CPMG-$^{1}$H NMR data with significantly different levels in plasma from patients with EOC and without.

| | | **State of disease** | |
|---|---|---|---|
| | | Disease + | Disease – |
| **Results of LOOCV model** | Test + | True positive<br>46 | False positive<br>10 |
| | Test – | False negative<br>16 | True negative<br>54 |

ROC analysis of the cross-validated step-wise logistic regression model produced an AUC of 0.819, see Figure 4.17.

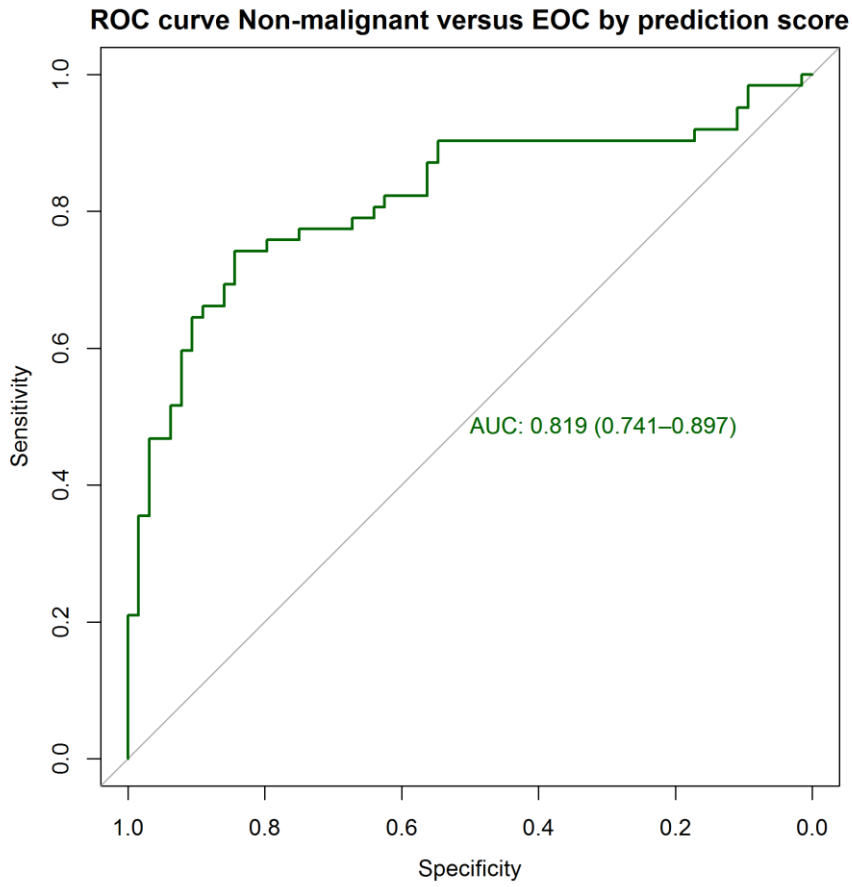



Figure 4.17 ROC curve for non-malignant 'Normal' versus EOC by cross-validated prediction score based on step-wise logistic regression model of 3 metabolites from the CPMG-$^{1}$H NMR data.

#### 4.3.2.4 Variation of selected metabolites in plasma

Biological replicates were taken from 14 patients at different time points by different people. To compare the replicates, concentrations of the predictive metabolites between the duplicate samples were compared. The reproducibility (precision) using the coefficient of variation (%CV) of each metabolite between each replicate is shown in Table 4.5. A prediction score for the biological replicates that were removed from the original analysis was also produced using the LOOCV model.

Table 4.5. Coefficient of variation (%) between biological replicates for significantly different metabolite integrals from CPMG-$^{1}$H-NMR data used in stepwise logistic regression model (TN- True negative, FN – False negative, TP – True positive, FP – False positive)

| Duplicate | Histidine (%CV) | Unknown. 1.81ppm (%CV) | Unknown. 2.39ppm (%CV) | Unknown. 2.9ppm (%CV) | Time between sampling (days) | Class | Prediction score from cross validated model of (binary vector) | | Model performance | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Biological replicate included in model | Biological replicate excluded from model | Biological replicate included in model | Biological replicate excluded from model |
| D2 | 0.66 | 0.18 | 1.12 | 0.26 | 0 | Benign | 0.16 | 0.14 | TN | TN |
| D4 | 3.06 | 3.79 | 1.22 | 1.86 | 1 | Malignant | NA | NA | NA | NA |
| D5 | 7.67 | 10.32 | 6.54 | 8.00 | 0 | EOC | 0.21 | 0.21 | FN | FN |
| D7 | 1.28 | 0.49 | 3.14 | 1.62 | 0 | EOC | 0.60 | 0.62 | TP | TP |
| D8 | 1.76 | 2.27 | 0.24 | 1.79 | 1 | Benign | 0.43 | 0.31 | TN | TN |
| D10 | 1.98 | 3.16 | 5.46 | 2.31 | 1 | Borderline | NA | NA | NA | NA |
| D12 | 3.83 | 4.75 | 6.05 | 5.90 | 2 | Benign | 0.84 | 0.79 | FP | FP |
| D13 | 15.78 | 12.35 | 16.76 | 12.66 | 0 | Borderline | NA | NA | NA | NA |
| D14 | 3.02 | 0.13 | 8.68 | 1.67 | 0 | EOC | 0.34 | 0.53 | FN | TP |
| D15 | 9.13 | 1.01 | 1.43 | 0.21 | 0 | EOC | 0.59 | 0.81 | TP | TP |
| D17 | 1.46 | 7.12 | 6.62 | 4.40 | 0 | Benign | 0.26 | 0.495 | TN | TN |
| D18 | 30.75 | 43.27 | 57.07 | 44.01 | 5 | Borderline | NA | NA | NA | NA |
| D20 | 1.03 | 5.14 | 6.58 | 0.16 | 0 | EOC | 0.31 | 0.20 | FN | FN |
| D21 | 4.32 | 1.22 | 8.43 | 1.00 | 3 | Benign | 0.36 | 0.28 | TN | TN |

Comparison of the prediction scores across the duplicate samples revealed that one sample classification changed. Duplicate D14 changed from false negative to true positive, see Table 4.5.

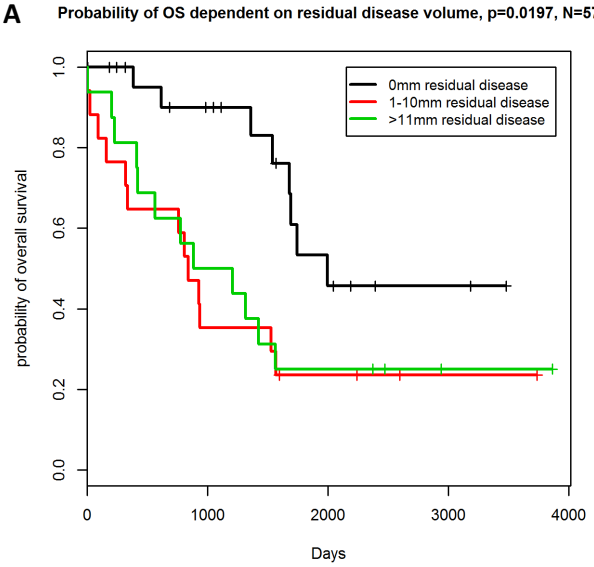
### *4.3.3 Survival analysis*

Overall survival (OS) and progression-free survival (PFS) data was available on 62 EOC patients. Univariate logrank statistical tests were performed. There was a significant difference in OS and PFS dependent on residual disease volume post-surgery divided into 0mm, 1-10mm and >11mm, log-rank p-value= 0.0197 and p=0.00978 respectively (Figure 4.18 A & B). Although there was no significant difference between OS in patients defined as optimally (≤10mm residual disease) and suboptimally debulked (>10mm residual disease), p=0.197 (Figure 4.18 C) the greatest survival advantage came in those patients where complete macroscopic debulking was achieved (0mm residual disease) (Figure 4.18 E). This is in keeping with previously published data (Wimberger *et al.*, 2010). Overall median survival time in the group with 0mm residual disease was 1997 days (95% CI 1678-NA) compared to 883 days (95% CI 565-1565) in the group with any macroscopic residual disease whereas patients with optimal surgical debulking had a median overall survival time of 1678 days (95% CI 1359-NA) compared to 1045 days (95% CI 419-NA) in those who had suboptimal debulking.

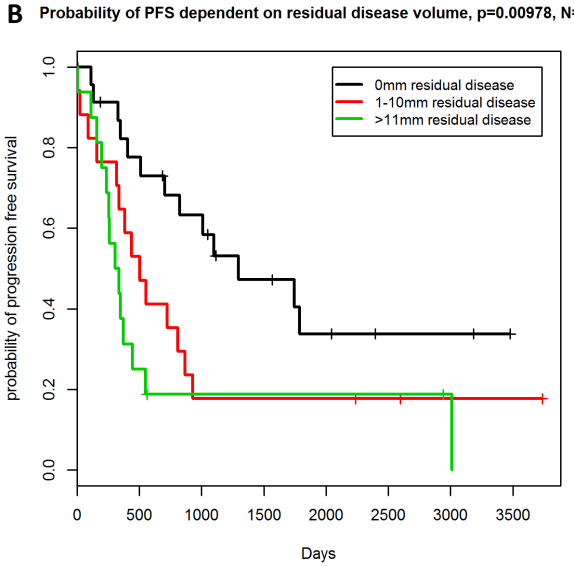
Progression free median survival time in the group with 0mm residual disease was 1297 days (95% CI 829-NA) compared to 371 days (95% CI 308-722) in the group with any macroscopic residual disease (Figure 4.18 F). Patients with optimal surgical debulking had a progression free survival time of 826 days (95% CI 510-1789) compared to 323 days (95% CI 236-NA) in those who had suboptimal debulking highlighting that those patients with the most residual disease (>10mm) have the shortest time to progression (Figure 4.18 D).

The significant difference in overall survival dependent on volume of residual disease was lost in multivariate cox analysis when adjusting for age, stage and grade (p=0.356, HR=1.22, 95%CI 0.80-1.88) but was still significant dependent on presence (>1mm) versus absence of disease (0mm) when adjusted for confounding factors (p=0.041, HR=2.44, 95%CI 1.04-5.74). The significant difference in progression free survival dependent on residual disease volume was also lost in multivariate cox analysis when adjusting for age, stage and grade (p=0.162, HR=1.34, 95%CI 0.89-2.02) and presence of residual disease (p=0.089, HR=1.87, 95%CI 0.91-3.86). Univariate analysis demonstrated that stage and grade were significantly associated with PFS, p=0.022 and p=0.044 respectively and therefore were adjusted for in subsequent models.

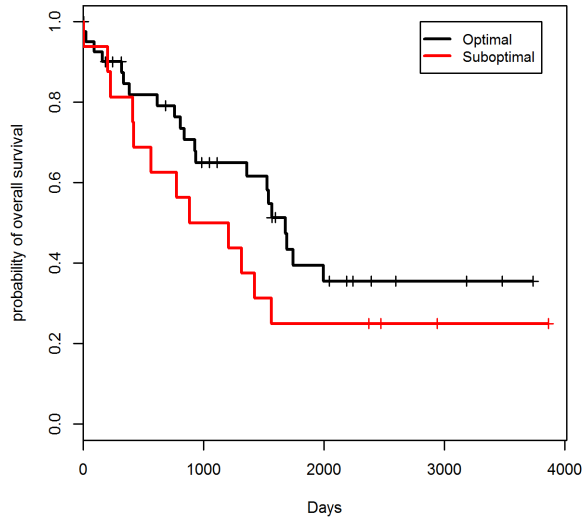
A


Probability of OS dependent on residual disease volume, p=0.0197, N=57
0mm residual disease
1-10mm residual disease
>11mm residual disease
probability of overall survival
0.0
0.2
0.4
0.6
0.8
1.0
0
1000
2000
3000
4000
Days

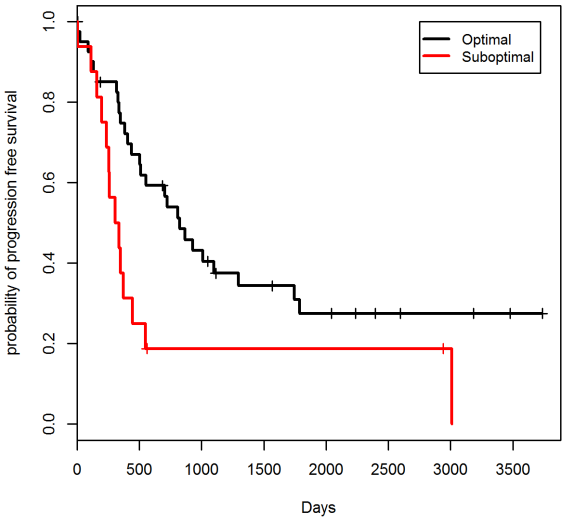
B


Probability of PFS dependent on residual disease volume, p=0.00978, N=57
0mm residual disease
1-10mm residual disease
>11mm residual disease
probability of progression free survival
0.0
0.2
0.4
0.6
0.8
1.0
0
500
1000
1500
2000
2500
3000
3500
Days

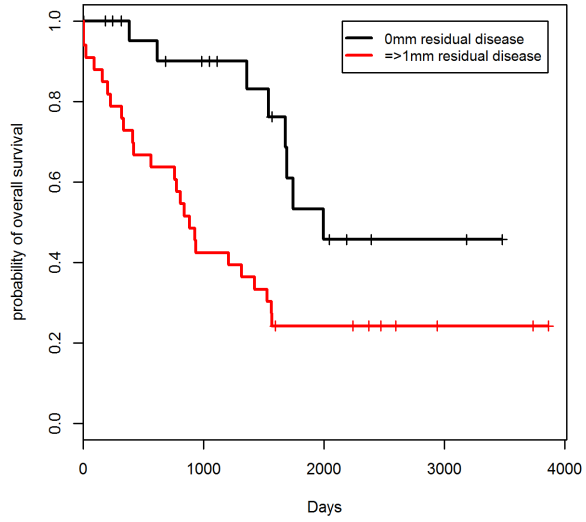
C


Probability of OS, optimal vs. suboptimal, p=0.197, N=57
Optimal
Suboptimal
probability of overall survival
0.0
0.2
0.4
0.6
0.8
1.0
0
1000
2000
3000
4000
Days

D


Probability of PFS, optimal vs. suboptimal, p=0.0128, N=57
Optimal
Suboptimal
probability of progression free survival
0.0
0.2
0.4
0.6
0.8
1.0
0
500
1000
1500
2000
2500
3000
3500
Days

E


Probability of OS, total vs. any residual disease, p=0.00552, N=57
0mm residual disease
=>1mm residual disease
probability of overall survival
0.0
0.2
0.4
0.6
0.8
1.0
0
1000
2000
3000
4000
Days

F

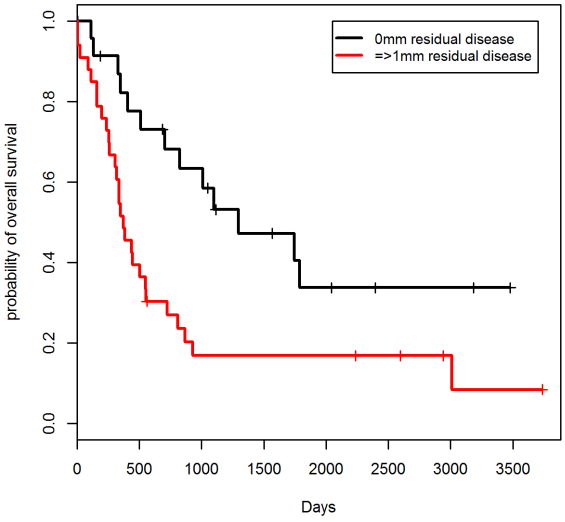

Probability of PFS, total vs. any residual disease, p=0.00574, N=57
0mm residual disease
=>1mm residual disease
probability of overall survival
0.0
0.2
0.4
0.6
0.8
1.0
0
500
1000
1500
2000
2500
3000
3500
Days

Figure 4.18. Kaplan-Meier survival plots for OS and PFS in the NMR ovarian study of patients with EOC. **A**, OS and **B**, PFS dependent on residual disease volume. **C**, OS and **D**, PFS dependent on optimal or suboptimal debulking. **E**, OS and **F**, PFS dependent on complete macroscopic debulking (0mm residual disease) vs. any amount of remaining residual disease (>1mm).

#### 4.3.3.1 Predicting survival in EOC patients to aid surgical stratification using $^{1}$H CPMG NMR peak integrals

Currently there are no predictive models or biomarkers that can be used to identify patients with the poorest prognosis despite optimal surgical treatment (Borley *et al.*, 2012). Identification of treatment stratification biomarkers in patients pre-treatment would enable personalised treatment pathways to be developed to optimise patient care. If a subset of patients could be identified who would respond poorly to treatment regardless of achieving optimal debulking, this may warrant withholding supraradical surgery to reduce the morbidity and mortality associated with this type of surgery and decrease the time to adjuvant chemotherapy. In addition, this group of patients may also benefit from upfront neoadjuvant chemotherapy with interval debulking. If a blood based diagnostic signature could identify these groups of patients, randomised control trials could be designed to prove this hypothesis and provide an opportunity for the development of personalised treatment stratification in patients with ovarian cancer.

##### 4.3.3.1.1 Significantly different metabolite integrals associated with overall survival in all patients with epithelial ovarian cancer

In order to identify metabolites in blood plasma that are associated with survival, data from integration of $^{1}$H CPMG NMR peaks was used from all patients diagnosed with epithelial ovarian cancer. The nonparametric logrank test was used to identify if patients had different survival distributions above and below the median of each metabolite concentration that was previously shown to have a significantly different concentration between EOC and benign/healthy controls in the initial phase of this study. Multivariable cox proportional hazard models adjusting for age, stage, grade and presence of residual disease (none vs. >1mm) were then performed on these metabolites.

One metabolite peak integral was found to be associated with overall survival when adjusted for confounding factors (p=0.00972). Values below the median concentration of beta-hydroxybutyrate at 1.2ppm were found to be associated with improved survival: HR 2.76 (95% CI 1.28-5.97) (Figure 4.19). The overall median survival time in the group with beta-hydroxybutyrate integral value below the median was 1744 days (95% CI 1528-NA) compared to 927 days (95% CI 713-1678) in the group with peak integral value above the median.

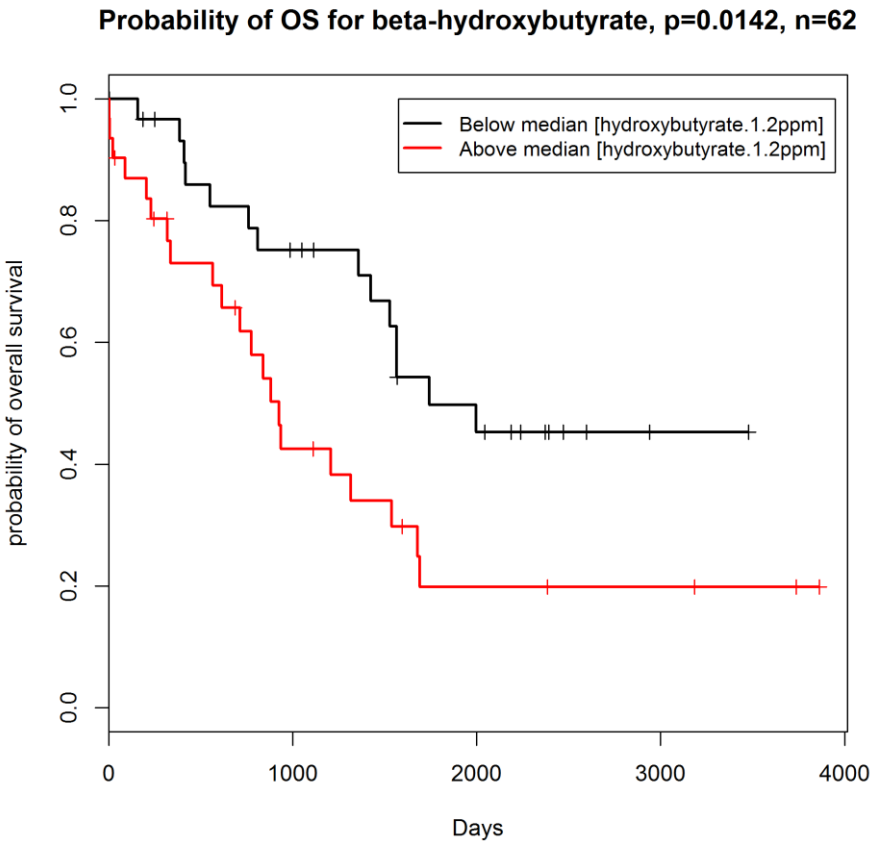



Figure 4.19 Kaplan-Meier survival curve for OS dependent on beta-hydroxybutyrate concentration.

4.3.3.1.2 Significantly different metabolite integrals associated with progression-free survival in all patients with epithelial ovarian cancer

The same analysis was repeated with progression free survival data in order to identify any peak integral values of $^{1}$H CPMG NMR peaks in blood plasma that were associated with progression free survival in all patients diagnosed with epithelial ovarian cancer.

The nonparametric logrank test was used to identify if patients had different survival distributions above and below the median of each metabolite concentration that was previously shown to have a significantly different concentration between EOC and benign/healthy controls in the initial phase of this study. Multivariable cox proportional hazard models adjusting for age, stage, grade and presence of residual disease (none vs. >1mm) were then performed on these metabolites. Concentrations of one metabolite, the peak integral at 2.39ppm, was found to be associated with progression free survival when adjusted for confounding factors (p=0.0376). Integral values of this peak below the median were associated with improved progression-free survival: HR 2.21 (95% CI 1.05-4.66) (Figure 4.20). The median progression free survival in the group with concentration below

the median was 1297 days (95% CI 722-NA) compared to 338 days (95% CI 257-705) for concentrations above the median.

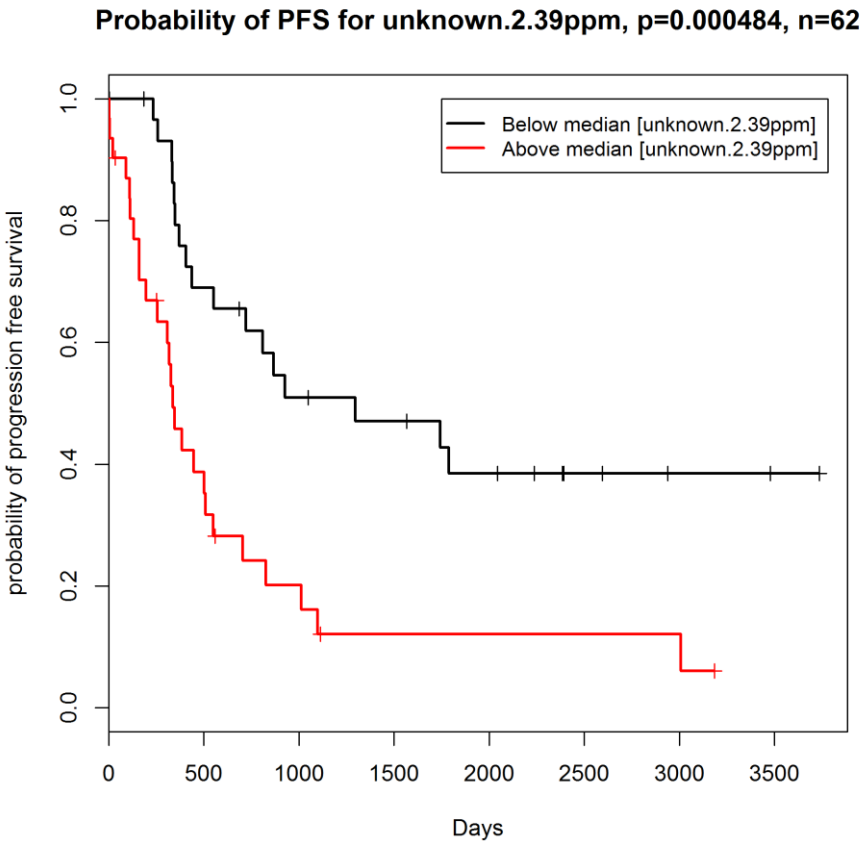



Figure 4.20 Kaplan-Meier survival curve for PFS dependent on the peak integration of metabolite feature at 2.39ppm.

### *4.3.4 $^{1}$H NOESY NMR analysis of plasma samples from EOC patients and controls*

In addition to the $^{1}$H CPMG NMR, data was also collected using the NOESY pre-saturation sequence. The same samples as in Section 4.3.2 were analysed with one additional outlier removed for issues relating to spectral alignment.

Examples of plasma spectra from the healthy, benign and EOC classes are shown in Figure 4.21. Figure 4.21A shows the 600 MHz $^{1}$H-NOESY NMR spectra of a plasma sample from a healthy volunteer with no known malignancy, Figure 4.21B shows the spectra from a patient with benign cystadenofibroma and Figure 4.21C shows the spectra from a patient with early stage EOC.

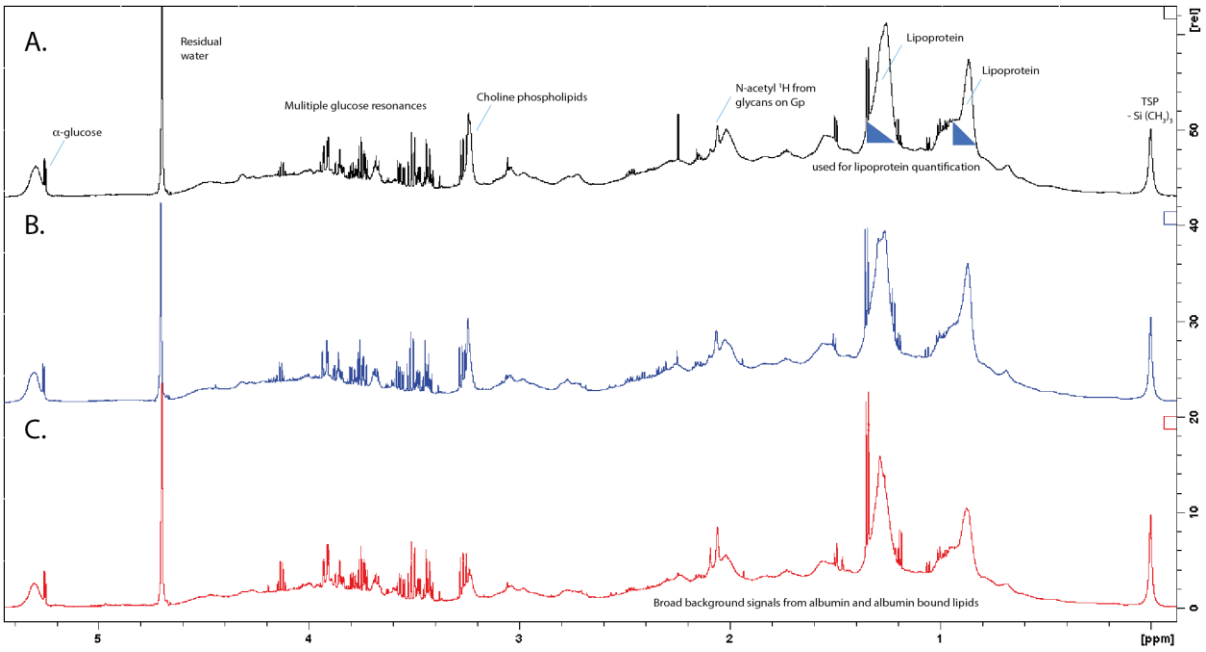



Figure 4.21 Comparison of $^{1}$H NOESY NMR plasma spectra from subjects with **A**, healthy control, **B**, benign cystadenofibroma and **C**, early stage high grade serous papillary EOC.

As with the $^{1}$H CPMG spectra, plasma samples from patients with EOC (n=62), benign gynaecological disease (n=50) and healthy volunteers (n=14) were averaged and the mean spectra visually inspected for differences in regions between the classes. Figure 4.22 shows regions between 1.9ppm and 3.3ppm that appear to show differences in glycoprotein acetyls at 2.04ppm, consistent with the CPMG averaged spectra in Figure 4.3 and in choline phospholipid resonances at 3.215ppm.

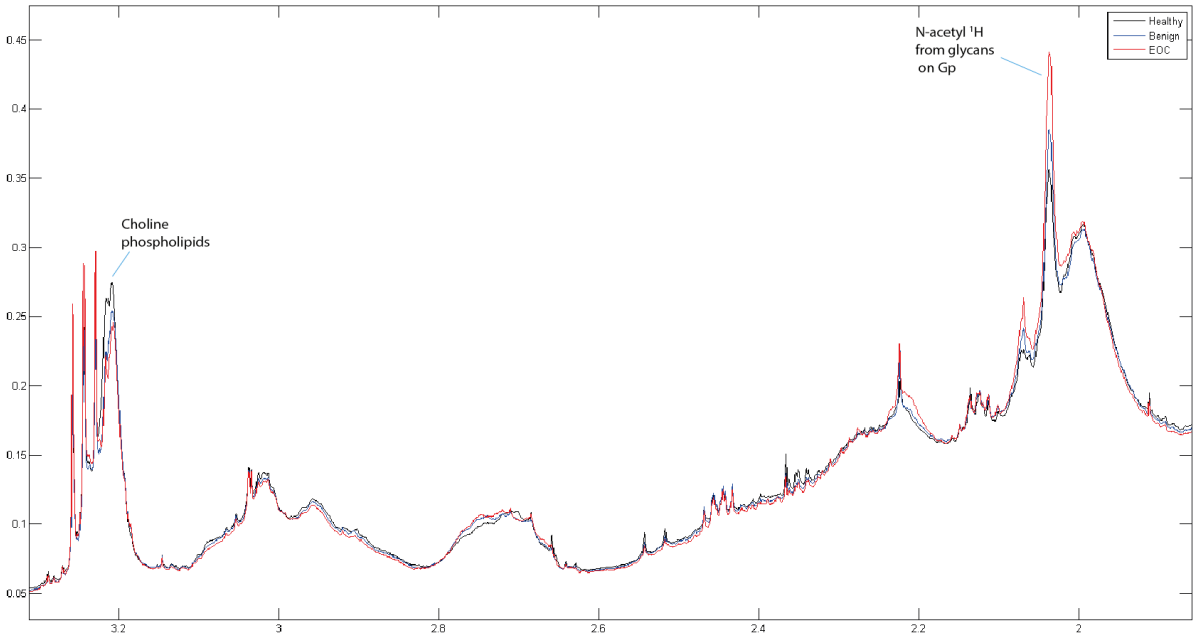
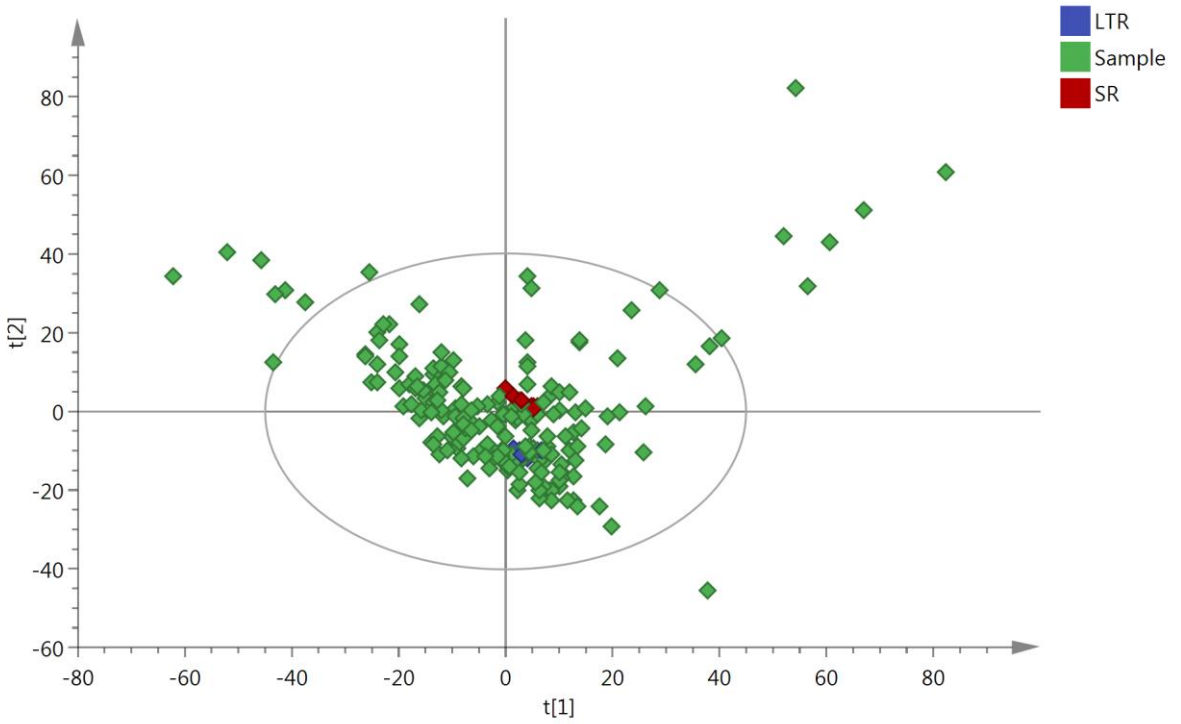



Figure 4.22 Averaged $^{1}$H NMR NOESY plasma spectra highlighting differences in glycoprotein acetyls (2.04ppm) and choline phospholipids (3.215ppm).

The peak areas corresponding to these chemical resonances are further discussed in Section 4.3.4.2.

4.3.4.1 Multivariate data analysis of NOESY-presat-$^{1}$H NMR spectra of plasma and quality control samples.

All samples were visualised using principal component analysis (PCA) to assess assay performance including clustering and inter-rack reproducibility of IQC samples.

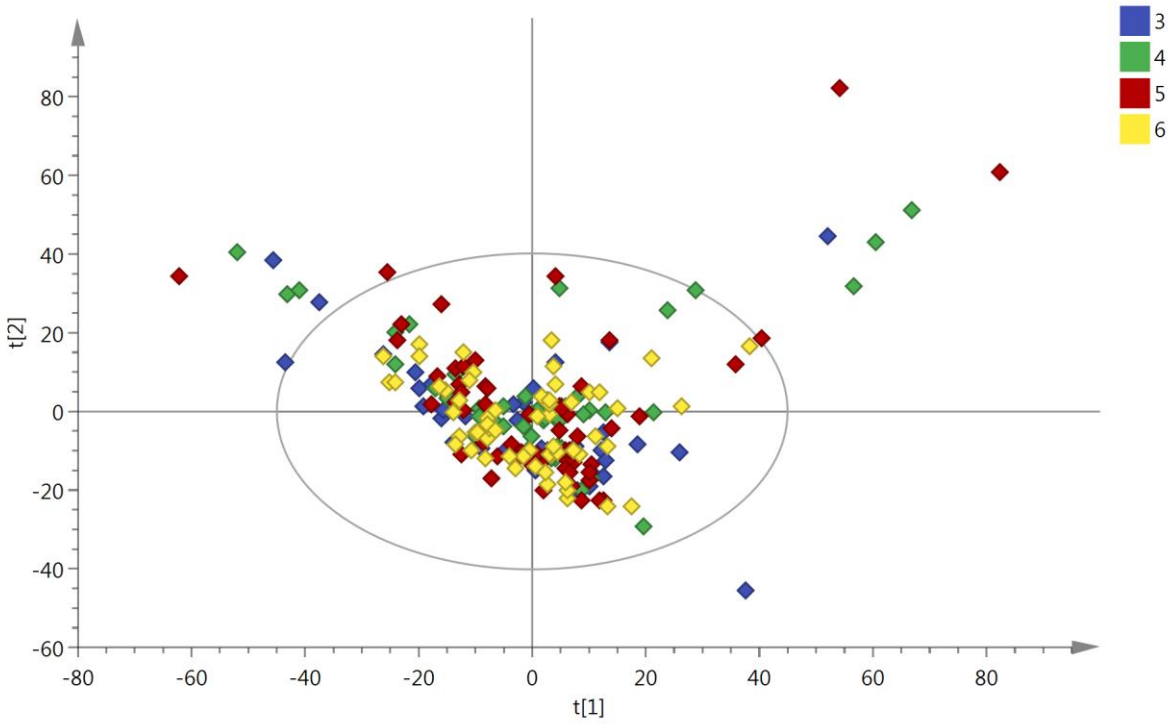



R2X[1] = 0.267 R2X[2] = 0.213 Ellipse: Hotelling's T2 (95%)

Figure 4.23 Unsupervised PCA scores plot of all samples (n=199) and IQC samples; SR (n=8) and LTR (n=8) in the $^{1}$H NMR NOESY data.

The PCA scores plot in Figure 4.23 shows clustering of the LTR and SR IQC samples in the metabolic space and demonstrates a good repeatability of the IQC samples across the experiment.

To assess if any variability was introduced over time during the experiment, the PCA was coloured according to Rack number to assess the inter-rack variability, shown in Figure 4.24.

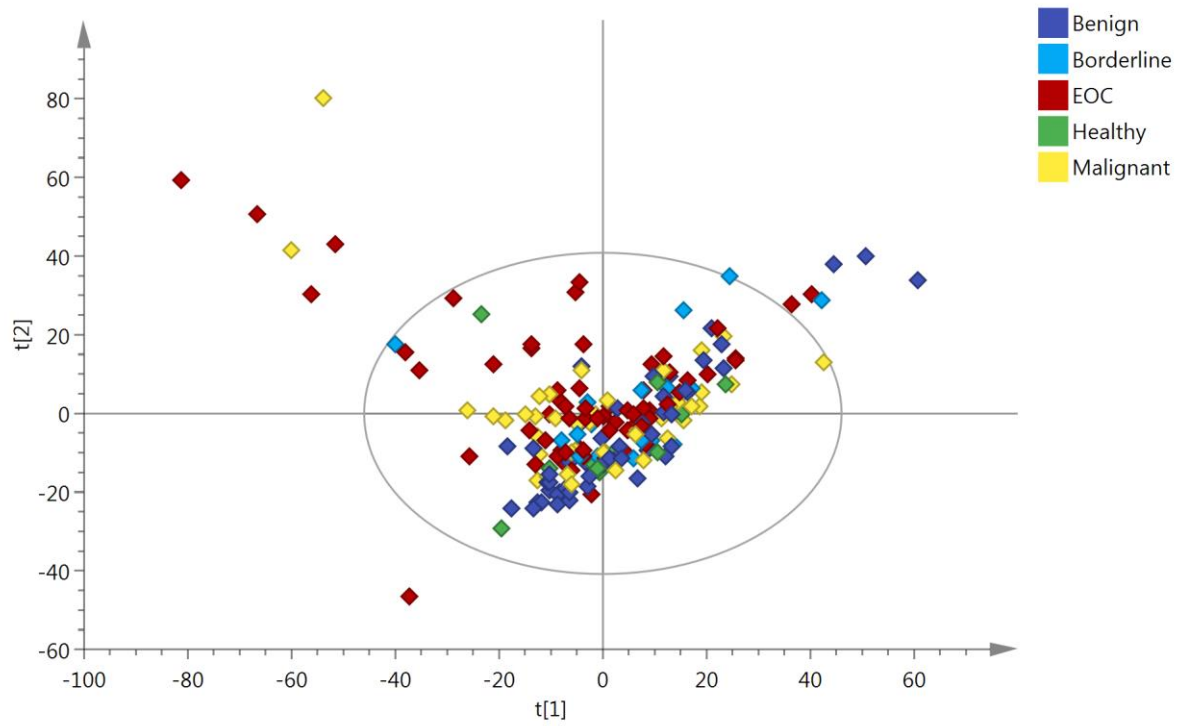



R2X[1] = 0.267 R2X[2] = 0.213 Ellipse: Hotelling's T2 (95%)

Figure 4.24 Unsupervised PCA scores plot of all samples (n=199) and IQC samples (n=16) coloured by rack number (3-6) in the $^{1}$H NMR NOESY data.

In Figure 4.24, PC1 on the x axis explains 26.7% of the variance, PC2 explains 21.3% of the variance. From visual assessment there appears to be minimal bias introduced across the experiment as there is no clustering of samples according to Rack number in the PCA scores plot.

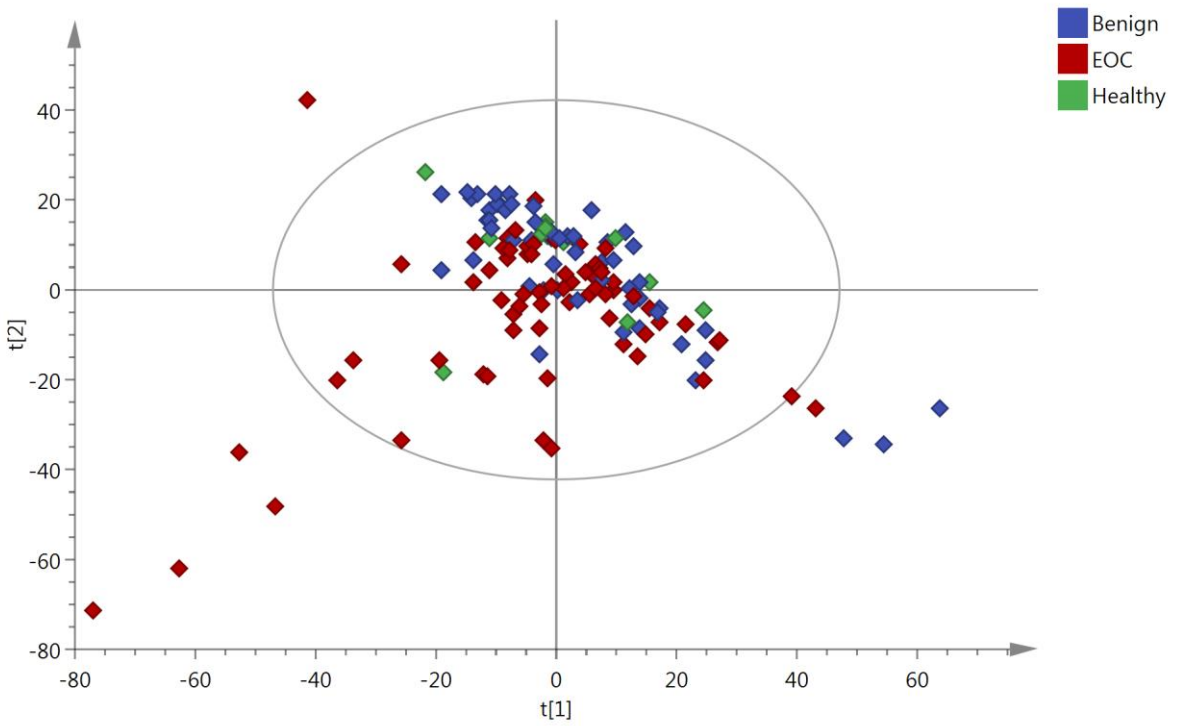
To identify differences in the metabolic signature generated from $^{1}$H NOESY NMR analysis of plasma samples, PCA analysis of all the samples in the study is shown in Figure 4.25.




R2X[1] = 0.276 R2X[2] = 0.217 Ellipse: Hotelling's T2 (95%)

Figure 4.25 Unsupervised PCA scores plot of plasma samples coloured according to class in the $^{1}$H NMR NOESY data. Benign – benign gynaecological pathologies (n=55), Borderline – Borderline ovarian mass (n=18), EOC – epithelial ovarian cancer (n=67), Healthy – No known cancer (n=14), Malignant – Cancer not arising from the ovary (n=45).

The samples are tightly packed but the benign and healthy samples appear to collect in the lower quadrants away from the EOC samples. Principal component analysis showed that the largest variance in the model was not explained by age, ethnicity or histological sub-type alone. Further analysis on just benign, EOC and healthy subjects by PCA analysis is shown in Figure 4.26.




R2X[1] = 0.309 R2X[2] = 0.248 Ellipse: Hotelling's T2 (95%)

Figure 4.26 Unsupervised PCA scores plot of plasma samples from patients with EOC (n=67), benign ovarian disease (n=55) and healthy volunteers (n=14) in the $^{1}$H NOESY NMR data.

In concordance with the 1H-CPMG analysis of the same samples, samples from the healthy and benign groups appear to cluster away from the EOC group with one healthy outlier in the lower left quadrant.

Supervised analysis of the EOC and the 'Normal' group which combines the healthy and benign groups was performed to optimise separation between the groups and identify specific metabolic features/regions that may be different between the EOC group and the 'Normal' group. The healthy and benign groups were combined to reflect the clinical situation where the biomarkers would be used to distinguish patients with an ovarian malignancy from a group of patients with or without a benign ovarian mass.

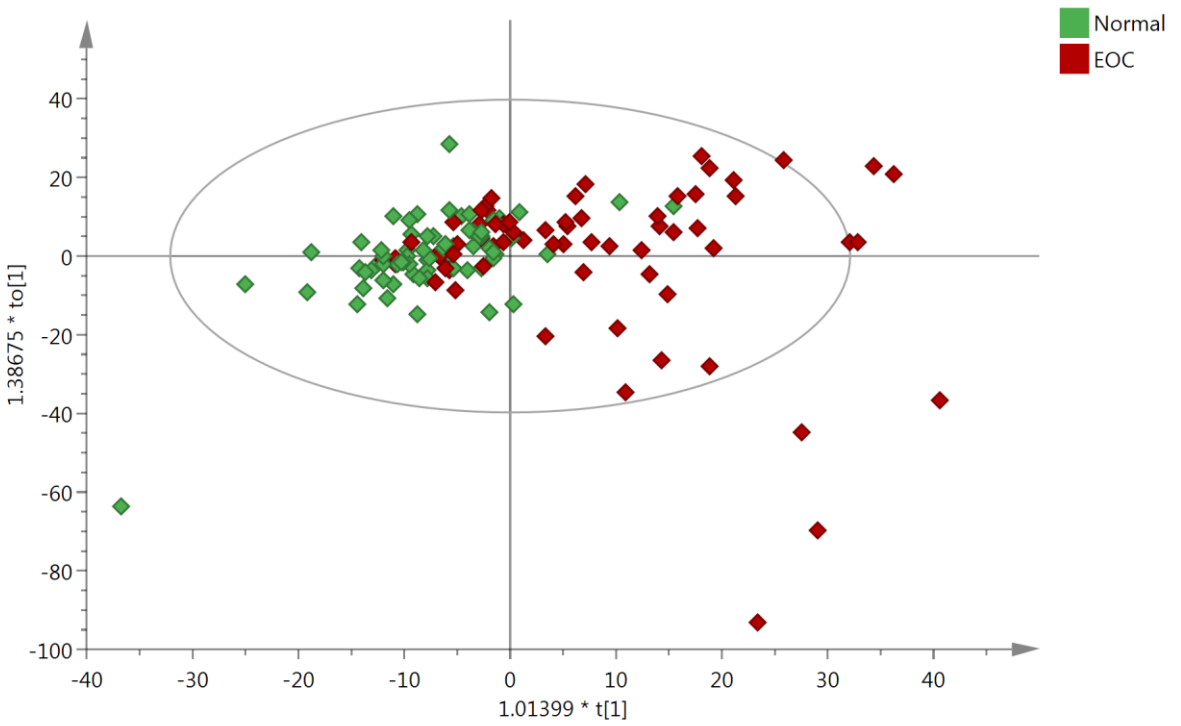



Figure 4.27 Supervised OPLS-DA scores plot of plasma samples from patients with EOC (n=67) and 'Normal' group (n=69) combining patients with benign gynaecological disease and healthy volunteers in the $^{1}$H NMR NOESY data. R2Y=0.351, Q2Y=0.237, R2X=0.614 using one predictive component plus two orthogonal components;

Permutation of the model, shown in Figure 4.28 suggests that the model is valid as all permuted Y-vectors are lower than the original Y-vector.

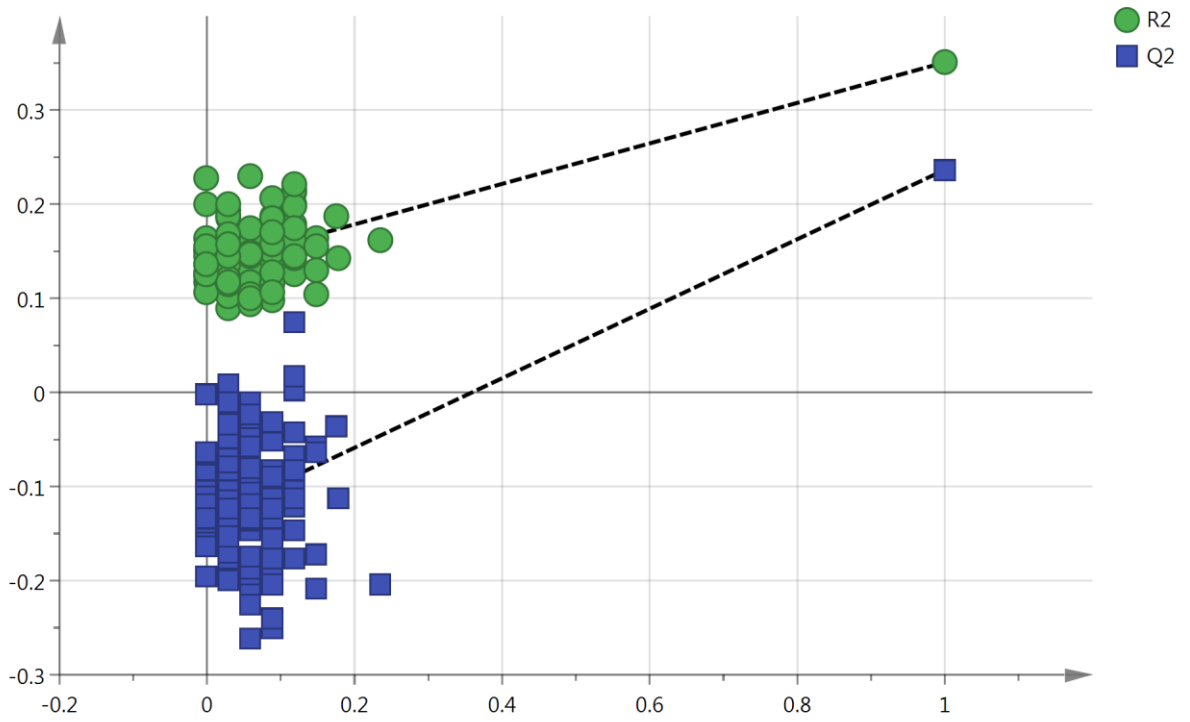



Figure 4.28 Permutation plot of 100 permutations of OPLS-DA model between EOC and 'Normal' groups in the $^{1}$H NOESY NMR study of plasma samples.

To further elucidate the differences in metabolites of case vs. control the influential loadings responsible for spectral differences from the supervised OPLS-DA was produced, Figure 4.29.

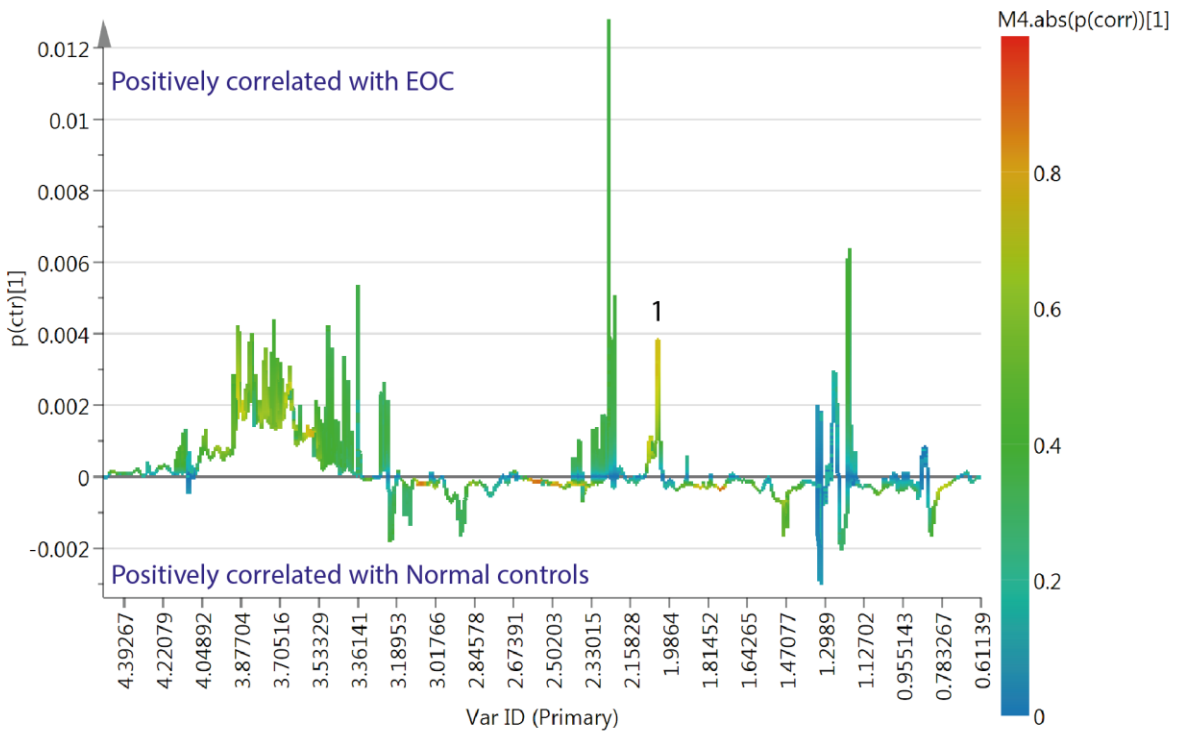



Figure 4.29 Correlation coefficient plot corresponding to $^{1}$H NOESY NMR OPLS-DA model of EOC vs. Normal controls of plasma samples. OPLS-DA coefficient plot labels: **1**, glycoprotein acetyls (one predictive component plus two orthogonal components; R2Y=0.351, Q2Y=0.237, R2X=0.614). Figure highlights spectral region between 0.6ppm to 4.40ppm only.

The OPLS-DA correlation plot allows identification of metabolites that correlate with a particular class. The colour coding refers to the magnitude of correlation and the direction of the peak indicates to which class the variable is positively correlated. In Figure 4.29 the peaks pointing upwards are positively correlated in EOC and the peaks pointing downwards are positively correlated in plasma of control patients.

As previously seen the chemical shift resonance at 2.04ppm corresponding to glycoprotein acetyls is influential in the spectral differences between the EOC and normal groups in the $^{1}$H CPMG and $^{1}$H NOESY NMR spectra.

To assess the performance of the global NMR profile in the form of the supervised OPLS-DA model in distinguishing patients with EOC from those without ROC analysis was performed, Figure 4.30.

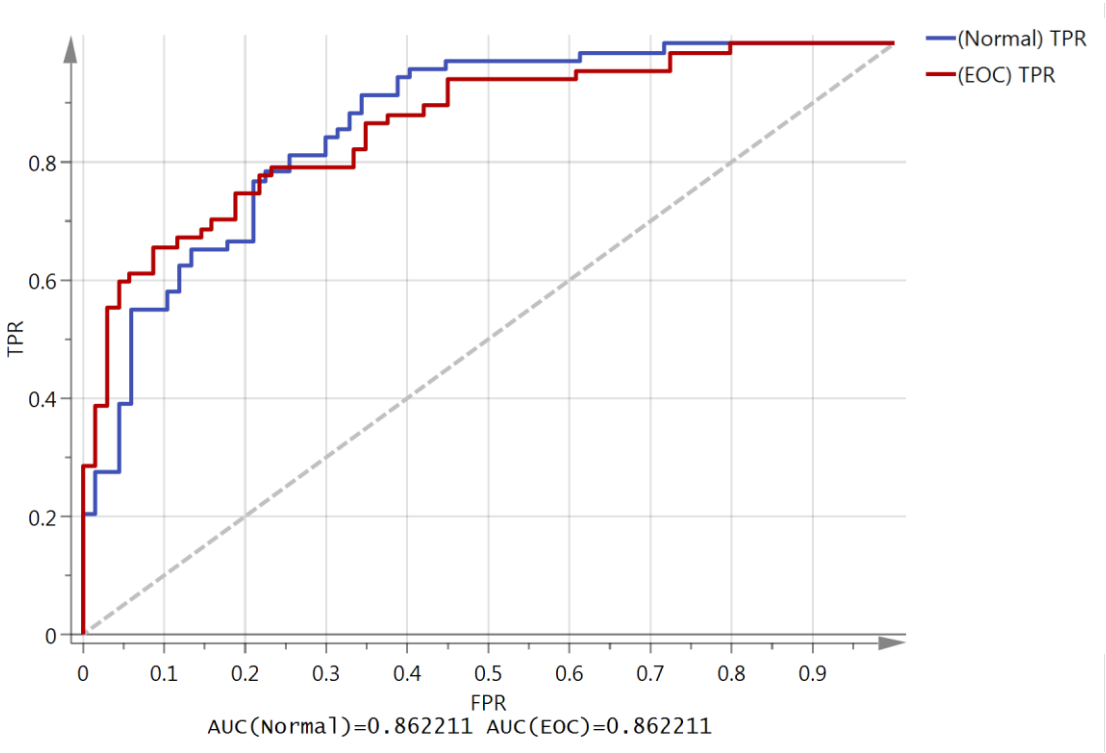



Figure 4.30 ROC curve for 'Normal' versus EOC by metabolic signature based on $^{1}$H NMR NOESY data. Note: each curve is a mirror image of the other and therefore AUC represents the ability of the metabolic signature to discriminate between the two groups.

### 4.3.4.2 Univariate data analysis of peak integrals of $^{1}$H NOESY NMR spectra of plasma samples

Visual inspection of the averaged spectra (Figure 4.22) and correlation coefficients loading plot (Figure 4.29) suggested differences in resonances at 0.84ppm, 2.04ppm and 3.22ppm. Four individual peaks were integrated from the NOESY $^{1}$H-NMR spectra for each plasma sample to investigate the phenotype-associated changes in these metabolites. The chemical shift, upper and lower boundaries of the peaks and the tentative metabolite assignment are shown in Table 4.6. The %CV in the IQC samples for the 4 metabolites is shown in Table 4.7.

Table 4.6 Table of peaks from NOESY $^{1}$H NMR spectra of plasma samples that were integrated, the chemical shift and upper and lower boundaries for integration.

| **Metabolite** | **Resonance chemical shift (ppm)** | **Peak lower bound (ppm)** | **Peak upper bound (ppm)** |
|---|---|---|---|
| Lipid resonances 0.85pm | 0.85 | 0.80 | 0.9 |
| Choline phospholipids | 3.215 | 3.202 | 3.228 |
| Unsaturated lipid | 5.3 | 5.25 | 5.35 |
| Glycoprotein acetyls | 2.0425 | 2.035 | 2.05 |

Table 4.7 Table of the %CV values for the peak integration in the NOESY $^{1}$H NMR spectra of the QC samples.

| **Metabolite** | **%CV in SR (n=8)** | **%CV in LTR (n=8)** |
|---|---|---|
| Lipid resonances 0.85pm | 2.61 | 3.11 |
| Choline phospholipids | 3.48 | 5.67 |
| Unsaturated lipid | 3.53 | 2.88 |
| Glycoprotein acetyls | 3.69 | 4.10 |

The Wilcoxon rank sum test was used to investigate which peak integrals from the $^{1}$H NOESY NMR spectra were significantly different between both the EOC group and the Normal group. A total of 4 peak integrals were investigated in EOC (n=62), benign (n=50) and healthy groups (n=14). Duplicate samples were removed to ensure each sample was from a unique patient. The p-values from the Wilcoxon-rank sum test were FDR adjusted using the Benjamini-Hochberg method and an adjusted P-value <0.05 was considered significant.

A total of 2 metabolites were significantly different between the EOC and Normal groups including choline phospholipids and glycoprotein acetyls supporting the visual differences seen in the spectra in Figure 4.22. Box and whisker plots for these two metabolites are shown in Figure 4.31.

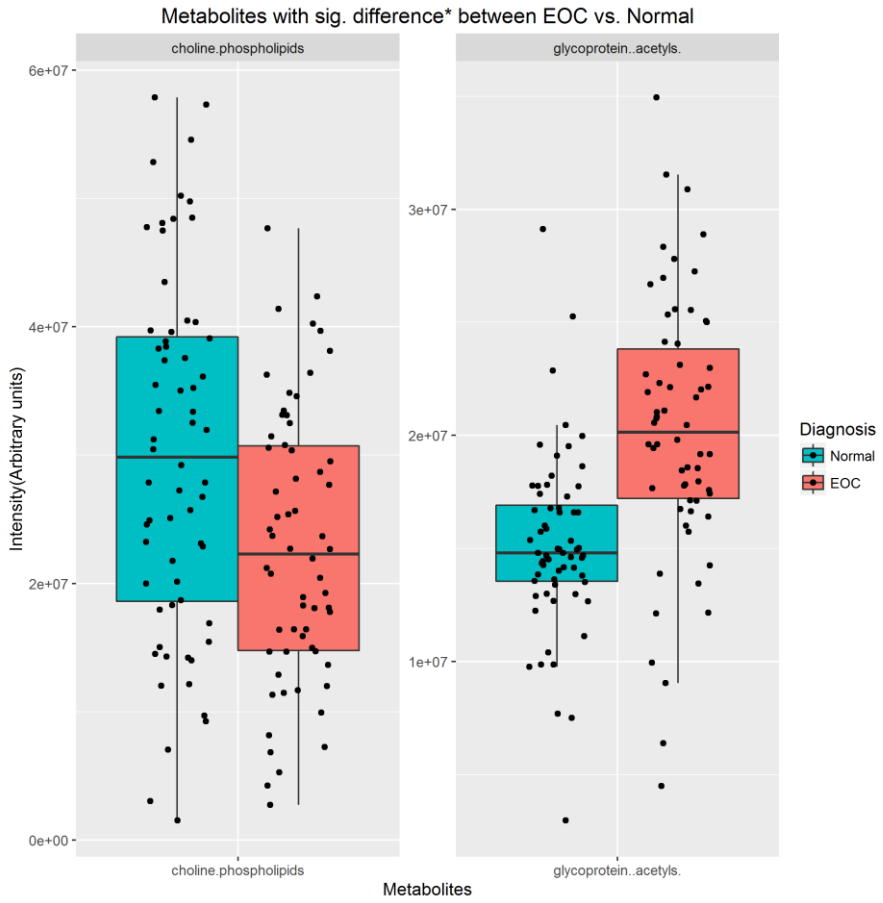



Figure 4.31 Metabolite integrals from NOESY-presat-$^{1}$H NMR data that are significantly different between EOC and 'Normal' groups. *FDR adjusted p-value <0.05.

#### 4.3.4.2.1 Association of metabolites with grade and stage of disease

To further elucidate the potential role of these metabolites in the pathogenesis of the disease, rank correlation using kendall's tau was used to determine whether these metabolites are also significantly associated with stage or grade.

Rank correlation of the significantly different metabolites between the EOC and 'Normal' group showed that resonances corresponding to choline phospholipids was associated with stage of disease. Following the pattern seen between the EOC group and normal group; choline phospholipids showed a decreasing concentration based on stage of disease. Figure 4.32 shows the box and whisker plot with the scatter of samples dichotomised to high stage (III/IV) and low stage (I/II). Values for the benign/healthy samples are shown for comparison.

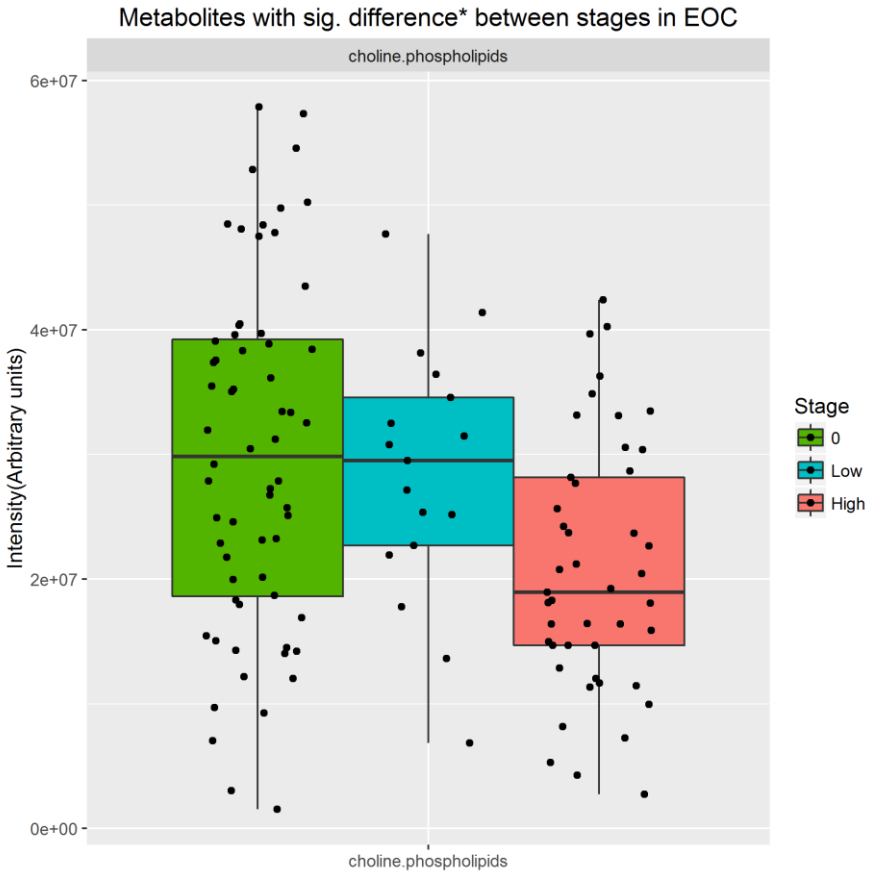



Figure 4.32 Box and whisker plots of metabolite integrals from NOESY-presat-$^{1}$H NMR data that show an association between stage (Low = Stage I,II and High = Stage III,IV) and metabolite integral, in the EOC group only. Values for samples with stage 0 (benign/healthy) are also shown for comparison.

There was no association between metabolite peak integrals from the NOESY-presat-$^{1}$H NMR of plasma and grade of disease in any of the significantly different metabolites between EOC and the normal group.

#### 4.3.4.3 Predictive capacity of metabolite integrals for diagnosis of EOC in plasma

To assess the performance of each of the significantly different metabolites in diagnosing EOC, ROC curves were produced and the AUC determined for each metabolite peak integral. The AUC based on metabolite peak integrals of choline phospholipids was 0.649 and for glycoprotein acetyls was 0.792. A combination of markers in a predictive model may further increase the predictive power of the metabolites and therefore a stepwise additive regression was used from known predictors to find the best model fit according to Akaike information criterion (AIC). As the model requires all metabolites to be independent of one another, Pearson correlation coefficient between the two metabolite was calculate and was r=0.0043740, suggesting the metabolites and not correlated.

As there was no prospectively collected patient cohort to test the model, the combined predictive model was internally validated using the Leave one out cross validation (LOOCV) method. Using the Leave one out Cross Validation (LOOCV) method based on metabolite integral of the 2 metabolites, 90 out of 126 (71%) samples were correctly predicted and all samples were classified, as there was no missing data for the peak integrals from $^{1}$H NOESY NMR.

A confusion matrix of the predictive ability of the model is shown in Table 4.8. The PPV = 73% and NPV = 70% and the sensitivity = 66% and specificity = 77%.

Table 4.8 Confusion matrix for the performance of the cross-validated step-wise logistic regression model for the 2 metabolite integrals from NOESY-presat-$^{1}$H NMR data with significantly different levels in plasma from patients with EOC and without.

| | | **State of disease** | |
|---|---|---|---|
| | | Disease + | Disease – |
| **Results of LOOCV model** | Test + | True positive<br>41 | False positive<br>15 |
| | Test – | False negative<br>21 | True negative<br>49 |

ROC analysis of the cross-validated step-wise logistic regression model produced an AUC of 0.812, see Figure 4.33.

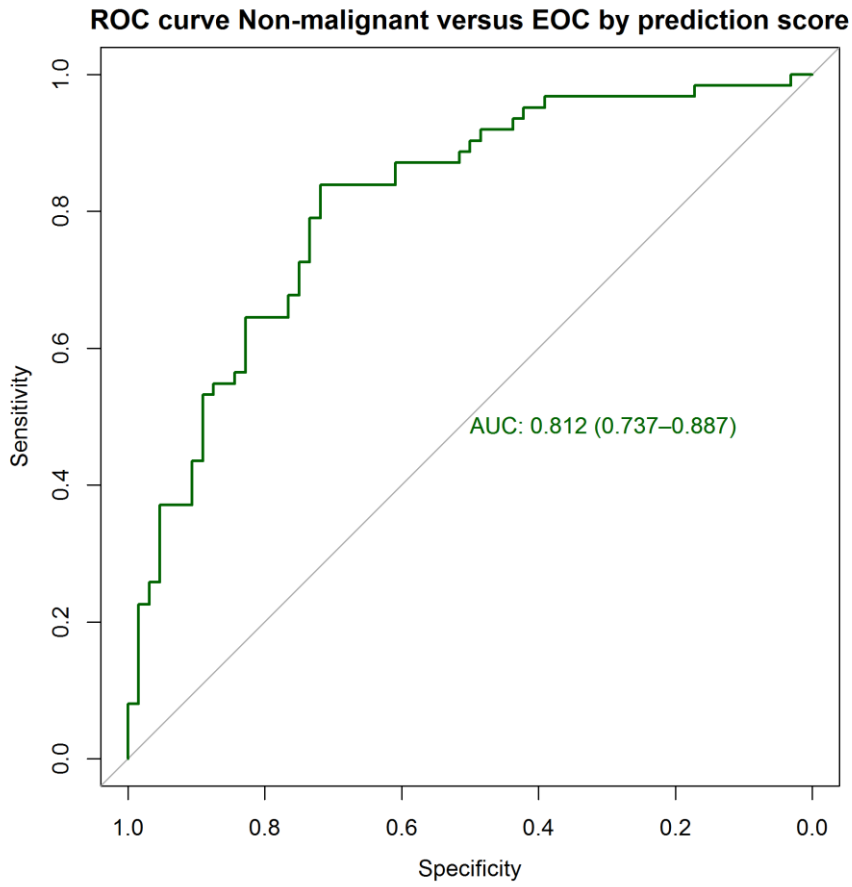



Figure 4.33 ROC curve for non-malignant versus EOC by cross-validated prediction score based on step-wise logistic regression model of 2 metabolites from NOESY-presat-$^{1}$H NMR data.

#### 4.3.4.4 Variation of selected metabolites in plasma

Biological replicates were taken from 14 patients at different time points by different people. To compare the replicates, concentrations of the predictive metabolites between the duplicate samples were compared. The reproducibility (precision) using the coefficient of variation (%CV) of each metabolite between each replicate is shown in Table 4.9. A prediction score for the biological replicates that were removed from the original analysis was also produced using the LOOCV model.

Table 4.9 Coefficient of variation (%) between biological replicates for significantly different metabolite integrals from NOESY-presat $^{1}$H-NMR data used in stepwise logistic regression model.

| Duplicate | Choline phospholipids (%CV) | Glycoprotein (acetyls) (%CV) | Time between sampling (days) | Class | Prediction score from cross validated model of (binary vector) | | Model performance | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Biological replicate included in model | Biological replicate excluded from model | Biological replicate included in model | Biological replicate excluded from model |
| D2 | 5.67 | 3.99 | 0 | Benign | 0.27 | 0.25 | TN | TN |
| D4 | 9.07 | 3.13 | 1 | Malignant | NA | NA | NA | NA |
| D5 | 12.70 | 9.93 | 0 | EOC | 0.61 | 0.49 | TP | FN |
| D7 | 1.76 | 2.77 | 0 | EOC | 0.67 | 0.58 | TP | TP |
| D8 | 1.37 | 0.12 | 1 | Benign | 0.42 | 0.42 | TN | TN |
| D10 | 1.02 | 2.56 | 1 | Borderline | NA | NA | NA | NA |
| D12 | 12.89 | 6.40 | 2 | Benign | 0.86 | 0.80 | FP | FP |
| D13 | 11.43 | 15.78 | 0 | Borderline | NA | NA | NA | NA |
| D14 | 15.74 | 8.68 | 0 | EOC | 0.46 | 0.77 | FN | TP |
| D15 | 4.89 | 0.28 | 0 | EOC | 0.50 | 0.46 | FN | FN |
| D17 | 1.21 | 1.25 | 0 | Benign | 0.36 | 0.37 | TN | TN |
| D18 | 58.92 | 47.68 | 5 | Borderline | NA | NA | NA | NA |
| D20 | 1.37 | 0.79 | 0 | EOC | 0.29 | 0.29 | FN | FN |
| D21 | 26.52 | 0.13 | 3 | Benign | 0.18 | 0.32 | TN | TN |

A prediction score for the biological replicates that were removed from the original analysis was produced using the LOOCV model. Comparison of the prediction scores across the duplicate samples revealed that the classification for two samples changed. The prediction score for D5 a sample from a patient with EOC was incorrectly classified compared to the analysis on the original sample and sample D14 again from a patient with EOC was correctly classified compared to the analysis of the original sample, see Table 4.9.

#### 4.3.4.5 Predicting survival in EOC patients to aid surgical stratification using $^{1}$H NOESY NMR plasma peak integrals

There was no association with overall or progression-free survival for either of the two metabolite peak integrals.

### *4.3.5 LipoFIT® analysis of lipoprotein subclasses in plasma samples from EOC patients and controls*

Lipoprotein subfractions were quantified using the *lipo*FIT® test: a software based analysis of NMR spectra (numaresHEALTH, Regensburg, Germany).

#### 4.3.5.1 Univariate data analysis of lipoprotein sub-fractions of NMR spectra from plasma samples

Data for all plasma and IQC samples analysed by $^{1}$H NMR was received from numaresHEALTH, Germany. The nomenclature of the lipoprotein subfractions is as follows: the first 2 letters depict the matrix e.g. TP – Total plasma, VL – VLDL, ID – IDL, LD – LDL and HD – HDL and the last two letters depict the analyte e.g. TG – triglycerides, CH – cholesterol, FC – free cholesterol, AB – Apo-B, PL – phospholipids. A complete list of the lipoprotein subfractions can be found in Appendix 2.

The Wilcoxon rank sum test was used to identify the lipoprotein subfractions from the NMR spectra that were significantly different between the EOC and the Normal group. A total of 105 lipoprotein subfractions were calculated in EOC (n=62), benign (n=50) and healthy groups (n=14). Duplicate samples were removed to ensure each sample was from a unique patient. The p-values from the Wilcoxon-rank sum test were FDR adjusted using the Benjamini-Hochberg method and an adjusted p-value $<0.05$ was considered significant.

A total of 35 lipoprotein subfractions were significantly different between the EOC and Normal groups including LDL-triglycerides, HDL-free cholesterol and total plasma ApoA1 and ApoA2. Box and whisker plots for these subfractions are shown in Figure 4.34.

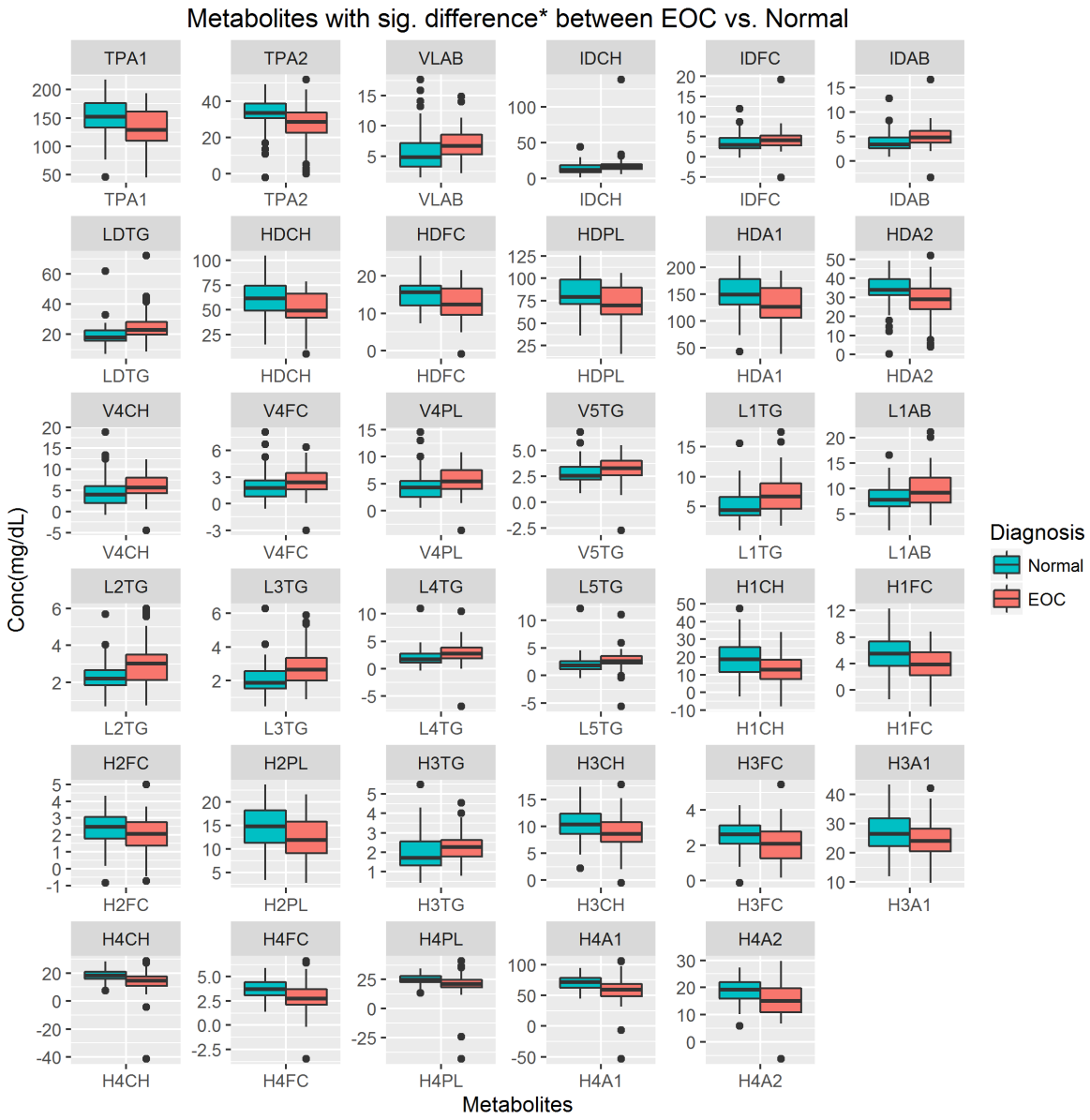



Figure 4.34 Metabolite integrals from lipoprotein sub-fractions data that are significantly different between EOC and 'Normal' groups. *FDR adjusted p-value <0.05.

#### 4.3.5.1.1 Association of metabolites with grade and stage of disease

To further elucidate the potential role of these metabolites, rank correlation using kendall's tau was used to determine whether these metabolites are also significantly associated with stage or grade.

Rank correlation of the significantly different metabolites between the EOC and 'Normal' group showed that lipoprotein subfractions: TPA2, LDTG, HDA2, L1TG, L3TG, H2FC, H3FC, H4A2 were associated with stage of disease. The box and whisker plots with associated scatter of the samples are shown in Figure 4.35.

In addition logistic regression analysis of stage (dichotomised into high (Stage III/IV) and low (Stage I/II) also showed a significant difference in lipoprotein subfractions of HDFC, L2TG and H2PL where higher stage disease was associated with higher values of L2TG and lower values of HDFC andH2PL.

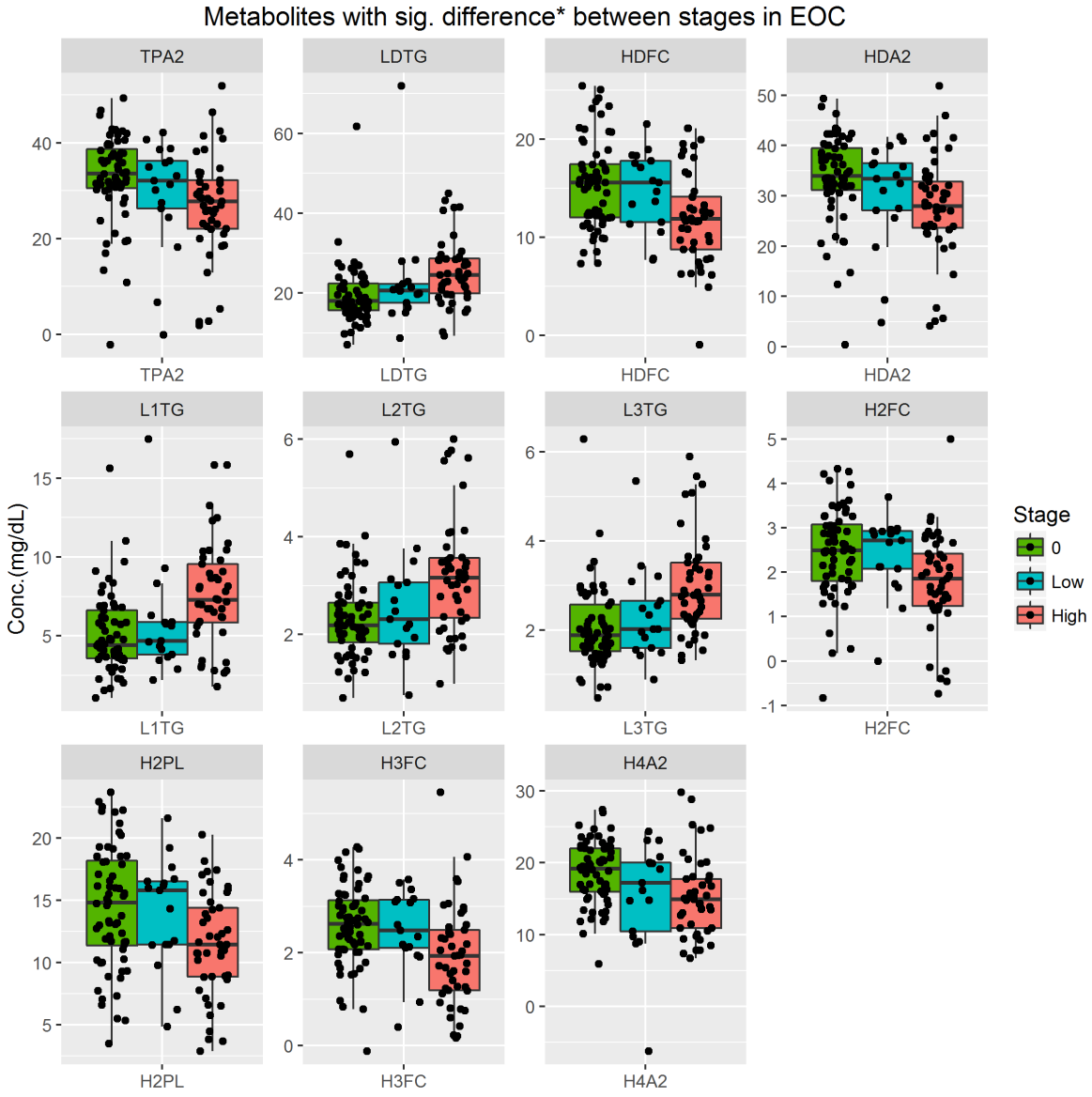



Figure 4.35 Box and whisker plots of lipoprotein sub-fractions data that show an association between stage (Low = Stage I,II and High = Stage III,IV) and metabolite integral, in the EOC group only. Values are also shown for samples with stage 0 (benign/healthy) for comparison.

There was no association between lipoprotein subfractions and grade of disease in any of the significantly different subfractions between EOC and the normal group.

4.3.5.2 Predictive capacity of lipoprotein sub-fractions for diagnosis of EOC in plasma

To assess the performance of each of the significantly different metabolites in diagnosing EOC, ROC curves were produced and the AUC determined for each lipoprotein subfraction. Amongst the highest AUC were based on lipoprotein subfractions of L3TG (AUC=0.721), H4CH (AUC=0.720) and IDAB (AUC=0.969). The %CV for these 3 markers in the SR and LTR IQCs in the experiment were all <5%. A combination of markers in a predictive model may further increase the predictive power of the metabolites and therefore a stepwise additive regression was used from known predictors to find the best model fit according to Akaike information criterion (AIC). The model requires all metabolites to be independent of one another and therefore Pearson correlation between each metabolite was determined, see Figure 4.36.

Pearson's correlation coefficient was calculated for the subfractions of each of the 35 metabolites that were significantly different between samples from patients with EOC compared to healthy and benign patients against the others. Only metabolites with an $R^2$ <0.7 compared to the other metabolites were included.

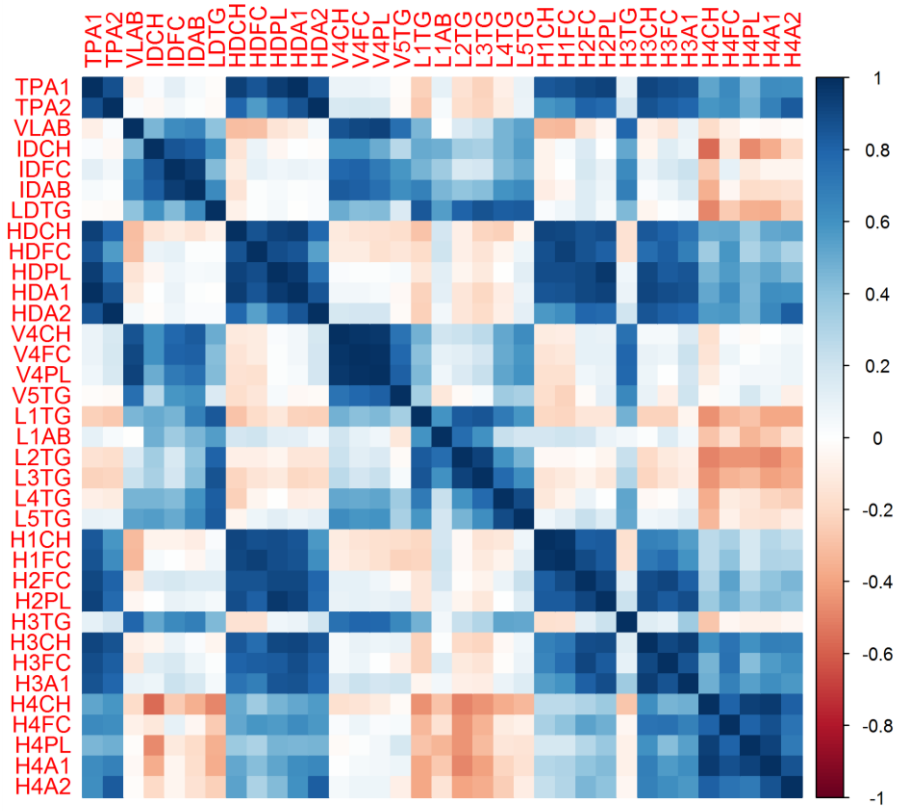



Figure 4.36 Correlation plot of the 35 lipoprotein sub-fractions that were significantly different between the EOC and 'Normal' groups.

The correlation plot in Figure 4.36 highlights that many of the lipoprotein subfractions are correlated, which is unsurprising as many of the subfractions are different lipids related to a specific lipoprotein particle e.g. LDL. The concentration of the following uncorrelated subfractions: L3TG, H4CH, IDAB, VLAB and HDA1 were combined in a step-wise logistic regression model to produce a final model consisting of L3TG and H4CH. This step-wise logistic regression model produced an AUC of 0.766 and correctly classified 73% of samples. However cross validation using the LOOCV method reduced the AUC to 0.610 and correctly classified only 56% of samples, suggesting the original model based on concentration of L3TG and H4CH was not robust for predicting EOC in plasma samples.

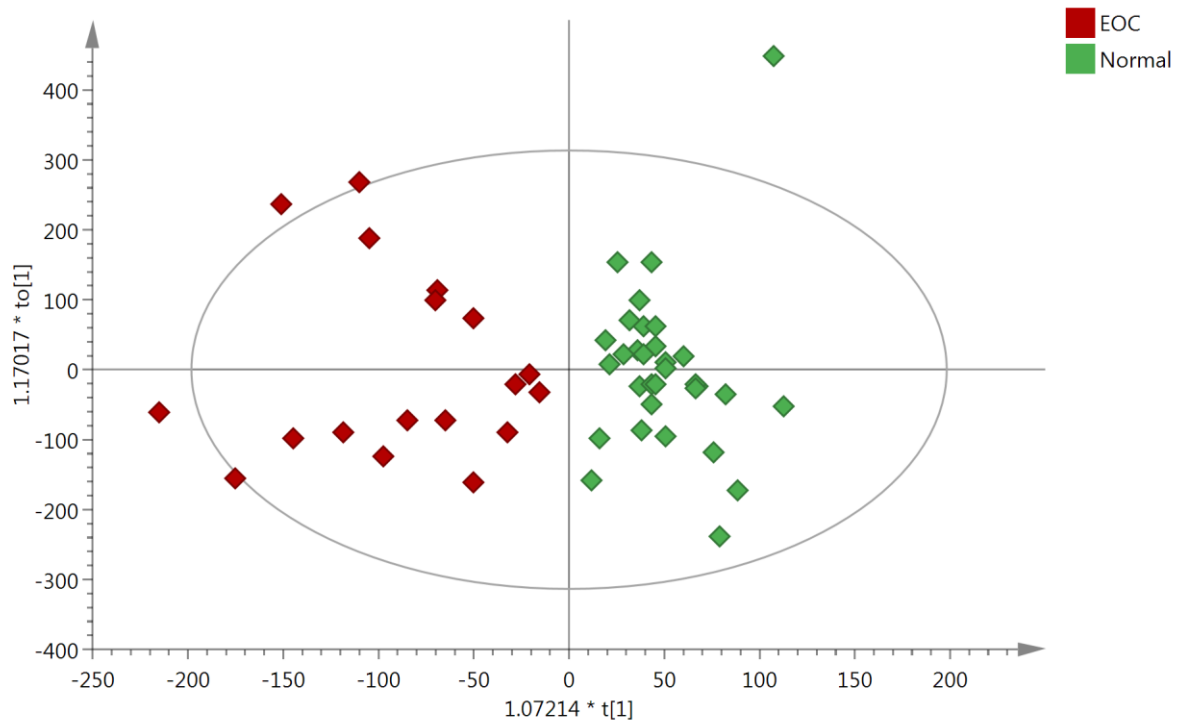
#### 4.3.5.3 Predicting survival in EOC patients to aid surgical stratification using lipoprotein sub-fraction concentration

In order to identify metabolites in blood plasma that are associated with survival, data from lipoFIT analysis of the NMR spectra was used from all patients diagnosed with epithelial ovarian cancer. The nonparametric logrank test was used to identify if patients had different survival distributions above and below the median of each metabolite concentration that was previously shown to have a significantly different concentration between EOC and benign/healthy controls in the initial phase of this study. Total plasma Apo-A1 concentration was associated with both overall and progression free survival in patients with epithelial ovarian cancer, however, this significance was lost when adjusting for confounding factors including age, stage, grade and presence of residual disease (none vs. >1mm). There were no lipoprotein subfractions that predicted survival when adjusting for these confounding factors.

### *4.3.6 $^{1}$H NOESY NMR analysis of urine samples from EOC patients and controls*

#### 4.3.6.1 Multivariate data analysis of NOESY-presat-$^{1}$H NMR spectra of urine and quality control samples

A small number of urine samples were also collected from patients undergoing gynaecological surgery at Hammersmith Hospital. Eighteen EOC samples, 23 benign samples and 9 healthy control samples were analysed using the NOESY pre-saturation sequence. Due to the larger fluctuation in pH in urine, spectra were aligned and then normalised. An OPLS-DA model for EOC compared to the Normal group was produced and the plot is shown in Figure 4.37.




R2X[1] = 0.0482 R2Xo[1] = 0.121 Ellipse: Hotelling's T2 (95%)

Figure 4.37 Supervised OPLS-DA scores plot of urine samples from patients with EOC (n=18) and 'Normal' group (n=32) combining patients with benign gynaecological disease and healthy volunteers in the $^{1}$H NMR NOESY data.

There appears to be complete separation between the two groups in the OPLS-DA model. To assess the validity of the supervised model, a permutation plot was produced, Figure 4.38.

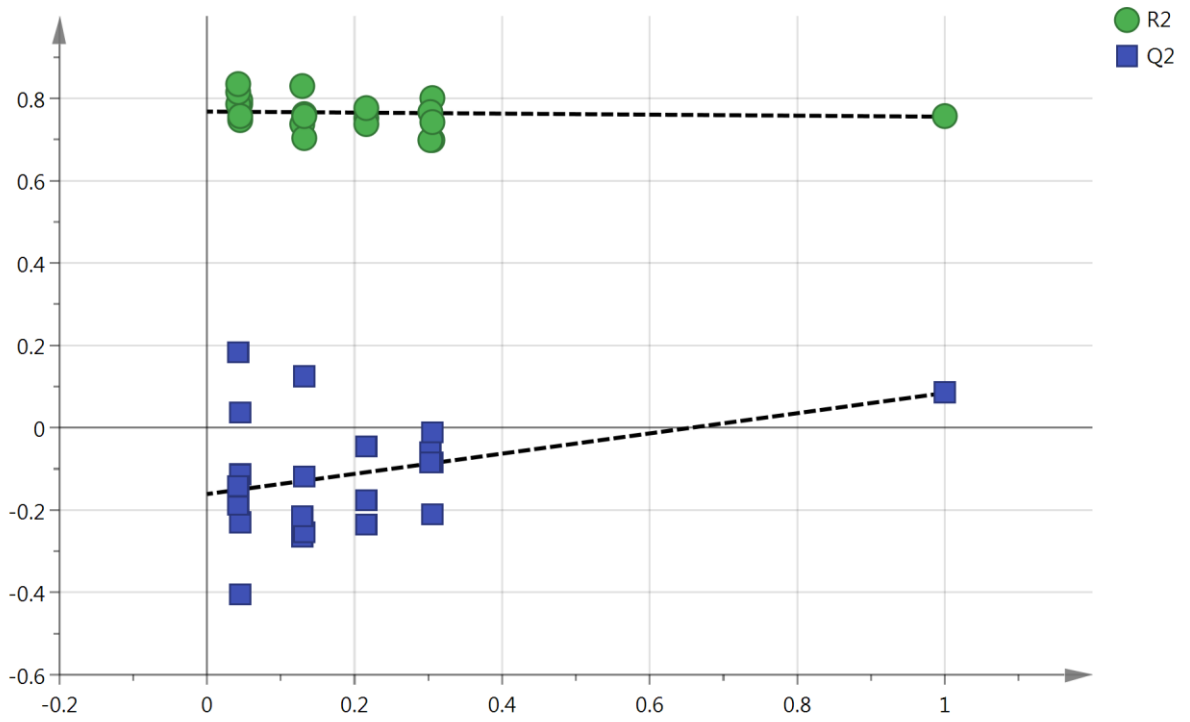



Figure 4.38 Permutation plot of 20 permutations of the OPLS-DA model between the EOC and 'Normal' groups in the $^{1}$H NOESY NMR data of urine samples.

The permutations plot (n=20) suggests that the model is spurious, i.e. it just fits the training set well but does not predict Y well for new observations.

The plot indicates that the model is not valid as not all Q2 (blue) values to the left are lower than the original points to the right and the same for the R2 meaning when Y is permuted the goodness of fit is better than the original model indicating the original model is not valid.

## 4.4 Discussion

### *4.4.1 $^{1}$H NMR metabolic signatures in the diagnosis of epithelial ovarian cancer*

Metabolic signatures were generated for plasma and urine samples from patients with ovarian and other gynaecological cancers and benign/healthy controls using $^{1}$H NMR with two pre-saturation techniques to provide complementary information. The global metabolic profile was analysed using multivariate data analysis techniques and peak integrals were analysed using univariate statistical analysis techniques. Although there was little clustering of samples based on class, together the discriminatory performance of the global metabolic profile generated an AUC of 0.868 for the CPMG spectra and 0.862 for the NOESY spectra.

Univariate analysis of the peak integrals, with correction for multiple testing, identified 10/30 metabolites in the CPMG spectra and 2/4 metabolites in the NOESY spectra that were significantly different between the EOC group and the Normal group; Wilcoxon rank sum test, FDR adjusted p-value <0.05. These predictors were then incorporated into a leave one out cross-validated stepwise logistic regression model to find the best model fit for diagnosis or prediction of EOC according to AIC using only independent predictors (Pearson correlation coefficient <0.7). A combination of 3 integrals from the CPMG spectra corresponding to histidine, unknown peak at 1.81ppm and at 2.39ppm yielded an AUC of 0.819 and a combination of the 2 metabolites from the NOESY spectra corresponding to choline phospholipids and glycoprotein acetyl groups yielded an AUC of 0.812.

It should be noted here that the AUC for the multivariate models contains duplicate samples from the same patient taken at different time points by different people (n=14) therefore the AUC is probably slightly overestimated. Analysis of the peak integrals in a stepwise logistic regression model showed that the majority of duplicate samples classifications were the same however there were 2 samples in the NOESY spectra that changed classification; one correctly and one incorrectly and 1 sample in the CPMG spectra that changed to the correct classification. This suggests that there may be biological variation between metabolites which might be expected as each sample reflects a cross section of all the metabolic processes occurring across the entire system. Alternatively, it might reflect systematic error or bias introduced by sampling and pre-processing due to stability of metabolites after sample collection and the degradation of metabolites due to the action of cells within the sample (Yin *et al.*, 2015). Either way, there is a need for robust sample collection to avoid poor sample quality and introduction of bias with appropriate record keeping such as conditions and time to centrifugation and storage.

Supervised analysis of $^{1}$H NMR NOESY spectra of a small number of urine samples showed that the model produced was not valid and therefore any further analysis would yield spurious results. Future studies exploring possible biomarkers of EOC in urine would require a larger number of samples to generate valid models.

#### 4.4.1.1 Variation in biological replicates within the study

In the univariate analysis of the peak integrals from the CPMG and NOESY spectra, 14 biological replicates were also analysed and the precision between the duplicates assessed. There are two possible explanations for the variation, either biological or analytical. As the duplicates were taken by two different people at two different times and no time to centrifugation or storage is recorded it is impossible to say what the likely cause is. However, it is important to find out whether the changes in the classification in the model are due to biological or analytical variation. This could

simply be done by performing the same experiment on multiple samples on patients across a number of days by the same person and keeping the time to centrifugation and storage consistent to elucidate whether biological variation is seen in the candidate metabolic biomarkers or it is analytical and related to the pre-analytical factors.

### *4.4.2 Beta-hydroxybutyrate as a diagnostic and prognostic biomarker of EOC*

Visual inspection of the averaged $^{1}$H CPMG NMR spectra and supervised multivariate analysis suggested a difference in the amount of beta-hydroxbutyrate, acetone and acetoacetate in patients with EOC compared to control subjects. Univariate analysis of the peak integral of one of the peaks corresponding to beta-hydroxybutyrate at 1.2ppm showed this metabolite to be significantly higher in patients with EOC. This is consistent with previous studies that identified increased levels of 3-hydroxybutyrate in the serum of patients with EOC (Odunsi *et al.*, 2005, Garcia *et al.*, 2011, Hilvo *et al.*, 2016) detailed in Chapter 1: Table 1.2. Increased levels were also seen in tumour tissue of patients with EOC (Hilvo *et al.*, 2016, Fong *et al.*, 2011) where the latter identified an 8.63 fold increase in 3-hydroxybutyrate in metastatic EOC tissue compared to normal subjects (Fong *et al.*, 2011). In addition, increased levels of other hydroxybutyrate (HBA) metabolites have been reported in EOC including: 2-hydroxybutyrate in primary and metastatic ovarian cancer tissue (Fong *et al.*, 2011) and 3,4-dihydroxybutyric and 2,4-dihydroxybutyric acid concentrations in serum of EOC patients (Hilvo *et al.*, 2016). They hypothesise that the observed accumulation in HBAs is likely to arise from the deregulation of multiple metabolic pathways including up regulation of fatty acid oxidation as an alternative energy source which has previously been described in ovarian cancer tumours (Fong *et al.*, 2011) and in metastasis (Nieman *et al.*, 2011).

Statistical total correlation spectroscopy (STOCSY) analysis in this study identified numerous other peaks that are correlated to the peak at 1.2ppm including resonances corresponding to other ketone body metabolites; acetone and acetoacetate. 3-hydroxybutyrate, acetone and acetoacetate were also found to be upregulated in EOC compared to controls in a study using $^{1}$H NMR (Garcia *et al.*, 2011) supporting this finding in the present study.

On retrospective analysis of the integral peaks from the $^{1}$H CPMG data, the correlation coefficient plot from the OPLS-DA model and the STOCSY analysis, the peak integral in Section 4.3.2.2 termed Unknown.2.39ppm is thought to be one of the peaks of a doublet of doublets corresponding to beta-hydroxybutyrate. This peak is also significantly higher in patients with EOC, correlated to the peak at 1.2ppm and correlated with increasing stage of disease.

Survival analysis of the peak integrals that were significantly different between case and control groups identified that beta-hydroxybutyrate at 1.2ppm is associated with overall survival in patients

with EOC. In addition, the unknown peak at 2.39ppm, now identified as a peak corresponding to beta-hydroxybutyrate, was shown to be associated with progression free survival and this is independent of confounding factors such as age, stage and presence of residual disease.

A recent study from Hilvo *et al*. reported metabolites of gamma-hydroxybutyrate (GHB) to be associated with tumour burden and patient survival (Hilvo *et al.*, 2016).The authors noted that accumulation of HBA metabolites, specifically dihydroxyutyric acid from the metabolism of GHB occurs in patients affected by SSADH deficiency caused by mutations in the *ALDH5A1* gene and showed that low mRNA expression of this gene was found to be a significant prognostic biomarker for HGSOC in a number of datasets including TCGA, suggesting that the prognostic signature observed in the study is linked to the downregulation of *ALDH5A1* causing a lowered activity of the SSADH enzyme. Furthermore, gene set enrichment analysis (GSEA) revealed a correlation between the loss of the *ALDH5A1* gene and the presence of epithelial-mesenchymal transition (EMT), another hallmark of cancer, associated with invasion and metastasis (Hanahan and Weinberg, 2011). This link is only correlative however and further work is required to elucidate the molecular link between energy metabolism, SSADH deficiency and EMT.

### *4.4.3 Altered amino acid metabolism in ovarian cancer*

The amino acids, alanine and histidine were significantly lower in patients with EOC relative to the control group. Correlation and logistic regression analysis also identified these amino acids to be associated with stage of disease. Our results are in agreement with other publications reporting decreases in various amino acids in blood of cancer patients relative to controls. Slupsky et al., using $^{1}$H NMR of urine showed a decrease in alanine in patients with ovarian and breast cancers relative to healthy controls (Slupsky *et al.*, 2010). Garcia *et al*., showed a decrease in concentrations of alanine in early stage EOC relative to controls in serum (Garcia *et al.*, 2011)and Tiziani *et al*., showed a decrease in alanine in blood samples of patients with oral cancer compared to controls using $^{1}$H NMR (Tiziani *et al.*, 2009). Elevated levels of metabolites such as histamine and alanine were found to be elevated in tumour tissue compared to normal tissue in squamous cell head and neck carcinomas (Mukherji *et al.*, 1997) suggesting that this change in concentration in the blood may reflect a shift in energy production and uptake of these amino acids from the blood into the tissue as alternative substrates for gluconeogenesis to ultimately feed into the TCA cycle. However, it is difficult to draw conclusions relating to perturbations of specific pathways as metabolic profiling of blood is merely a snapshot of multiple metabolic events occurring within an organism at one point in time whereas labelled flux analysis of metabolites could better determine this by monitoring production and consumption rates of metabolites within a biological system.

Visual inspection of the averaged $^{1}$H CPMG plasma spectra and the correlation coefficient plot generated from the OPLS-DA model in both CPMG and NOESY spectra all identify that the peak resonance at 2.04ppm, was consistently found to be in differential abundance in cases compared to controls. Univariate analysis of this peak in the NOESY spectra found a significantly higher abundance in EOC compared to the control group. This is consistent with another study using $^{1}$H NMR that found this region to be one of the most influential regions in the spectra for distinguishing case from control, which they were able to do with 100% sensitivity and specificity (Odunsi *et al.*, 2005). A study by Kolwijck *et al*., aimed to investigate the identity of this unassigned and prominent resonance in this region (δ2.0-2.1ppm), frequently found in the NMR spectra of cancer patients. They concluded that both N-acetylaspartate (NAA) and N-acetyl groups from glycoproteins and/or glycolipids may contribute to the resonance seen at δ2.0-2.1ppm (Kolwijck *et al.*, 2009).

### *4.4.4 Altered lipid metabolism in ovarian cancer*

Another peak integral from the $^{1}$H NOESY plasma spectra that was found to be significantly different between EOC and the control group was the peak at 3.215ppm corresponding to choline phospholipids. This was found to be higher in the normal controls compared to patients with EOC. The choline moiety of phospholipids was also shown to be down regulated in a study using $^{1}$H NMR in ovarian cancer (Garcia *et al.*, 2011). This metabolite was also negatively associated with stage but there was no association with survival for any of the metabolites from the $^{1}$H NOESY spectra that were assessed.

Visual inspection of the peak at 0.84ppm in the averaged NOESY spectra suggested that there may be differences in resonances relating to lipids/lipoproteins in this region; however integral analysis of this broad peak was not significant in univariate analysis. This is perhaps not surprising as this peak is a broad peak with contribution from many subclasses of lipoproteins. LipoFIT analysis of the lipoprotein subfractions further elucidated the differences in lipoprotein subclasses between the two groups and suggests that LDL and VLDL subclasses and associated lipids are higher in patients with EOC compared to controls whereas HDL subclasses and associated lipids are higher in the control group. Total plasma ApoA1 and ApoA2 levels were also lower in EOC with levels of ApoA2 negatively associated with stage of disease. Survival analysis showed that Total plasma ApoA1 was associated with overall and progression free survival however this association was lost after adjusting for confounding factors, such as presence of residual disease. Interestingly, plasma ApoA1 is one of the tests used in both the OVA1 and Overa® multivariable index assays for diagnosis of ovarian cancer in the US.

Combination of these markers in a cross-validated stepwise logistic regression for diagnosis of EOC was not robust suggesting that these subfractions may not have utility as diagnostic or prognostic biomarkers but may provide potential insights into differences in lipid/lipoprotein metabolism in patients with EOC.

It is important to recognise that the levels of lipoproteins, lipids and ketones can be affected by co-morbidities such as diabetes, heart disease and fasting. Ketone bodies in particular are elevated during periods of fasting. In this study, fasting status was not available for the samples collected so a systematic bias in the sample collection cannot be ruled out between the patients with malignant and benign disease and healthy controls, highlighting the need for robust bio-banking procedures for metabonomic studies. However, Hilvo *et al*. showed a linear correlation of hydroxybutyric acid metabolites between tumour and serum samples supporting the hypothesis that this is due to altered metabolism of the tumour and not from another physiological process (Hilvo *et al.*, 2016).

### 4.5 Summary

The amino acids, alanine and histidine were shown to be lower in plasma of patients with EOC relative to the control group suggesting a shift in energy production and uptake of these amino acids into the tissue as alternative substrates for gluconeogenesis to feed into the TCA cycle.

Concentrations of N-acetylaspartate (NAA) and N-acetyl groups from glycoproteins at 2.04ppm are higher in patients with EOC compared to the control group. This may reflect proteins related to the immunological response in these patients.

Beta-hydroxybutyrate concentrations before surgery were found to be strong predictors of OS and PFS and can predict survival independently of the success of tumour reduction surgery.

Robust biobanking procedures with time to centrifugation and storage are required to minimise pre-analytical effects on metabolite concentration.

CA 125 is gold standard biomarker for detection of patients with advanced disease but there are contradictory results for its use as a prognostic biomarker. The utility of HBAs as disease biomarkers is in their prognostic power and future studies should aim to assess their potential in clinical use in isolation and in combination with other prognostic biomarkers.

# CHAPTER 5. Targeted MS-based quantitative profiling of plasma to identify metabolites associated with epithelial ovarian cancer

## 5.1 Introduction

This chapter focusses on identifying metabolites in plasma that are associated with disease in epithelial ovarian cancer patients. Metabolites that can discriminant patients with epithelial ovarian cancer from those with benign gynaecological disease and healthy unaffected volunteers will be investigated using a commercially available quantitative mass-spectrometry (MS) based assay, the Biocrates Absolute*IDQ*® p150 Kit. Metabolites associated with particular characteristics of the disease including stage, grade and prognosis will also be investigated.

## 5.2 Background

The advantage of MS-based metabonomics over NMR is its sensitivity. NMR however is extremely reproducible (Dona *et al.*, 2014) so the two techniques are complementary. Targeted MS based analysis using a kit based assay removes the requirement for feature selection and identification, which can be laborious and time-consuming, and produces absolute quantities of metabolites.

Plasma samples were analysed in the National Phenome Centre using the Biocrates Absolute*IDQ*® p150 Kit, a commercially available targeted metabolomics assay designed to be used with LC-MS/MS triple quadrupole instruments covering four compound classes including; acylcarnitines, amino acids, phospho – and sphingolipids and hexoses. A full list of the metabolites can be found in Appendix 2. Samples were analysed by flow injection analysis (FIA) coupled to tandem mass spectrometry (MS/MS) on Waters triple-quadrupole mass spectrometers. A one-point internal standard calibration using isotope-labelled internal standards ensures the metabolites can be absolutely quantified. Those metabolites referred to as 'semi-quantitative' had no commercial standards available and are therefore relative concentrations. The simultaneous derivatisation, extraction and quantification of metabolites can be fully automated in a 96-well plate format and requires as little as 10µl of plasma per injection. As well as the 3 plasma-based internal quality control (IQC) levels provided with each kit, the pooled study IQC samples; the sample reference (SR) and long term reference (LTR) were also analysed on each plate. A total of 310 samples including blanks, standards and IQCs were analysed over 4 plates as part of this study. A total of 208 individual patients and control subjects were included in the study and the clinical characteristics of these subjects are detailed in Table 5.1. Sixteen patients had duplicate samples and these are discussed in Section 5.3.4.

All samples were visualised using principal component analysis (PCA) to assess assay performance including clustering and intra-plate reproducibility of IQC samples.

## 5.3 Results

### *5.3.1 Characteristics of the patients in the study*

**Table 5.1** Clinical characteristics of participants used in the ovarian Biocrates study.

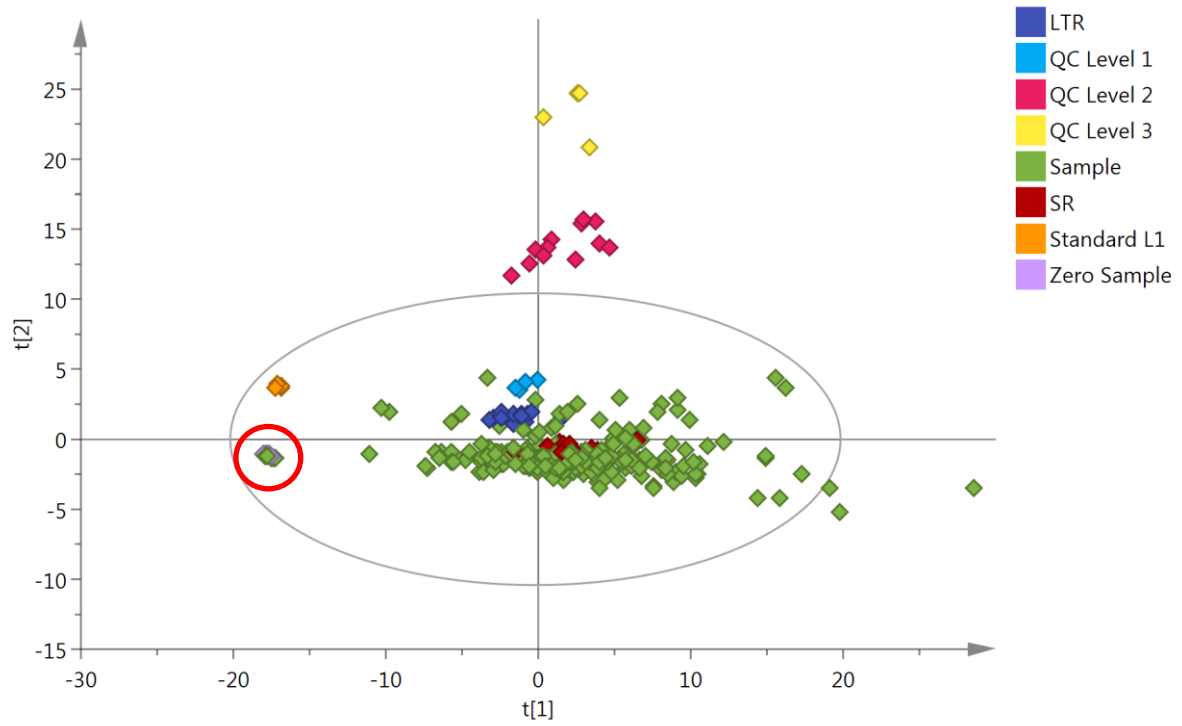
| | | Healthy (n=14) | Benign (n=55) | Borderline (n=19) | EOC (n=73) | Malignant (n=47) |
|---|---|---|---|---|---|---|
| Age | Mean (range) | 30 (22-43) | 51 (20-84) | 50 (25-72) | 60 (36-85) | 61 (32-83) |
| Grade | 1 | | | | 5 | 8 |
| | 2 | | | | 14 | 17 |
| | 3 | | | | 51 | 17 |
| | Undifferentiated | | | | 1 | 0 |
| | Unknown | | | | 2 | 5 |
| Stage | 1A,1B,1C | | | 8, 1, 4 | 6, 0, 7 | 13, 7, 2 |
| | 2A,2B,2C | | | 1 | 3, 2, 1 | 1, 2, 2 |
| | 3A | | | 1 | 1 | 1 |
| | 3B | | | | 8 | 1 |
| | 3C | | | | 29 | 7 |
| | 4 | | | | 16 | 3 |
| | Unknown | | | 4 | 0 | 8 |
| Histological subtype | Serous | | | 9 | 48 | 3 |
| | Mucinous | | | 9 | 3 | 2 |
| | Endometrioid | | | 1 | 4 | 27 |
| | Clear cell | | | | 10 | 2 |
| | Mixed | | | | 6 | 1 |
| | Carcinosarcoma | | | | 1 | 0 |
| | Unknown | | | | 1 | 12 |
| Residual disease | 0mm | | | | 27 | 3 |
| | 1-10mm | | | | 22 | 0 |
| | >11mm | | | | 18 | 3 |
| | Unknown | | | | 6 | 41 |

### *5.3.2 Analysis of plasma samples from EOC patients and controls using the Biocrates AbsoluteIDQ® p150 Kit*

Plasma samples were processed using the Met*IDQ*™ software provided with the Biocrates Absolute*IDQ*® p150 Kit. The Met*Val* module performed an automatic quality assessment of the data obtained. Any values outside of those defined in the operating procedure where the internal standard or standard/IQC was out of range were removed from the final analysis. In addition, any metabolite with a %CV >30% in any of the IQC groups was removed. In total, 36/158 metabolites were removed due to large variation in the quality control samples.

### 5.3.2.1 Multivariable data analysis of metabolites of plasma and quality control samples

To assess the performance of the assay, all samples, standards and IQC samples (n=306) were visualised using principal components analysis where each sample is plotted in the 'metabolic space' according to the concentration of the metabolites within the sample, see Figure 5.1.

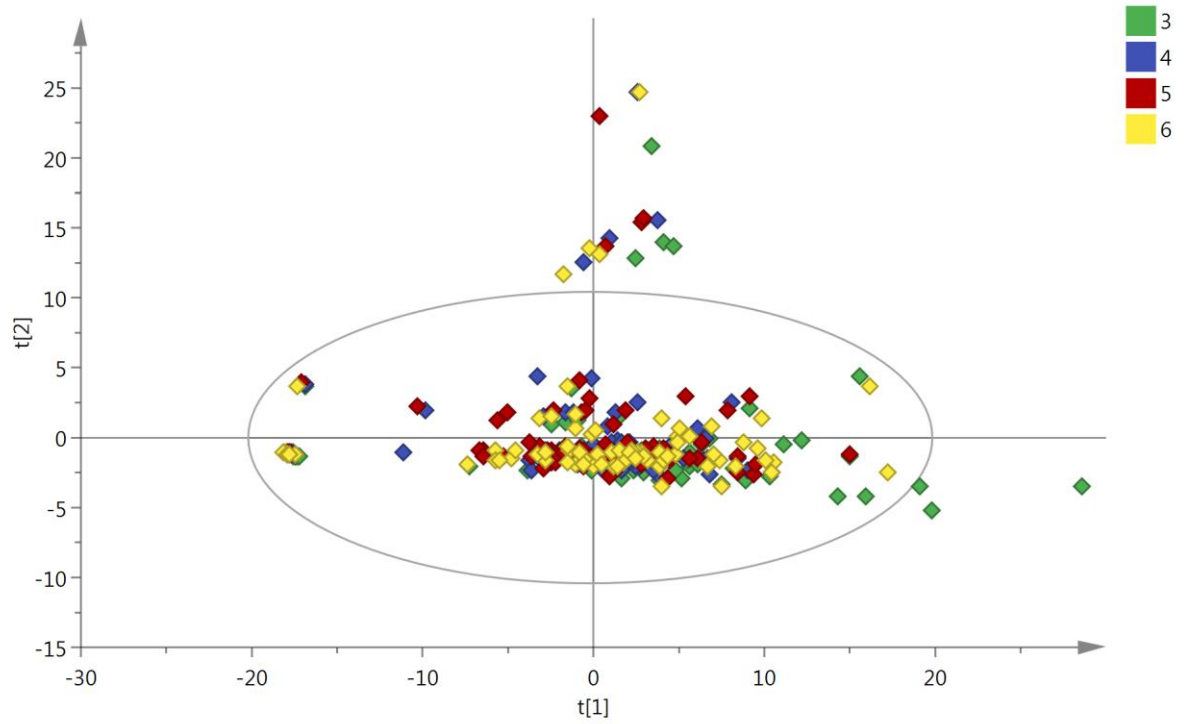



R2X[1] = 0.52 R2X[2] = 0.147 Ellipse: Hotelling's T2 (95%)

**Figure 5.1** Unsupervised PCA scores plot of all samples (n=246), standards (n=4), zero samples (n=12) and IQC samples (n=44) of the Biocrates data. LTR – Long Term Reference, QC Levels 1,2,3 – kit based assay QCs, SR – Sample Reference.

QCs and standards cluster reasonably well in the metabolic space in Figure 5.1 suggesting consistency of analysis across the run. A subset of samples (n=22) appear to cluster with the 'zero samples' (circled in red in Figure 5.1) which contain no biological sample suggesting they also contain no biological sample and were therefore removed as outliers.

Although the IQC samples appear to cluster well in Figure 5.1, there may be some consistent variability introduced as the samples were analysed across 4 plates. To assess whether there was any inter-plate variability, IQC and patient samples were also visualised using PCA but coloured by plate number, see Figure 5.2.

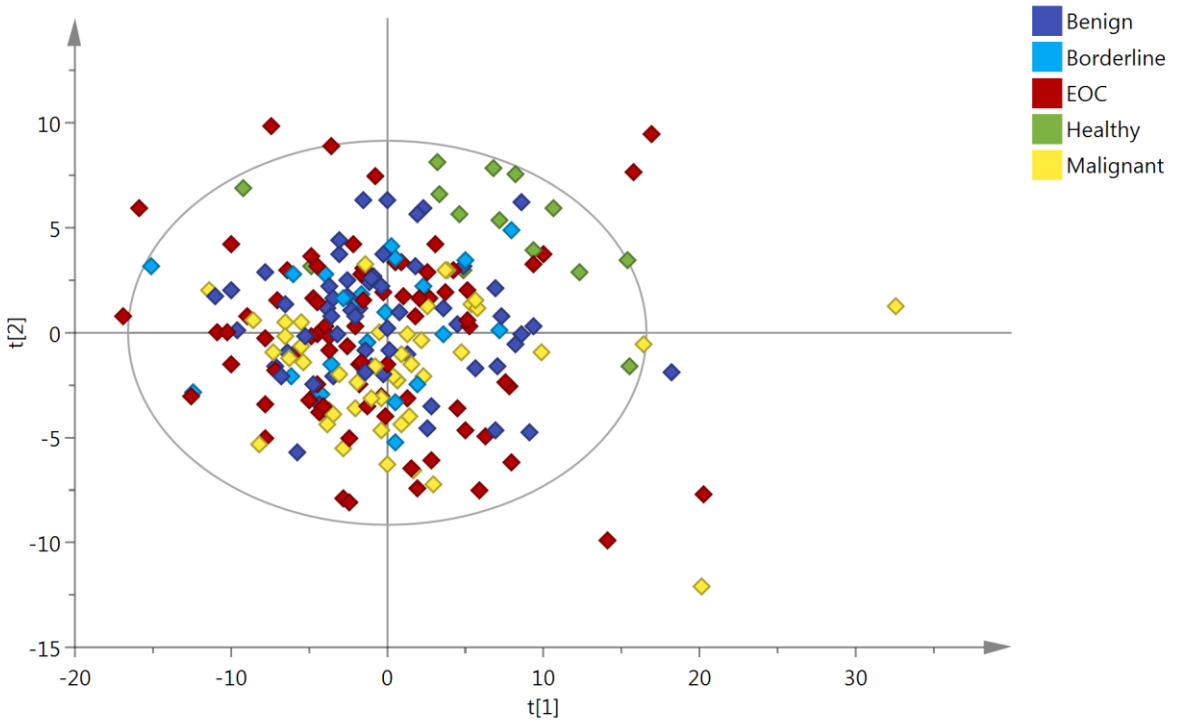



R2X[1] = 0.52 R2X[2] = 0.147 Ellipse: Hotelling's T2 (95%)

**Figure 5.2** Unsupervised PCA scores plot of all samples and IQC samples coloured by plate number (3-6) of the Biocrates data.

No distinct pattern according to plate number was seen across patient samples and QC samples in Figure 5.2 and therefore there was no requirement for normalisation according to metabolite concentrations for IQC samples. The need for normalisation across plates would be detrimental to the routine use of this assay or any metabolic mass-spectrometry based quantitative assay for diagnostic use in a clinical setting as diagnostic cut-offs would be plate specific and not absolutely quantified. Principal component analysis showed that the largest variance in the model was not explained by age, ethnicity or histological sub-type alone.

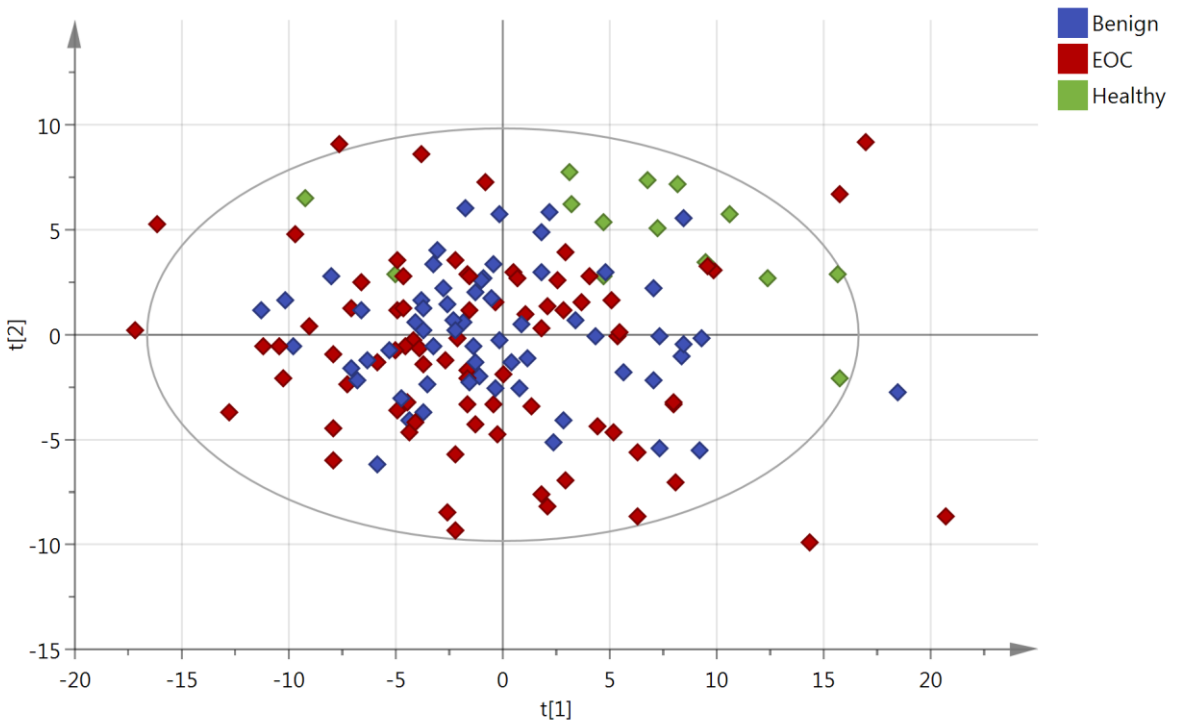
To identify the metabolic signature associated with epithelial ovarian cancer, multivariable principal components analysis was performed on all 122 metabolites from the Biocrates Absolute*IDQ*® p150 Kit in 224 plasma samples in patients with EOC, cancer of non-ovarian origin (primarily arising from the endometrium), benign disease, no disease (healthy) or a borderline tumour, see Figure 5.3.




R2X[1] = 0.371 R2X[2] = 0.113 Ellipse: Hotelling's T2 (95%)

**Figure 5.3** Unsupervised PCA scores plot of plasma samples coloured according to class of the Biocrates data. EOC – Epithelial ovarian cancer, Benign – benign gynaecological pathologies, Malignant – Cancer not arising from the ovary, primarily gynaecological, Healthy – No known cancer, Borderline – Borderline ovarian masses.

Other than the relatively small number of healthy samples (n=14), coloured green in Figure 5.3, there appears to be no obvious clustering of samples based on the signature of 122 metabolites from the assay. To further elucidate the metabolic signature associated with EOC, Figure 5.4 shows the PCA scores plot of just the 155 plasma samples from patients with EOC, benign ovarian pathologies and healthy controls.

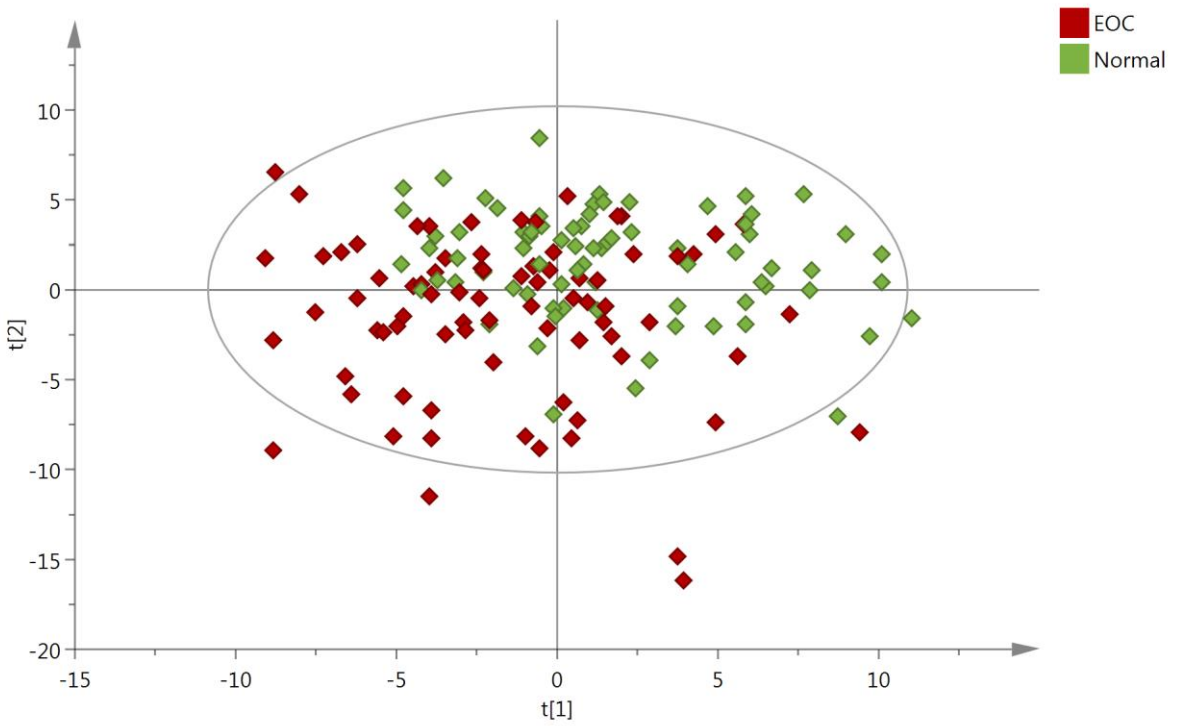



R2X[1] = 0.37 R2X[2] = 0.129 Ellipse: Hotelling's T2 (95%)

**Figure 5.4** Unsupervised PCA scores plot of plasma samples from patients with EOC, benign gynaecological disease and healthy volunteers of the Biocrates data.

As the plasma samples from healthy subjects appeared to cluster together in the original PCA; EOC, benign and healthy class samples only are shown in Figure 5.4. The samples from healthy controls were combined with the samples from patients with benign disease termed the 'Normal' group to further attempt to identify metabolites capable of distinguishing patients with EOC and those without, more closely resembling the clinical setting as not all patients without disease will have a benign pathology.

To optimise separation between the two classes and identify a metabolic signature associated with each class, a supervised PLS-DA model was produced, Figure 5.5. Here the independent X variables are the individual metabolite concentrations and the dependent Y variable is the class and comprises a single vector, i.e. Benign =0 and EOC =1.




R2X[1] = 0.277 R2X[2] = 0.182 Ellipse: Hotelling's T2 (95%)

**Figure 5.5** Supervised PLS-DA scores plot of plasma samples from patients with EOC and benign ovarian disease and Healthy volunteers, grouped to 'Normal' of the Biocrates data, R2Y = 0.385, Q2Y=0.186, R2X=0.535 using three components.

The associated loadings plot for the PLS-DA is shown in Figure 5.6. The loadings plot will help to elucidate which metabolites are causing the separation between the two classes.

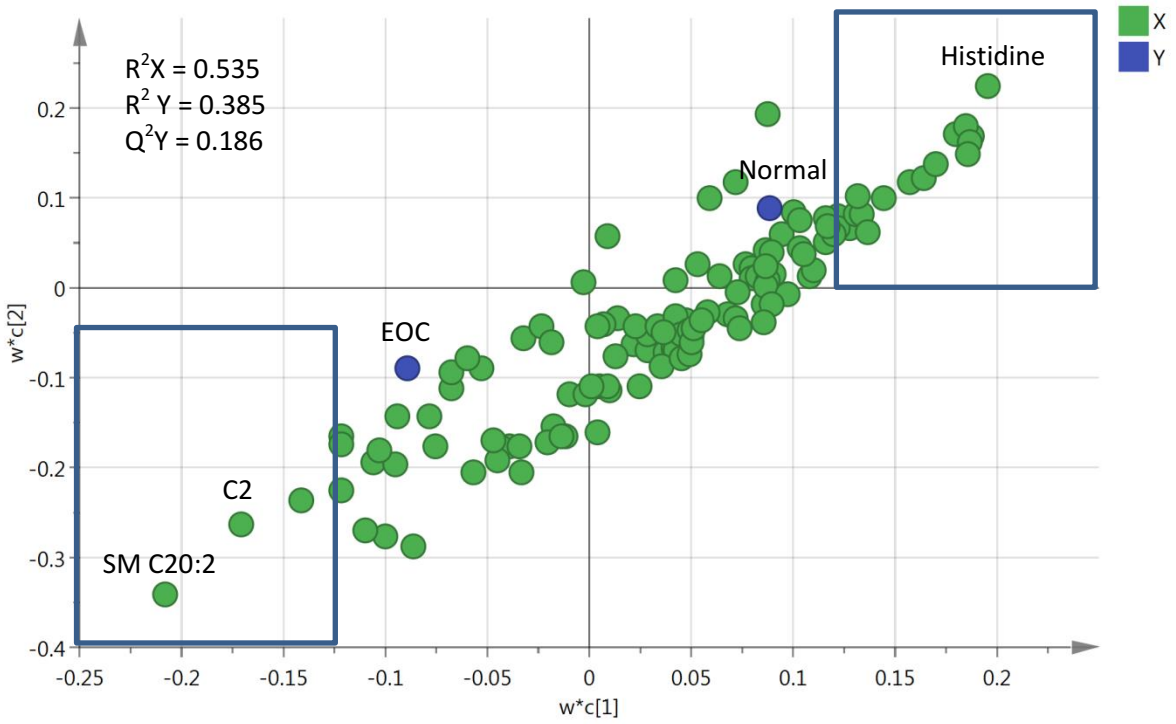



**Figure 5.6** Supervised PLS-DA loadings plot of plasma samples from patients with EOC and benign ovarian disease and Healthy volunteers, grouped to 'Normal' of the Biocrates data.

The PLS-DA loading weights scatter plot, Figure 5.6 displays the relation between the X- variables and the Y-variables in this case, EOC and benign disease. X-variables situated in the vicinity of the dummy Y-variables, the blue spots, have the highest discriminatory power between the classes. The highest discriminatory variable associated with EOC is the sphingomyelin; SM C20:2 and acylcarnitine; C2 and the highest discriminatory variable associated with the 'Normal' group is the amino acid; histidine.

The R2 and Q2 statistics describe the predictive ability of the classification models. R2X and R2Y represent the fraction of variance of the X and Y matrix explained by the model and Q2Y represents the predictive accuracy of the model.

The importance of each metabolite in the PLS-DA model was evaluated by variable importance in the projection (VIP) score. The VIP score positively reflects the metabolite's influence on the classification. VIP-values above 1 are considered important X-variables. The VIP plot is shown in Figure 5.7.

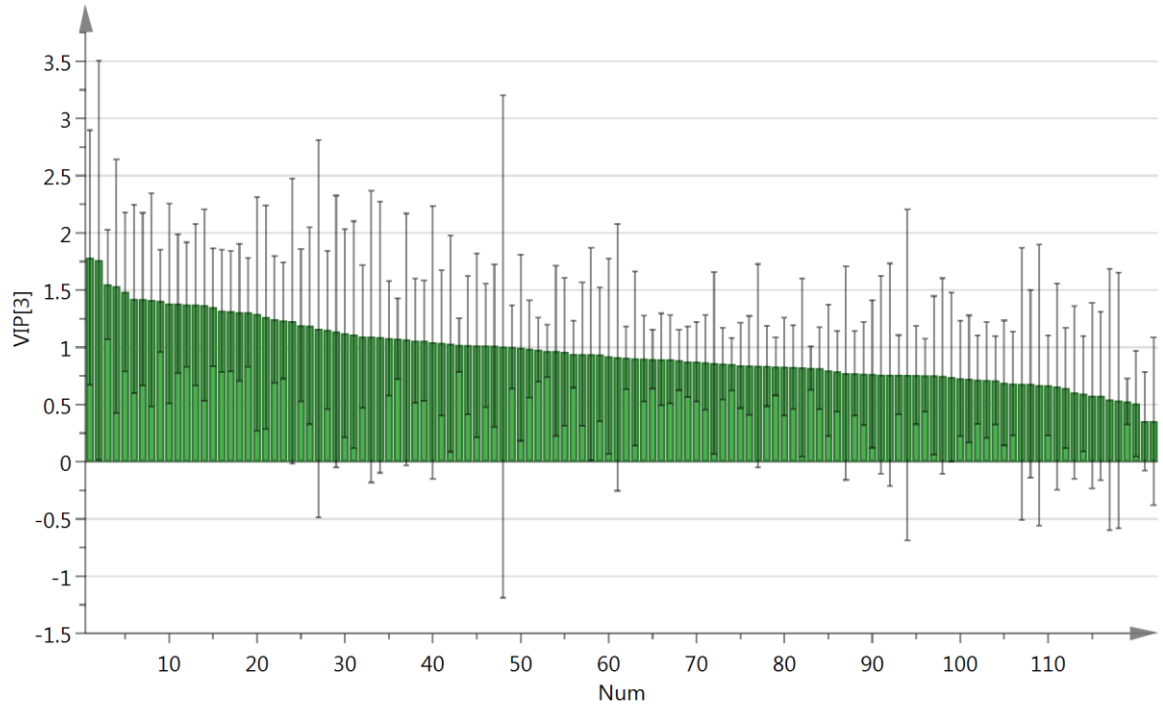



**Figure 5.7** Variable importance in the projection (VIP) plot summarising the importance of the variables to explain X and to correlate to Y (EOC/Normal) in the PLS-DA model of the Biocrates data.

Similarly to the loadings plot, in Figure 5.10, the metabolites with the largest VIP score were; SM C20:2 (VIP score 1.78), C18:2 (VIP score 1.76), PC ae C40:2 (VIP score 1.55), C18:1 (VIP score 1.53), C2 (VIP score 1.49), PC ae C34:2 (VIP score 1.42), PC ae C34:3 (VIP score 1.42) and histidine (VIP score 1.41). To determine the diagnostic performance of the metabolic signature a ROC curve was generated for the patient groups of those with and without EOC, the ROC curve is shown in Figure 5.8.

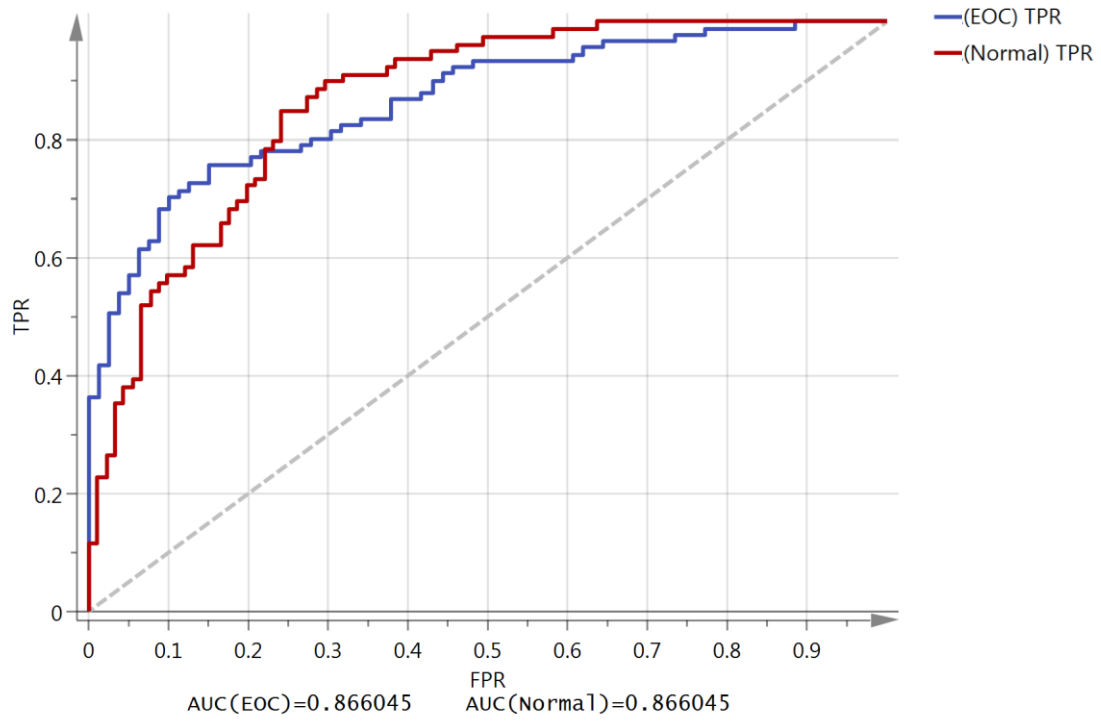



**Figure 5.8** ROC curve for non-malignant 'Normal' and EOC groups by metabolic signature based on the Biocrates Absolute*IDQ*® p150 Kit of 122 metabolites.

The AUC was 0.866 for prediction of EOC using all 122 metabolites from the Biocrates AbsoluteIDQ® p150 kit included in the analysis.

### 5.3.2.2 Univariable data analysis of metabolite concentrations in plasma using the Biocrates Absolute*IDQ*® p150 Kit

As there appears to be variation in the metabolic signature produced using the targeted assay not associated with epithelial ovarian cancer, the Wilcoxon rank sum test was used to identify the individual quantified metabolites that showed phenotype-associated changes that were significantly different between both the EOC and the benign groups and then between the EOC group and the 'Normal' group consisting of the benign and healthy groups combined. A total of 122 metabolites were investigated in EOC (n=73), benign (n=55) and healthy groups (n=14). Duplicate samples were removed to ensure each sample was from a unique patient. The p-values from the Wilcoxon-rank sum test were FDR adjusted using the Benjamini-Hochberg method and an adjusted p-value <0.05 was considered significant.

A total of 6 metabolites were significantly different between the benign and EOC groups. As predicted from the loadings plot in Figure 5.6 and in the VIP plot in Figure 5.7, SM C20:2 (VIP score 1.78) was significantly higher in the EOC group and PC.aa.C34:4 (1.19) was significantly higher in the

benign group as well as the amino acid; histidine (VIP score 1.41), lysoPC.a.C18.2 (VIP score 1.26), PC.ae.C36.3 (VIP score 1.38) and PC.ae.C40.1 (VIP score 1.31). The box and whisker plots of these metabolites are shown in Figure 5.9.

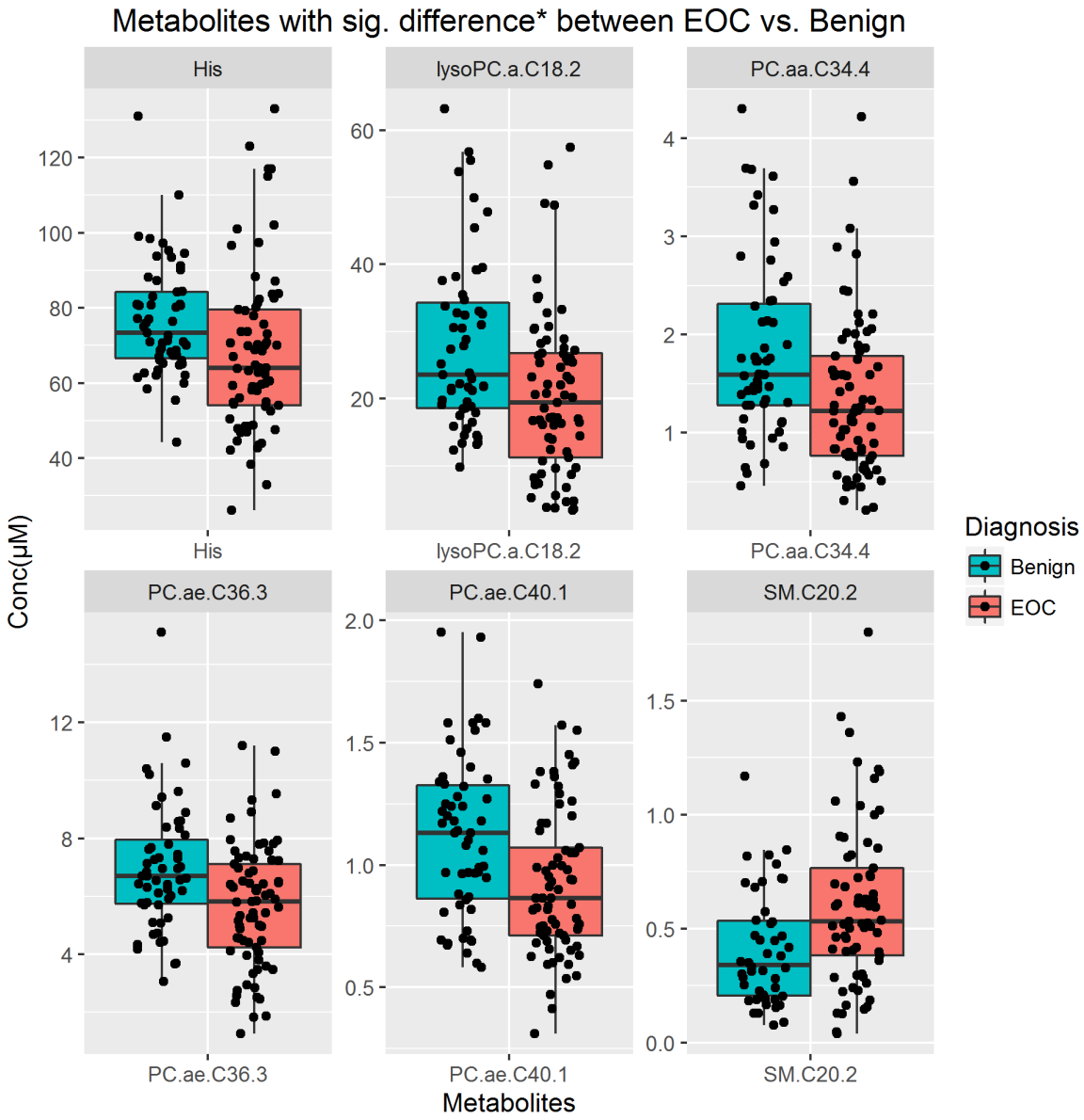



**Figure 5.9** Metabolites from Biocrates targeted assay that are significantly different between EOC and Benign groups. *FDR adjusted p-value <0.05.

A combination of the results from the benign and healthy classes to produce the 'Normal' group resulted in a total of 30 significantly different metabolites between the two groups, this included all the metabolites previously defined as significant between benign and EOC, Figure 5.10.

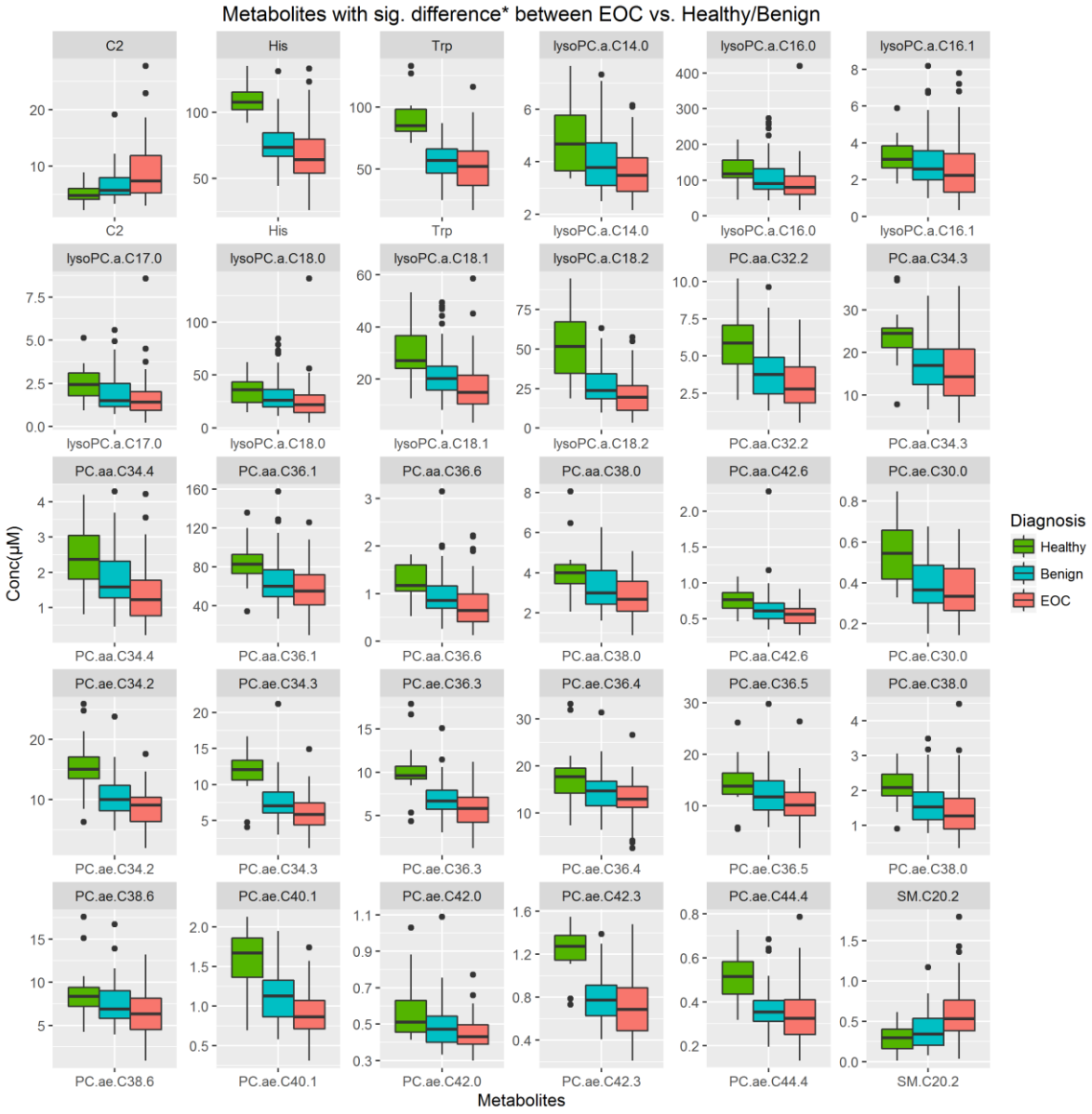



**Figure 5.10** Metabolites from Biocrates targeted assay that are significantly different between EOC and normal groups. *FDR adjusted p-value <0.05.

5.3.2.2.1 Association of metabolites with grade and stage of disease

To further elucidate the potential role of these metabolites, rank correlation using kendall's tau was used to determine whether these metabolites are also significantly associated with stage or grade.

Rank correlation of the significantly different metabolites between the EOC and 'Normal' group showed that C2, tryptophan, lyso.PC.a.C18:2, PC.aa.C32:2, PC.aa.C34:4 were associated with stage. Following the pattern seen between the EOC and normal group; C2 showed an increasing concentration based on stage while tryptophan, lyso.PC.a.C18:2, PC.aa.C32:2 and PC.aa.C34:4 all showed a decreasing concentration with increasing stage, Figure 5.11.

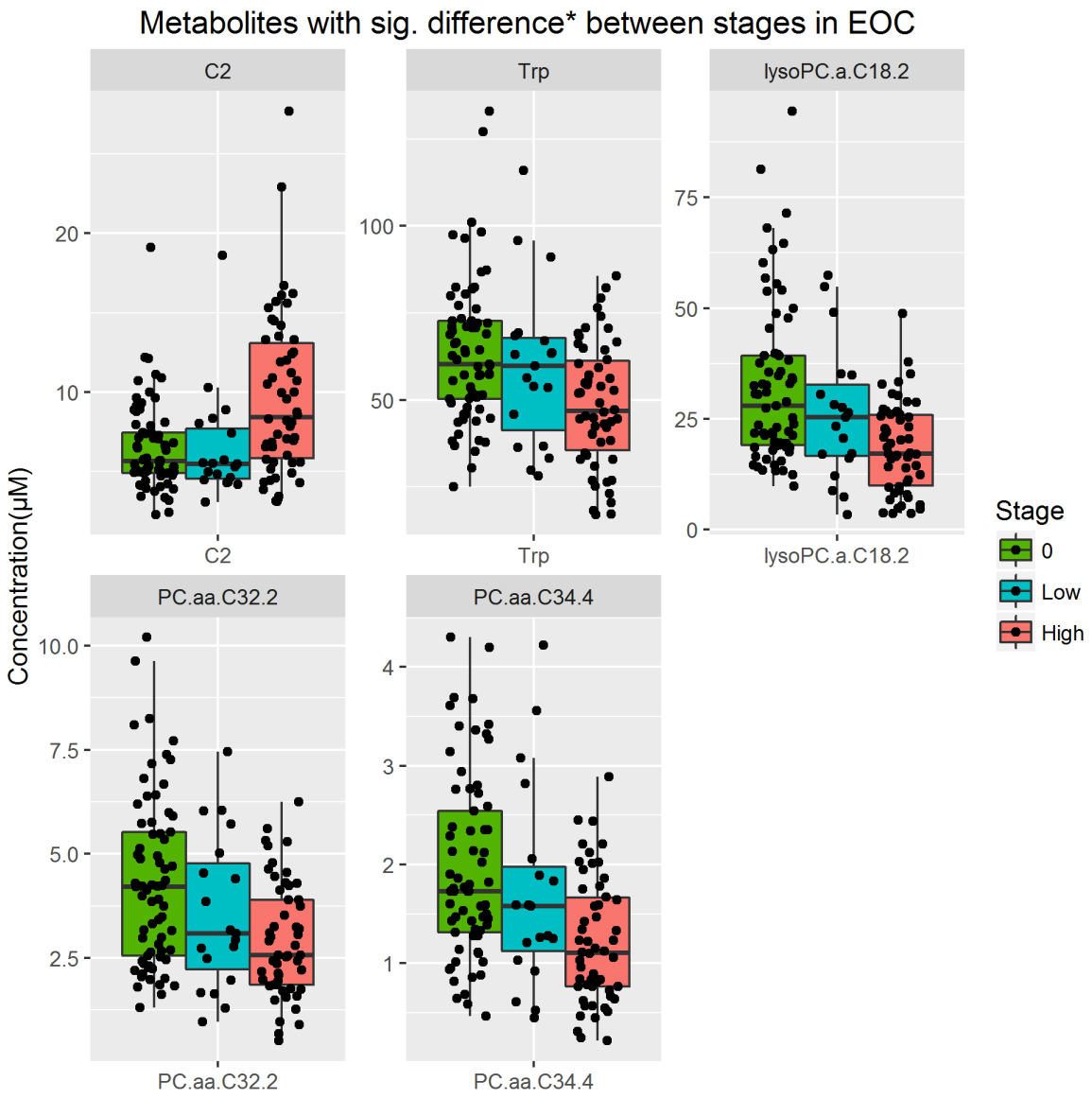



**Figure 5.11** Box and whisker plots of metabolites that show an association between stage (Low = Stage I, II and High = Stage III, IV) and metabolite concentration, in the EOC group only of the Biocrates data. Values for samples with stage 0 (benign/healthy) are also shown for comparison.

Rank correlation of the significantly different metabolites between the EOC and 'Normal' group showed that tryptophan, histidine, lyso.PC.a.C18:2, PC.aa.C36:1, PC.aa.C36:3 were associated with grade of disease. Following the pattern seen between the EOC and normal group; all metabolites showed a decreasing concentration with increasing grade, Figure 5.12.

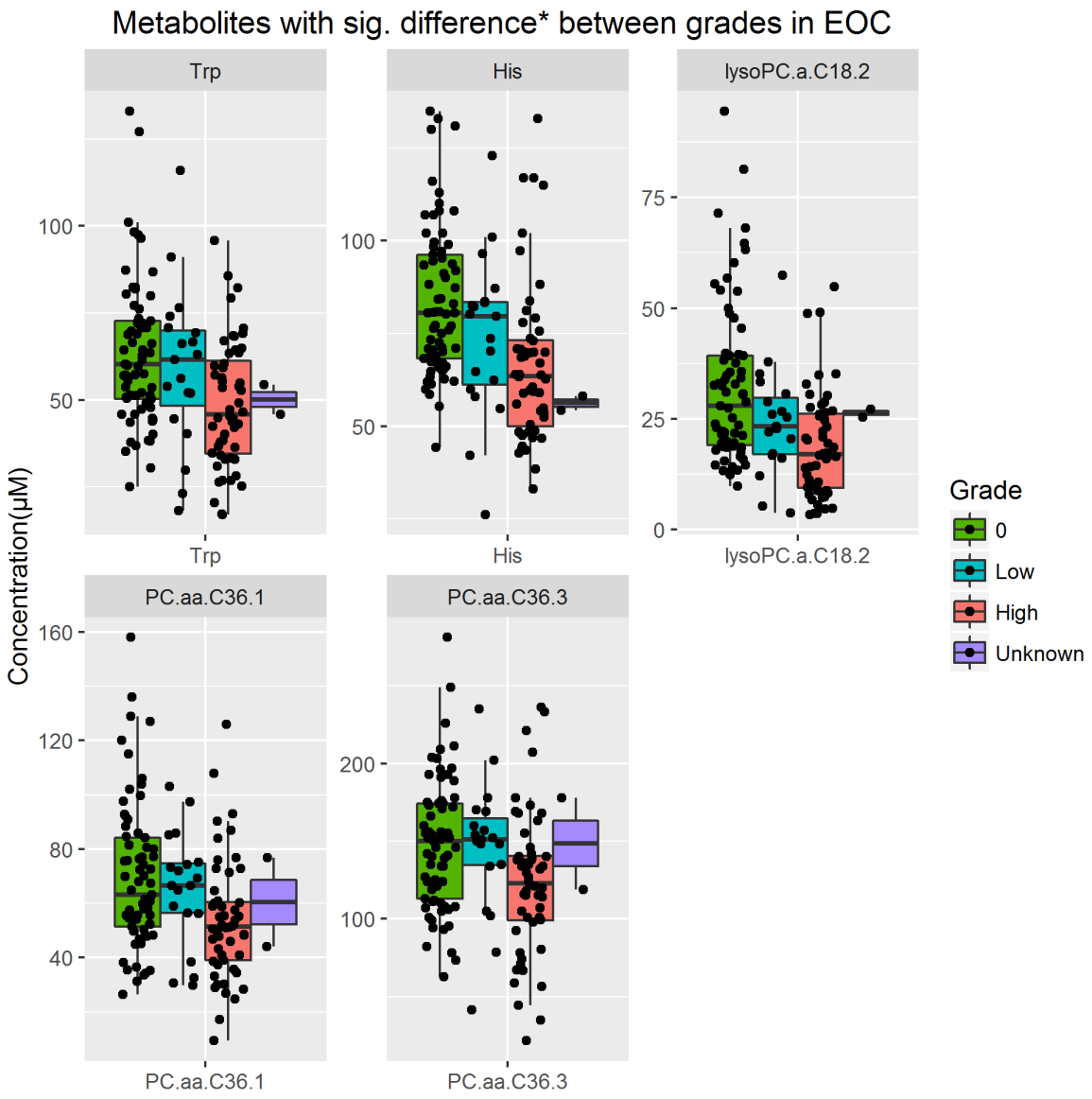



**Figure 5.12** Box and whisker plots of metabolites that are show an association between grade (Low = Grade I,II and High = Grade III, Undifferentiated) and metabolite concentration, in the EOC group only of the Biocrates data. Values are also shown for samples with grade 0 (benign/healthy) for comparison.

### 5.3.3 Predictive capacity of metabolite concentrations for diagnosis of EOC in plasma

To assess the performance of each of the significantly different metabolites in diagnosing EOC, ROC curves were produced and the AUC determined for each metabolite. The highest AUC was based on metabolite concentrations of histidine (AUC = 0.719), lyso.PC.a.C18.2 (AUC=0.704) and PC.ae.C40.1 (AUC=0.707). A combination of markers in a predictive model may further increase the predictive power of the metabolites and therefore a stepwise additive regression was used from known predictors to find the best model fit according to Akaike information criterion (AIC). The model

requires all metabolites to be independent of one another and therefore Pearson correlation between each metabolite was determined, see Figure 5.13.

Pearson's correlation coefficient was calculated for the concentration of each of the thirty metabolites that was significantly different between samples from patients with EOC compared to healthy and benign patients against the others. Only metabolites with an $R^2$ <0.7 compared to the other metabolites were included. The concentrations of these nine metabolites were combined in a step-wise logistic regression model to produce a final model consisting of five metabolites: C2, histidine, tryptophan, PC.ae.C34.3 and SM.C20.2.

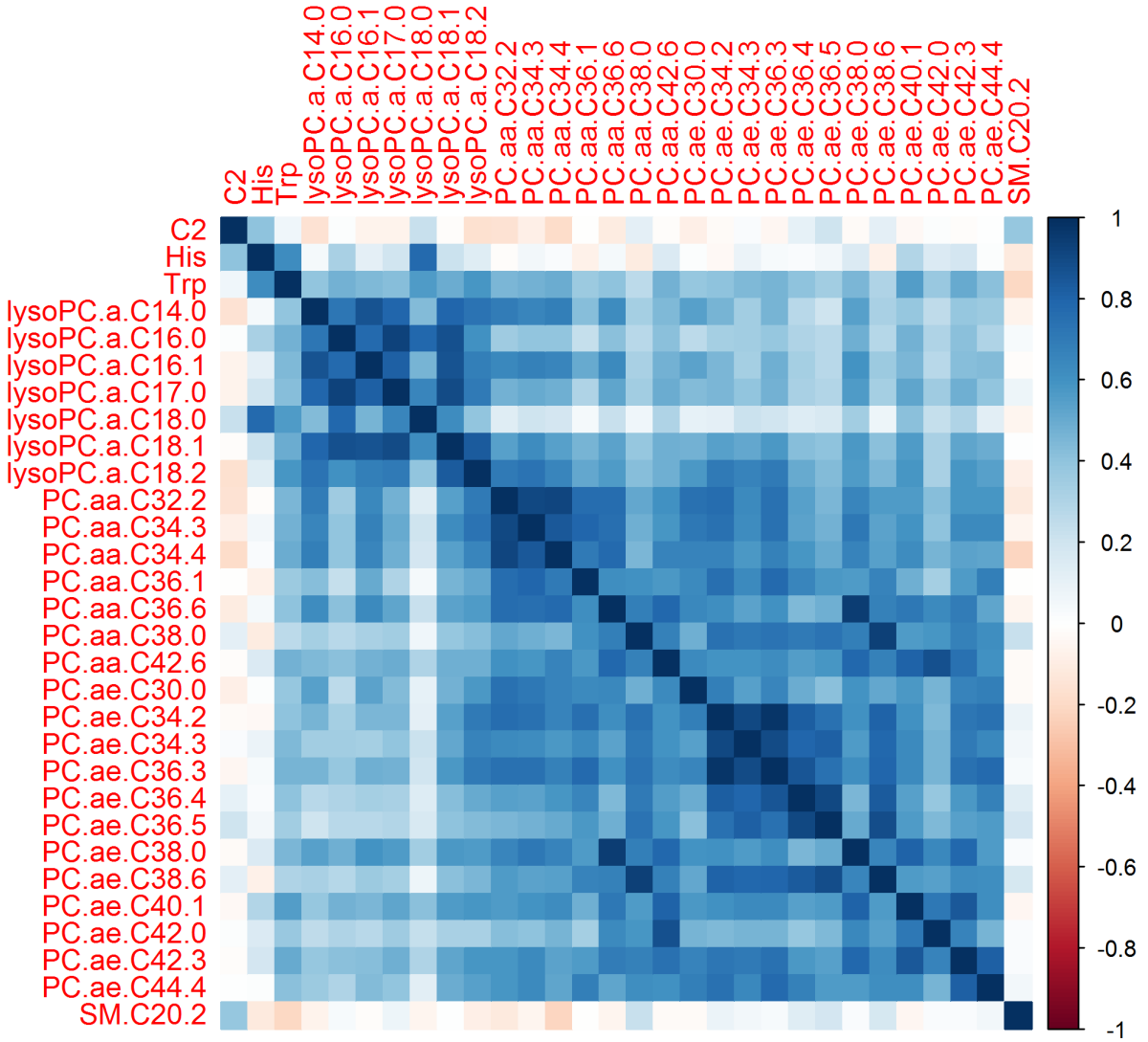



**Figure 5.13** Correlation plot of the 30 significantly different metabolites between the EOC and normal groups in the Biocrates data.

As there was no prospectively collected patient cohort to test the model, the combined predictive model was internally validated using the Leave one out cross validation (LOOCV) method.

Using the Leave one out Cross Validation (LOOCV) method based on concentration of the five metabolites, 89 out of 127 (70%) samples were correctly predicted and only 15 samples could not be classified due to missing data (11%).

A confusion matrix of the predictive ability of the model is shown in Table 5.2. The PPV = 72% and NPV = 68% and the sensitivity = 74% and specificity = 65%.

**Table 5.2** Confusion matrix for the performance of the cross-validated step-wise logistic regression model for 5 metabolites with significantly different levels in plasma of patients with EOC and without of the Biocrates data.

| | | **State of disease** | |
|---|---|---|---|
| | | Disease +<br>69 | Disease –<br>58 |
| **Results of LOOCV model** | Test +<br>71 | True positive<br>51 | False positive<br>20 |
| | Test –<br>56 | False negative<br>18 | True negative<br>38 |

ROC analysis of the cross-validated step-wise logistic regression model produced an AUC of 0.804, see Figure 5.14.

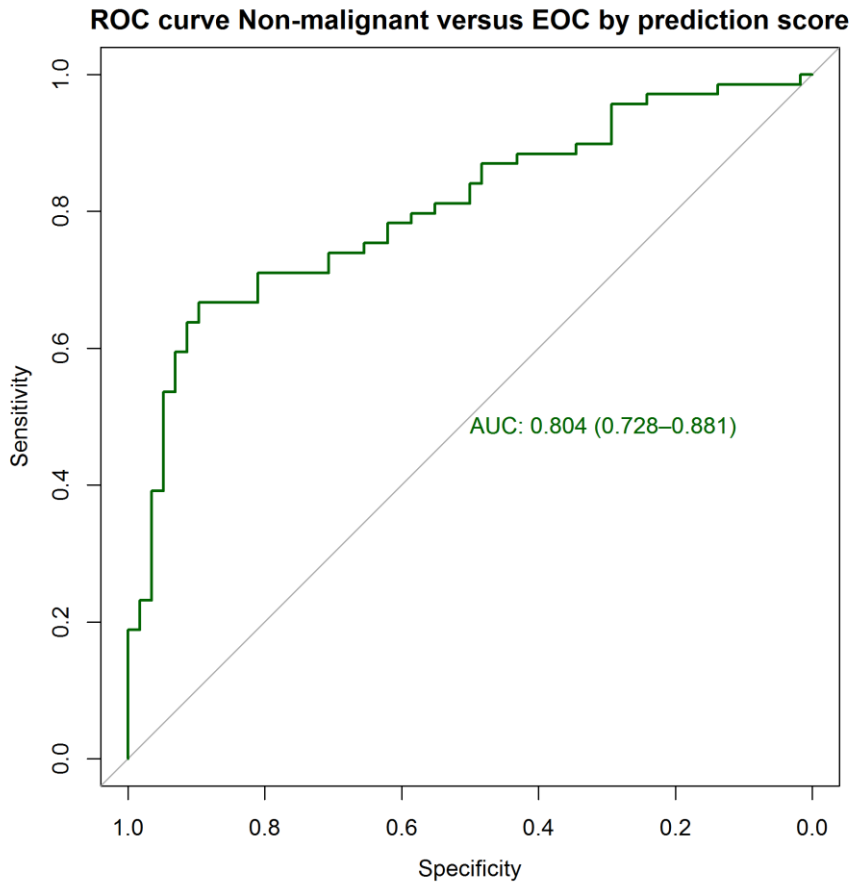



**Figure 5.14** ROC curve for non-malignant 'Normal' versus EOC by cross-validated prediction score based on step-wise logistic regression model of 5 metabolites of the Biocrates data.

### 5.3.4 Variation of selected metabolites in plasma

Biological replicates were taken from 15 patients at different time points by different people. To compare the replicates, concentrations of the predictive metabolites between the duplicate samples were compared. The reproducibility (precision) using the coefficient of variation (%CV) of each metabolite between each replicate is shown in Table 5.3. A prediction score for the biological replicates that were removed from the original analysis was also produced using the LOOCV model

**Table 5.3** Coefficient of variation (%) between biological replicates for significantly different metabolites used in stepwise logistic regression model of the Biocrates data.

| | C2 (%CV) | His (%CV) | Trp (%CV) | PC.ae.C34:3 (%CV) | SM C20:2 (%CV) | Time between sampling (days) | Class | Prediction score from cross validated model of (binary vector) | | Model performance | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Biological replicate included in model | Biological replicate excluded from model | Biological replicate included in model | Biological replicate excluded from model |
| D1 | 0.60 | 0.56 | 0 | 1.55 | 17.07 | 0 | EOC | 0.96 | 0.98 | TP | TP |
| D2 | 1.83 | 8.48 | 2.5 | 2.4 | 59.7 | 0 | Benign | 0.05 | 0.06 | TN | TN |
| D3 | 42.86 | 2.32 | 1.01 | 0.97 | 19.14 | 0 | EOC | 0.898 | 0.565 | TP | TP |
| D7 | 5.94 | 1.87 | 6.26 | 8.02 | NA | 0 | EOC | NA | 0.419 | NA | FN |
| D8 | 7.36 | 8.23 | 11.07 | 7.17 | 32.34 | 1 | Benign | 0.5 | 0.56 | FP | FP |
| D9 | 12.57 | 8.58 | 10.63 | 13.70 | 19.06 | 2 | Benign | 0.6 | 0.6 | FP | FP |
| D10 | 12.04 | 13.63 | 12.73 | 11.99 | 9.64 | 1 | Borderline | 0.764 | 0.744 | NA | NA |
| D12 | 17.65 | 0 | 1.25 | 2.20 | 15.70 | 2 | Benign | 0.65 | 0.73 | FP | FP |
| D13 | 7.51 | 4.97 | 6.98 | 5.42 | NA | 0 | Borderline | NA | NA | NA | NA |
| D14 | 21.56 | 13.71 | 11.42 | 22.96 | 16.61 | 0 | EOC | 0.411 | 0.122 | FN | FN |
| D15 | 13.33 | 7.73 | 15.30 | 15.21 | 11.33 | 0 | EOC | 0.81 | 0.85 | TP | TP |
| D17 | 14.60 | 14.45 | 12.91 | 15.22 | NA | 0 | Benign | NA | NA | NA | NA |
| D18 | 5.57 | 1.91 | 5.16 | 4.92 | 6.79 | 5 | Borderline | NA | NA | NA | NA |
| D20 | 11.11 | 9.89 | 14.24 | 14.16 | 0.70 | 0 | EOC | 0.199 | 0.1 | FN | FN |
| D21 | 5.81 | 2.78 | 2.24 | 1.28 | 15.67 | 3 | Benign | 0.29 | 0.32 | TN | TN |

D6 was removed from the analysis as one of the replicates was subsequently found to not have been taken pre-operatively. A prediction score for the biological replicates that were removed from the original analysis was produced using the LOOCV model. Comparison of the prediction scores across the duplicate samples revealed that no samples had a prediction score that was discrepant compared to the original analysis, see Table 5.3.

### *5.3.3 Survival analysis*

Although OS and PFS were previously reported for this cohort in Chapter 4: Section 4.3.3, an additional number of samples were included in the Biocrates study that were not included in the NMR study due to restrictions on sample volume. To check that this cohort is also consistent with the literature as well as consistent with the subset of samples used in the NMR study, the analysis was repeated on the samples included in the Biocrates study.

Overall survival (OS) and progression-free survival (PFS) data was available on 67 patients. Univariate logrank statistical tests were performed. There was a significant difference in OS and PFS dependent on residual disease volume post-surgery divided into 0mm, 1-10mm and >11mm, logrank p-value=0.049 and p=0.004 respectively. (Figure 5.15 A & B). Although there was no significant difference between OS in patients defined as optimally (≤10mm residual disease) and suboptimally debulked (>10mm residual disease) (Figure 5.15 C) the greatest survival advantage came in those patients where complete macroscopic debulking was achieved (0mm residual disease) (Figure 5.15 E). This is in keeping with previously published data (Wimberger *et al.*, 2010). Overall median survival time in the group with 0mm residual disease was 2036 days (95% CI 1691-NA) compared to 937 days (95% CI 776-1565) in the group with any macroscopic residual disease whereas patients with optimal surgical debulking had a median overall survival time of 1678 days (95% CI 1359-NA) compared to 1045 days (95% CI 419-NA) in those who had suboptimal debulking.

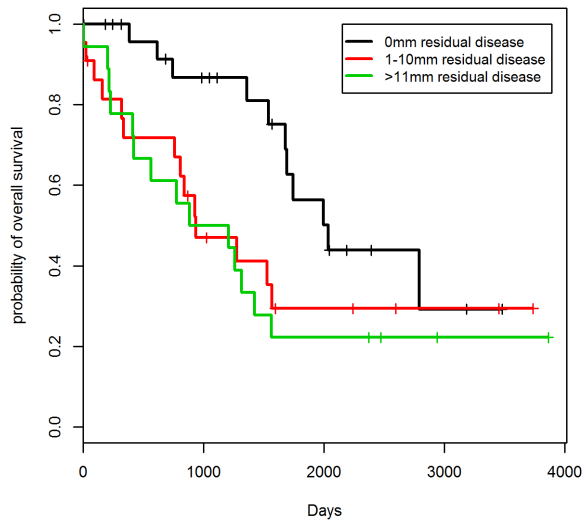
Progression free median survival time in the group with 0mm residual disease was 1297 days (95% CI 826-NA) compared to 377 days (95% CI 319-552) in the group with any macroscopic residual disease (Figure 5.15 F). Patients with optimal surgical debulking had a progression free survival time of 826 days (95% CI 511-1744) compared to 283 days (95% CI 196-549) in those who had suboptimal debulking highlighting that those patients with the most residual disease (>10mm) have the shortest time to progression(Figure 5.15 D).

The significant difference in overall survival dependent on residual disease volume was lost in multivariate cox analysis when adjusting for age, stage and grade (p=0.335, HR=1.21, 95%CI 0.82-1.80) and dependent on presence of residual disease (>1mm) versus absence (0mm) when adjusting for confounding factors (p=0.073, HR=1.99 95%CI 0.94-4.21).

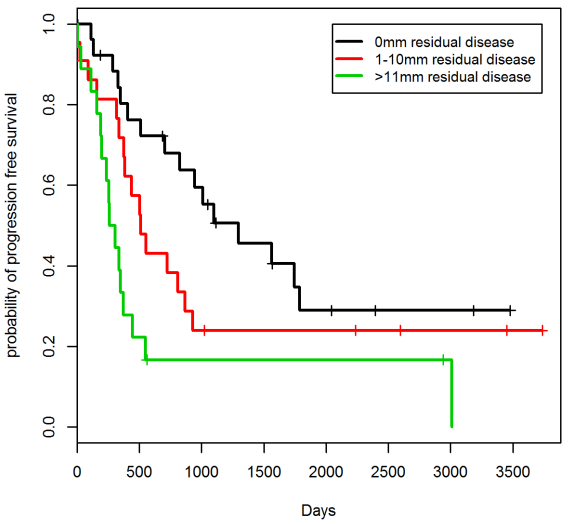
The significant difference in progression free survival dependent on residual disease volume was also lost in multivariate cox analysis when adjusting for age, stage and grade (p=0.137, HR=1.33, 95%CI 0.91-1.95) and dependent on presence of residual disease when adjusting for confounding factors (p=0.129, HR=1.66 95% CI (0.86-3.17). Univariate analysis demonstrated that age and stage were

significantly associated with survival, $p<0.001$ and $p=0.009$ respectively and therefore were adjusted for in subsequent models.

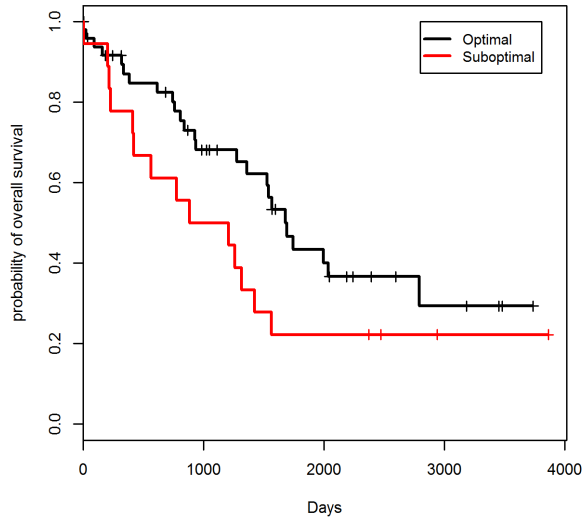
A


Probability of OS dependent on residual disease volume, p=0.0488, N=67
0mm residual disease
1-10mm residual disease
>11mm residual disease
probability of overall survival
Days

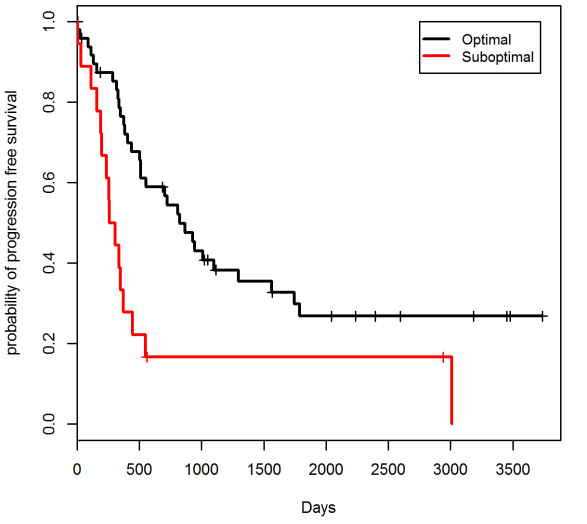
B


Probability of PFS dependent on residual disease volume, p=0.0041, N=67
0mm residual disease
1-10mm residual disease
>11mm residual disease
probability of progression free survival
Days

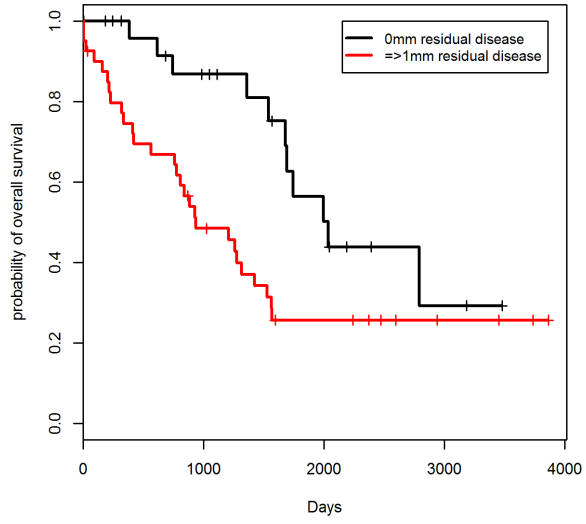
C


Probability of OS, optimal vs. suboptimal, p=0.0949, N=67
Optimal
Suboptimal
probability of overall survival
Days

D


Probability of PFS, optimal vs. suboptimal, p=0.00208, N=67
Optimal
Suboptimal
probability of progression free survival
Days

E


Probability of OS, total vs. any residual disease, p=0.0156, N=67
0mm residual disease
=>1mm residual disease
probability of overall survival
Days

F

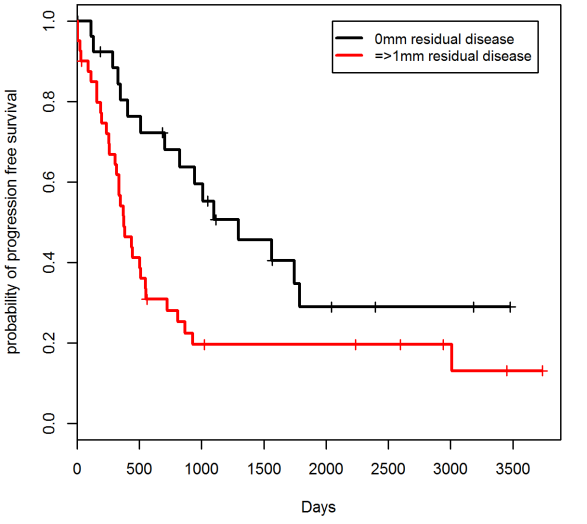

Probability of PFS, total vs. any residual disease, p=0.0132, N=67
0mm residual disease
=>1mm residual disease
probability of progression free survival
Days

**Figure 5.15** Kaplan-Meier survival plots for OS and PFS in the Biocrates ovarian study. **A**, OS and **B**, PFS dependent on residual disease volume. **C**, OS and **D**, PFS dependent on optimal or suboptimal debulking. **E**, OS and **F**, PFS dependent on complete macroscopic debulking (0mm residual disease) vs. any amount of remaining residual disease (>1mm).

#### 5.3.3.1 Significantly different metabolites associated with overall survival in all patients with epithelial ovarian cancer

In order to identify metabolite concentrations in blood plasma that are associated with survival, data from the Biocrates Absolute*IDQ*® p150 Kit was again used from all patients diagnosed with epithelial ovarian cancer. The nonparametric logrank test was used to identify if patients had different survival distributions above and below the median of each metabolite concentration that was previously shown to have a significantly different concentration between EOC and benign/healthy controls in the initial phase of this study. Multivariable cox proportional hazard models adjusting for age, stage, grade and presence of residual disease (none vs. >1mm) were then performed on these metabolites.

Six out of 30 significantly different metabolites were found to be associated with overall survival when adjusted for confounding factors ($p<0.05$) (Table 5.4). 5 out of the 6 metabolites; tryptophan, lysoPC.a.C18.0, PC.aa.C36.6, PC.ae.C38.0 and PC.ae.C42.3 demonstrated a concentration above the median to be associated with improved survival whereas C2 demonstrated a concentration below the median to be associated with improved survival. This pattern matches the pattern seen in patients with EOC compared to benign/healthy controls where a higher concentration of C2 was seen in the plasma of patients with EOC and vice versa for the other 5 metabolites. Increasing concentrations of C2 were also seen with higher stage disease and decreasingly concentrations of tryptophan were also associated with higher grade and stage disease potentially suggesting a role of these metabolites in the pathogenesis of the disease. As the cox proportional hazard models were adjusted for stage and grade this difference in survival was not due to these confounding factors.

Kaplan Meier survival curves for these 6 metabolites are shown in Figure 5.16.

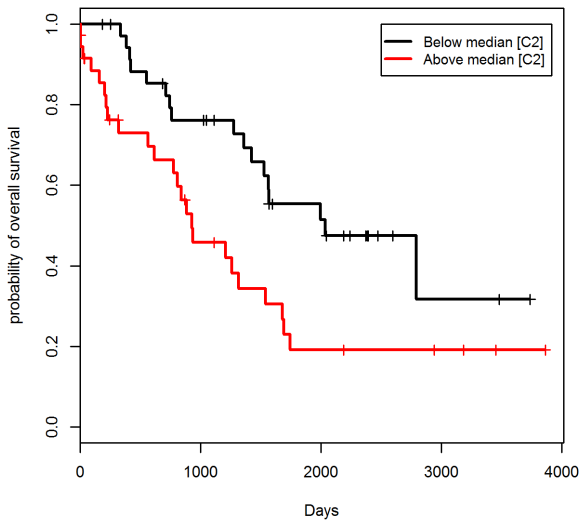
The metabolite, tryptophan showed the largest effect size equivalent to a HR of 0.39 (95% CI 0.18-0.83). The overall median survival time in the group with tryptophan concentration below the median (52µM) was 1207 days (95% CI 776-1678) compared to 2793 days (1425-NA) in the group with tryptophan concentration above the median.

**Table 5.4** Summary of 6 metabolites from the Biocrates Absolute*IDQ*® p150 targeted assay that are associated with survival.

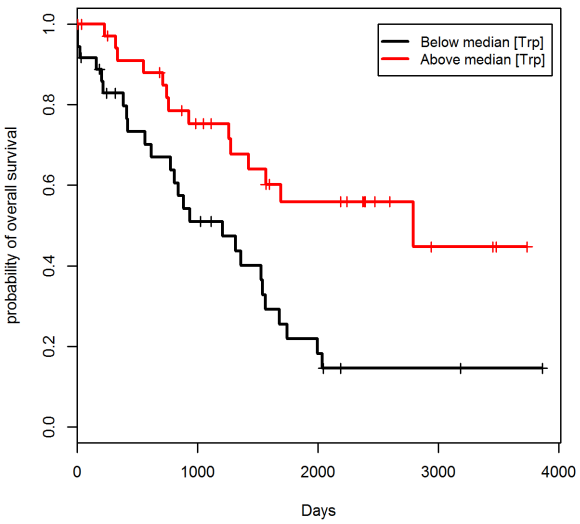
| Metabolite | Median concentration (μM) | Cox model (p-value*) | Hazard ratio | 95% CI | Logrank test (p-value) | Improved survival associated with |
|---|---|---|---|---|---|---|
| C2 | 7.4 | 0.0418 | 2.11 | 1.03-4.32 | 0.0096 | Lower concentration |
| Tryptophan | 52 | 0.0149 | 0.39 | 0.18-0.83 | 0.0024 | Higher concentration |
| LysoPC.a.C18.0 | 22.1 | 0.0156 | 0.41 | 0.20-0.84 | 0.0013 | Higher concentration |
| PC.aa.C36.6 | 0.645 | 0.0338 | 0.48 | 0.25-0.95 | 0.0063 | Higher concentration |
| PC.ae.C38.0 | 1.27 | 0.0320 | 0.48 | 0.24-0.94 | 0.0024 | Higher concentration |
| PC.ae.C42.3 | 0.686 | 0.0498 | 0.51 | 0.26-0.99 | 0.0295 | Higher concentration |

* p-value adjusted for age, stage, grade, residual disease.

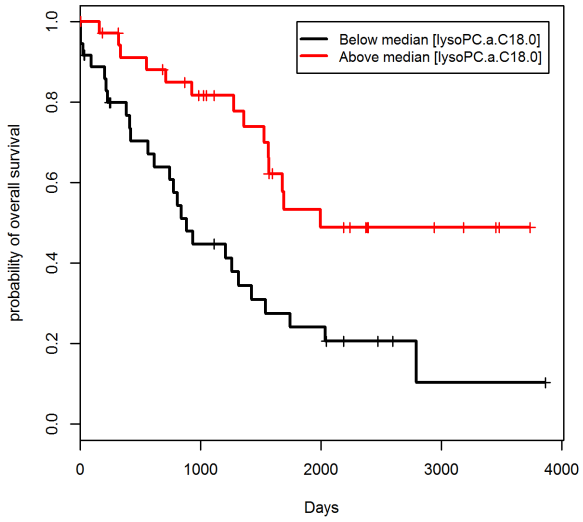
A


Probability of OS for C2, p=0.00957, N=72
Below median [C2]
Above median [C2]
probability of overall survival
Days

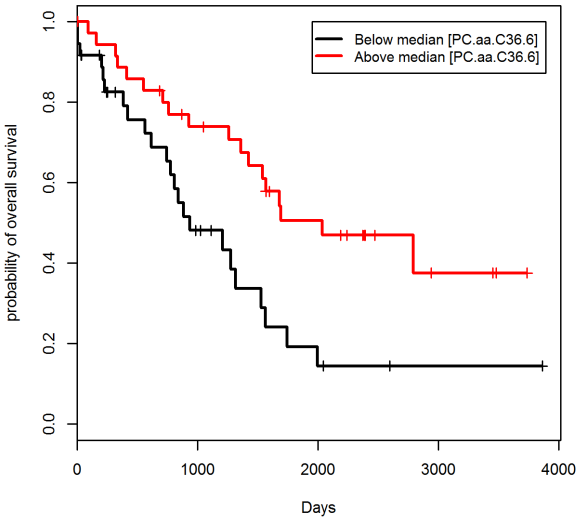
B


Probability of OS for Trp, p=0.00244, N=72
Below median [Trp]
Above median [Trp]
probability of overall survival
Days

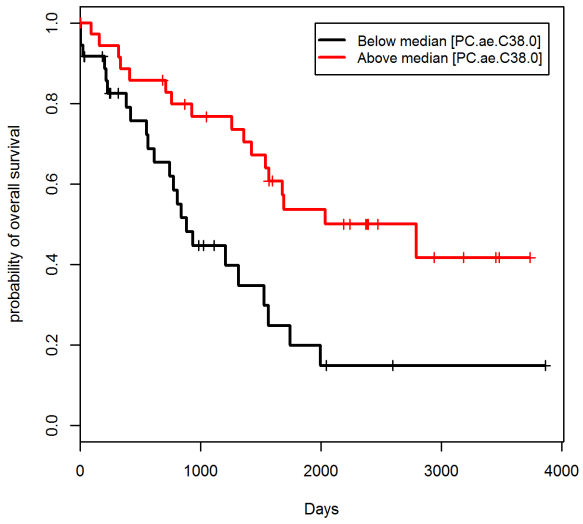
C


Probability of OS for lysoPC.a.C18.0, p=0.00128, N=72
Below median [lysoPC.a.C18.0]
Above median [lysoPC.a.C18.0]
probability of overall survival
Days

D


Probability of OS for PC.aa.C36.6, p=0.00627, N=72
Below median [PC.aa.C36.6]
Above median [PC.aa.C36.6]
probability of overall survival
Days

E


Probability of OS for PC.ae.C38.0, p=0.00235, N=72
Below median [PC.ae.C38.0]
Above median [PC.ae.C38.0]
probability of overall survival
Days

F

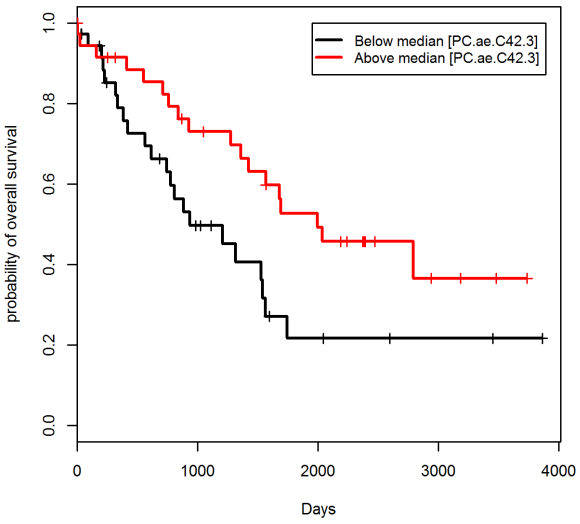

Probability of OS for PC.ae.C42.3, p=0.0295, N=72
Below median [PC.ae.C42.3]
Above median [PC.ae.C42.3]
probability of overall survival
Days

**Figure 5.16** Kaplan-Meier survival curves for OS dependent on metabolite concentration of the Biocrates Absolute*IDQ*® p150 targeted assay. **A**, C2. **B**, tryptophan. **C**, lysoPC.a.C18.0. **D**, PC.aa.C36.6. **E**, PC.ae.C38.0. **F**, PC.ae.C42.3.

#### 5.3.3.2 Significantly different metabolites associated with progression-free survival in all patients with epithelial ovarian cancer

The same analysis was repeated with progression free survival data in order to identify any metabolite concentrations in blood plasma that were associated with progression free survival using data from the Biocrates Absolute*IDQ*® p150 Kit in all patients diagnosed with epithelial ovarian cancer. The nonparametric logrank test was used to identify if patients had different survival distributions above and below the median of each metabolite concentration that was previously shown to have a significantly different concentration between EOC and benign/healthy controls in the initial phase of this study. Multivariable cox proportional hazard models adjusting for age, stage, grade and presence of residual disease (none vs. >1mm) were then performed on these metabolites. Concentrations of one metabolite, lysoPC.a.C18.0, was found to be associated with progression free survival when adjusted for confounding factors (p=0.000252). LysoPC.a.C18.0 concentrations above the median were associated with improved progression-free survival (Figure 5.17). Higher concentrations of this lysophospholipid were also associated with improved overall survival (Figure 5.16 C) suggesting a low concentration of this lysophospholipid is associated with progression and death with a HR of 0.44 (95% CI 0.23-0.86). The median progression free survival in the group with concentration above the median was 1100 days (95% CI 552-NA) compared to 348 days (95% CI 258-722) for concentrations below the median.

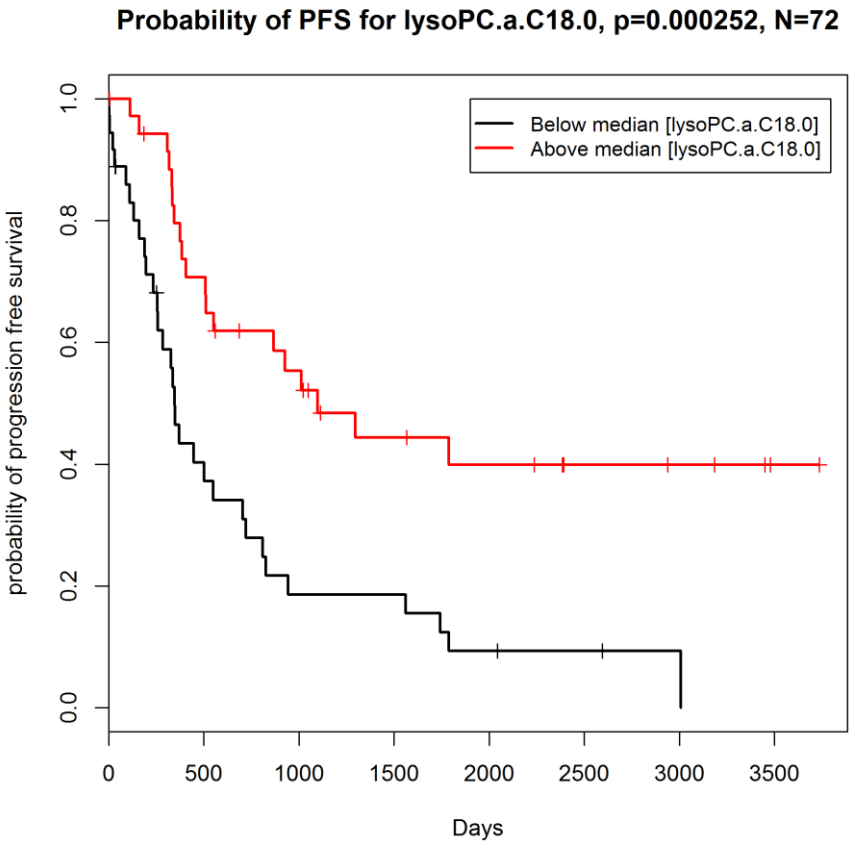



**Figure 5.17** Kaplan-Meier survival curve for PFS dependent on lysoPC.a.C18.0 concentration.

### *5.3.4 Predictive capacity of combining metabolites from analysis of plasma using $^{1}$H NMR and the Biocrates AbsoluteIDQ® p150 targeted assay for the diagnosis of EOC*

A total of 111 samples with both NMR spectral data and data from the Biocrates Absolute*IDQ*® p150 Kit were combined to assess whether combinations of biomarkers from two complementary techniques could further increase the diagnostic capability of the logistic regression model.

Using the Leave one out Cross Validation (LOOCV) method based on concentration of the five metabolites, 86 out of 111 (77%) samples were correctly predicted.

A confusion matrix of the predictive ability of the model is shown in Table 5.5. The PPV = 79% and NPV = 75%. The Sensitivity = 78% and Specificity = 77%.

Table 5.5 Confusion matrix for the performance of the cross-validated step-wise logistic regression model for the 5 metabolites in the combined model consisting of histidine and unknown.1.81ppm integrals from CPMG-$^{1}$H NMR, glycoprotein acetyls integral from NOESY-presat-$^{1}$H NMR and PC.ae.C44.4 and SM.C20.2 from the Biocrates Absolute*IDQ*® p150 Kit with significantly different levels in plasma from patients with EOC and without.

| | | **State of disease** | |
|---|---|---|---|
| | | Disease + | Disease – |
| **Results of LOOCV model** | Test + | True positive<br>46 | False positive<br>12 |
| | Test – | False negative<br>13 | True negative<br>40 |

ROC analysis of the cross-validated step-wise logistic regression model produced an AUC of 0.847, see Figure 5.18.

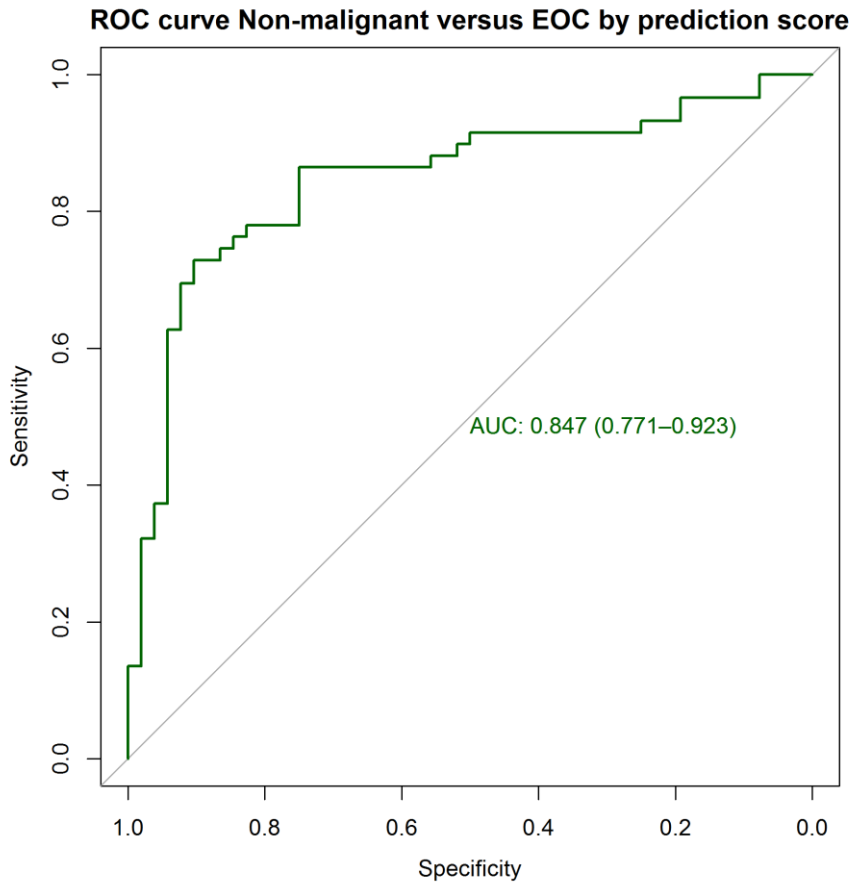



**Figure 5.18** ROC curve for non-malignant 'Normal' versus EOC by cross-validated prediction score based on step-wise logistic regression model of 5 metabolites from CPMG-$^{1}$H NMR, NOESY-presat-$^{1}$H NMR and the Biocrates Absolute*IDQ*® p150 Kit.

## 5.4 Discussion

### *5.4.1 Metabolic signature for the diagnosis of epithelial ovarian cancer*

A metabolic profile of 122 quantitated metabolites was generated for 141 plasma samples from patients with ovarian and other gynaecological cancers and benign/healthy controls using the Biocrates Absolute*IDQ*® p150 Kit. The data was analysed using multivariable and univariate statistical analysis techniques. Multivariable chemometric data analysis using principal component analysis (PCA) and partial least squares-discriminant analysis (PLS-DA) is normally used for highly multidimensional data such as that generated from untargeted LC-MS data. Little separation of classes in the scores plots was seen based on the metabolite concentrations from the biocrates targeted assay and this is most likely due to the limited number of X variables (metabolites) analysed. However it is an effective tool for quality control purposes and demonstrated that normalisation of metabolite concentrations was not necessary. Although little separation was seen between the classes, together the discriminatory performance of the metabolic signature generated

from 122 metabolites using the assay yields an AUC of 0.866 for diagnosis of EOC and Benign/Healthy controls, together termed the 'Normal' group, although importantly this calculation includes duplicate samples and therefore the AUC may be slightly overestimated. These duplicates were removed from the univariate analysis.

Univariate analysis of the 122 metabolites, with correction for multiple testing, identified 30 metabolites that were significantly different between the EOC group and the Normal group, Wilcoxon rank sum test, FDR adjusted p-value <0.05. These predictors were then incorporated into a leave one out cross-validated stepwise logistic regression model to find the best model fit for diagnosis or prediction of EOC according to AIC using only independent predictors (Pearson correlation coefficient <0.7). A combination of 5 metabolites: acetyl carnitine (C2), the amino acids; histidine (His) and tryptophan (Trp), the glycerophospholipid; PC ae C34:3 and the sphingomyelin; SM C20:2 yielded an AUC of 0.804. For all the significantly different metabolites in Figure 5.10, the most pronounced differences were consistently between healthy subjects and those with EOC, further highlighted by the correct classification of all 12 healthy subjects with valid concentrations using the cross validated predictive logistic regression model. However, patients in whom the test will be performed will be a mixture of people with benign, borderline, malignant disease and those with no ovarian disease, so the test is required to be suitably specific to rule-out all patients without the disease.

There is only a modest increase in discriminatory power when using the entire signature (AUC 0.866) compared to the 5 signature predictive model (AUC 0.804). However, the targeted plate-based kit assay from Biocrates Life Sciences AG is probably more suited to biomarker discovery than routine clinical use as batch testing for ovarian cancer diagnosis is inappropriate. A recent study by Siskos *et al.* however did demonstrate high inter-laboratory precision and quantitative accuracy and robustness, independent of the manufacture of the MS instrument (Siskos *et al.*, 2017). It would be more appropriate to develop an individual targeted assay using mass spectrometry of the 5 metabolites in the predictive model especially as 2/5 metabolites; SM C20:2 and PC ae C34:3 are only semi-quantitated in the Biocrates Absolute*IDQ*® p150 Kit assay, this would create a more robust and timely assay for routine clinical use within a diagnostic laboratory.

By combining the values from the Biocrates Absolute*IDQ*® p150 targeted assay and the NMR peak integrals from Chapter 4, a step-wise logistic regression model of 5 metabolites: histidine and unknown.1.81ppm integrals from CPMG-$^{1}$H NMR, glycoprotein acetyls integral from NOESY-presat-$^{1}$H NMR and PC.ae.C44.4 and SM.C20.2 from the Biocrates Absolute*IDQ*® p150 Kit assay increased the AUC to 0.847. A combined targeted assay or multivariate index assay as discussed in the

Introduction to this thesis can increase the sensitivity and specificity of the test, as it has in this case and provide a robust and cost effective diagnostic test for EOC.

#### 5.4.1.1 Variation in biological replicates within the study

Compared to the NMR study discussed in Chapter 4, there was no difference in classification in the duplicate samples assessed. This is likely due to the quantitative nature of the kit-based MS assay compared to integration of peaks in the NMR spectra.

However, SM C20:2 has the highest variability in concentration between biological replicates: 0.7 – 59.7 %CV, (Table 5.3) and is the metabolite with the highest number of missing values 34/246 (13.8%) of the samples included in the analysis, due to many samples having a concentration below the limit of detection (LOD). Analytical sensitivity of the assay for measuring this metabolite therefore requires improvement to ensure a valid measurement in every plasma sample.

There also appears to be a high variability in concentration of SM C20:2 between biological replicates. It is important that metabolites used for diagnosis/prognosis don't have high biological variability to allow for robust and repeatable measurement. One reason for the large variability may be due to metabolite degradation. As previously mentioned, it is well documented that metabolites are unstable and appropriate sample collection; handling and storage are required to avoid degradation of metabolites after venepuncture. Although samples in this study were collected to a defined SOP, the time between sample collection and centrifugation did vary and would have benefited from being more tightly controlled and documented so that the effect, if any, on the delay in separation could be assessed. A further validation dataset would therefore be able to address this issue and validate the findings seen in this study.

### *5.4.2 Metabolites associated with cancers of non-ovarian origin*

It is important to determine whether significant differences seen in metabolite concentrations are specific to EOC or are related to other malignancies as well. In this study, plasma samples were collected from patients undergoing gynaecological surgery and therefore, based on post-operative histological diagnosis, included patients with benign ovarian tumours, benign gynaecological disease, borderline tumours, epithelial ovarian cancer, endometrial cancer and metastatic disease from another primary origin. The group of patients with non-ovarian cancer diagnoses are termed the 'Malignant' group.

Comparison of the EOC and Malignant groups with the 'Normal' group showed a total of 23 significantly different metabolites between the two groups. 9/23 metabolites were significantly different in both groups; C2, His, PC.aa.C36.6, PC.aa.C42.6, PC.ae.C36.3, PC.ae.C38.0, PC.ae.C40.1,

PC.ae.C42.3 and SM C20:2. This suggests that this difference in concentration of these 9 metabolites is associated with malignancy rather than epithelial ovarian cancer, specifically. The other 14 significantly different metabolites between the Malignant and Normal groups are all acylcarnitines and are all higher in the malignant group then the normal group, suggesting a role of these metabolites in the pathogenesis of cancers of non-ovarian origin.

### *5.4.3 Acetyl carnitine is associated with EOC, stage of disease and survival*

Higher acetyl carnitine (C2) concentrations were observed in plasma from EOC patients and patients with other gynaecological cancers compared to benign/healthy controls. In the patients with EOC, significantly higher levels were also seen in patients with higher stage disease and lower concentrations were associated with improved overall survival. The physiological range in human plasma is 1-5µM (Engle *et al.*, 2009) which is consistent with results seen here for the non-malignant groups. However, concentrations of C2 in patients with early stage disease and stage 0 (benign/healthy) were similar, suggesting C2 is not, in isolation, a potential biomarker of early stage disease.

Carnitine can be absorbed from the diet and synthesised in the liver, brain and kidney (Siliprandi *et al.*, 1989). Acetyl carnitine has shown to have neuro-protective effects and is utilised in the treatment of peripheral neuropathies including that induced by chemotherapy. Acetyl-carnitine catalyses the transport of acyl groups across the inner mitochondrial membrane thus supporting energy production via beta-oxidation of long chain fatty acids (Bahl and Bressler, 1987). A study in serum samples from patients with multiple myeloma (MM) showed accumulation of acetyl carnitine both at diagnosis and at relapse and suggest that this could involve increased lipid oxidation in highly metabolically active myeloma cells (Lodi *et al.*, 2013). C2 has also been shown to be raised in plasma of patients with metastatic gastrointestinal cancers compared to healthy controls (Wang *et al.*, 2015). However, its lack of tumour potentiation was confirmed in an in-vitro study by Engle et al., showing lack of enhanced proliferation of OVCAR-3 and SKOV-3 cells on addition of acetyl-L-carnitine (Engle *et al.*, 2009).

### *5.4.4 Altered lipid metabolism in ovarian cancer and association with survival*

Altered lipid metabolism is recognised as an important hallmark of cancer cells. Increased *de novo* lipogenesis to fuel the unrelenting growth of cancer cells has an impact on the lipid and fatty acid composition of the entire cell (Menendez and Lupu, 2007). This potential shift between circulating lipids and fatty acids in the blood and consumption by cancer cells may be reflected in the blood of patients with cancer.

As reported in Table 1.2 in Chapter 1, lysoPC 18:2 has previously been shown to be downregulated in serum of patients with EOC compared to benign ovarian tumours/healthy controls in plasma between patients with advanced EOC pre and post-operatively (Ke *et al.*, 2016) and in plasma of patients with metastatic EOC compared to patients with benign ovarian tumours and uterine fibroids (Ke *et al.*, 2014). In the present study, lysoPC 18:2 was also shown to be significantly lower in the plasma of patients with EOC compared to patients with benign gynaecological disease or healthy controls however it was eliminated from the final logistic regression model for predicting diagnosis following stepwise additive regression. In addition, lysoPC 18:2 was shown to be significantly lower in patients with higher stage and grade of disease however; there was no association with survival in this cohort of patients. There was however, a significant negative association between lysoPC 18:0 and OS where the lower the concentration the more unfavourable the survival.

A prospective case-control study in three different cancer types; breast, prostate and colorectal identified that higher levels of lysophosphatidylcholine 18:0 were related to a lower risk of these cancers (Kuhn *et al.*, 2016) and this is consistent with another 3 case-control studies in which lower blood concentrations were found in patients with breast, prostate and colorectal cancer compared to controls (Qiu *et al.*, 2013, Osl *et al.*, 2008, Zhao *et al.*, 2007). In the present study, lyso PC 18:0 was significantly lower in patients with EOC (Figure 5.10) and a lower concentration was associated with poorer survival, Figure 5.16 C. In addition analysis of progression-free survival identified that this lysophosphatidylcholine was the only metabolite to be associated with progression-free survival where a lower concentration is again associated with poorer survival (Figure 5.17). Together the consistency of findings across multiple cancer types, including ovarian from this study suggest a global rather than cancer-type specific metabolic alteration and suggest lyso PC 18:0 may not only predict cancer survival but also cancer risk.

### *Altered amino acid metabolism in ovarian cancer and association with survival*

Histidine and tryptophan concentrations were significantly lower in plasma from EOC patients compared to plasma from patients with benign gynaecological disease and healthy controls. A lower level of histidine was also seen in patients with a cancer of non-ovarian origin compared to the control group. Both of these amino acids were also included in the predictive model for diagnosis of EOC and decreasing concentrations were also associated with increasing grade of EOC. In addition, lower levels of tryptophan were associated with increasing stage of EOC and a higher concentration was associated with improved overall survival in EOC patients.

Previous publications have identified the down-regulation of histidine in both urine (Zhang *et al.*, 2013) and plasma (Ke *et al.*, 2014) of patients with EOC compared to healthy controls. A concomitant

increase in Imidazol-5-yl-pyruvate, a metabolite of histidine in EOC patients compared to healthy controls was suggested to represent a disturbance of the histidine metabolic pathway in EOC patients which is linked to 5-phosphoribosyl-1-pyrophosphate (PRPP)-mediated nucleotide formation (Ke *et al.*, 2014).

At least four independent studies have reported a depletion of serum/plasma tryptophan concentrations in patients with ovarian cancer (Hilvo *et al.*, 2016, Ke *et al.*, 2014, Zhang *et al.*, 2012, Bachmayr-Heyda *et al.*, 2017). The likely mechanism for this is the catabolism of tryptophan to kynurenine via the enzyme indoleamine-2,3-dioxygenase (IDO) expressed in tumour cells. The result is suppression of the anti-tumour immune responses (Platten *et al.*, 2012). This up regulation of IDO is often associated with a poor prognosis in a number of cancer types (Lob *et al.*, 2009) including serous ovarian tumours (Takao *et al.*, 2007) which is supported by the negative association of tryptophan and survival seen in the present study, where lower tryptophan levels are associated with poorer survival Figure 5.16 B).

A recent study by Bachmayr-Heyda *et al*. using the Biocrates Absolute*IDQ*® p180 Kit identified a panel of 43 mainly (poly) unsaturated glycerophospholipids that were negatively associated with survival outcome (HR 0.46 p=0.007). Low concentrations of four amino acids including histidine and tryptophan were also associated with short compared to longer survival and healthy controls (Bachmayr-Heyda *et al.*, 2017). Moreover, metabolites were quantified in purified serum fractions and vesicles and the glycerophospholipid signature matched that of the LDL and vLDL signature suggesting that the metabolites measured in serum are derived from LDLs/vLDLs.

This data further supports the findings in the present study and the potential role of these biomarkers in monitoring and predicting disease outcome. In addition to biomarker identification, this data improves our understanding of the tumour biology but further analysis is required to define the mechanisms that lead to the identified changes in metabolism. Defining these processes aids in finding potential future targets for specific therapies such as inhibition of the enzyme involved in tryptophan metabolism; IDO or preventing tumour cell uptake of glycerophospholipids from (v)LDLs.

## 5.5 Summary

Targeted MS-based analysis of plasma samples from patients with EOC and those without using the Biocrates AbsoluteIDQ® p150 kit has identified potential blood based metabolic biomarkers for diagnosis and prognosis including acetyl carnitine (C2), tryptophan and lysoPC 18:0. Lower levels of tryptophan may be due to increased catabolism by the IDO enzyme in tumour cells and may be associated with the T-cell mediated immune response. Lysophosphatidylcholine 18:0 levels were lower in patients with EOC compared to non-affected controls and a number of other publications

suggest that this metabolite may not only be associated with ovarian cancer but other cancers and may be associated with cancer risk. By combining results from the targeted MS assay and the NMR peak integrals, the AUC was 0.847 with an associated sensitivity of 78% and specificity of 77%. Further validation in an independent dataset is required with samples collected under strict protocols.

# CHAPTER 6. General Discussion

## 6.1 Circulating tumour DNA methylation biomarkers and future directions

The work presented in this thesis has demonstrated that differentially methylated loci identified from bioinformatics analysis of genome wide methylation studies can be detected in cell-free circulating DNA in plasma using bisulphite pyrosequencing as a diagnostic biomarker of serous ovarian cancer. One of the methylation markers, cg16927606 in *TMEM149* correctly diagnosed 100% of women who did not have the disease. In contrast, 79% of patients with the disease were correctly classified.

Biomarkers used for screening a population should have a high sensitivity to minimise the number of false negative results, ensuring that a test doesn't rule out a diagnosis when in fact that individual has the disease. Reports on the sensitivity and specificity for CA 125 as a diagnostic biomarker vary widely with a sensitivity ranging from 61% to 90% and specificity ranging from 35% to 91% for distinguishing benign from malignant masses (Pepin *et al.*, 2014). Although it appears that results for ctDNA methylation detection in plasma are as sensitive and specific as CA 125 if not slightly better, there are a number of caveats to this. Firstly, the number of samples analysed in the study was small as plasma samples are required to be centrifuged twice, so-called 'double-spun' and therefore they had to be prospectively collected and could not be sourced from the bio bank, at that time. Secondly, there are limitations in terms of the analytical sensitivity of pyrosequencing where a number of samples did not have a detectable signal for cfDNA in plasma distinguishable from background noise, however this was assay dependent.

One of the markers identified from the bioinformatics analysis was methylation of *ZNF154*. During the course of this PhD another group also identified methylation of the exact same CpG site to be differently methylated in a number of cancers and proposed its use as a pan cancer biomarker (Sanchez-Vega *et al.*, 2013, Margolin *et al.*, 2016). Through in silico modelling, they also proposed that bisulphite sequencing would be able to detect ctDNA in a background of 99% non-tumour DNA (Margolin *et al.*, 2016). This study, however, is the first study to show the measurement of *ZNF154* methylation in blood in vitro.

### *6.1.1 Limitations and future work*

As previously eluded to the number of double spun plasma samples analysed in the study was too small to make firm conclusions based on the diagnostic power of the methylation markers. However, this was a proof of principle study to determine if bioinformatics analysis of methylation in tumour and non-tumour DNA could be translated into ctDNA biomarkers in plasma. Moving the project forward, plasma samples collected at ICHNT are now double-spun before storing in the bank thereby enabling further studies to be conducted on ctDNA in the future. In addition a collaboration with

Agena Bioscience and the Institute of Cancer Research will assess the performance of another method that combines restriction enzyme digestion with the epiTYPER® that can simultaneously sequence multiple CpG sites. This equipment was also recently purchased at ICHNT and therefore further analysis of future samples will also be conducted on site. Furthermore, a power calculation could be performed to identify the number of samples needed to achieve statistically significant results and this could help guide the number of samples required for future validation studies.

## 6.2 Metabolic biomarkers and future directions

In this thesis, untargeted and targeted metabolic profiling identified differences in beta-hydroxybutyrate and other ketone metabolites in the plasma of patients with EOC. Furthermore, beta-hydroxybutyrate was associated with survival; the higher the concentration of the metabolite, the worse the survival. During the course of this PhD another research group using a metabolomics based platform, GC-MS, also identified beta-hydroxybutyrate and ketone metabolites as being significantly different in patients with EOC (Hilvo *et al.*, 2016). The authors attempted to correlate the findings with the tumour biology and suggested that the increase in ketone metabolites was linked to SSADH deficiency, loss of *ALDH5A1* and epithelial to mesenchymal transition. However, this only relates to the dihydroxybutyric acid metabolites, which are linked to gamma-hydroxybutryic acids (GHB) in the body not beta-hydroxybutyrate, see Figure 6.1. The resultant increase in gamma-amino- and hydroxy-butyric acids due to SSADH deficiency also leads to a reduction in succinate; see Figure 6.1 which contradicts the finding of succinate as an 'oncometabolite'. Succinate has been shown to inhibit 2-oxoglutarate-dependent oxygenases which activates oncogenic HIF pathways and can also alter DNA methylation affecting the expression of genes involved in cell differentiation (Yang *et al.*, 2013). However, beta-hydroxybutyrate metabolism can increase acetoacetate which can further increase succinate contributing to this potential pathway in tumourigenesis. Further analysis is needed to prove the causal association between beta-hydroxybutyrate and ovarian cancer and the association with poor survival and whether this is linked to succinate and the TCA cycle. The metabolic pathways involving succinate, beta-hydroxybutyrate, gamma-aminobutyric acid (GABA) and gamma-hydroxybutyric acid are shown in Figure 6.1.

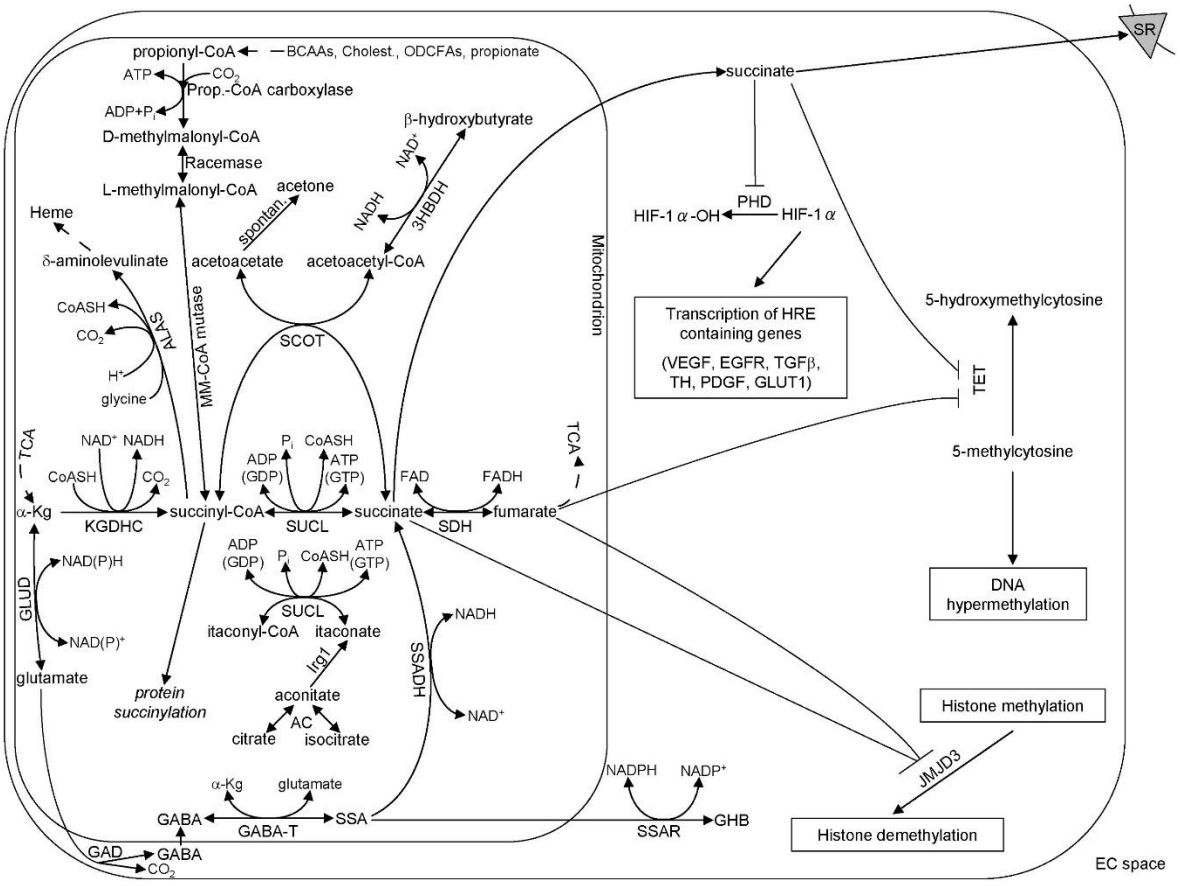



Figure 6.1. A schematic illustrating the metabolism of hydroxybutyric acid metabolites and succinate in metabolic and signal transduction pathways. (Reproduced with permission from (Tretter *et al.*, 2016).

ALAS: δ-aminolevulinate synthase, PHD: prolyl hydroxylases, TET: ten-eleven translocation (TET) family of 5-methylcytosine (5mC) hydroxylases, JMJD3: Jumonji C domain-containing histone lysine demethylases, VEGF: vascular endothelial growth factor, EGFR: epidermal growth factor receptor, TGFβ: transforming growth factor β, TH: tyrosine hydroxylase, PDGF: platelet-derived growth factor, GLUT1: glucose transporter 1, BCAAs: branched-chain amino acids, ODFCAs: odd-number of chain fatty acids, Cholest.: cholesterol, 3HBDH: 3-hydroxybutyrate dehydrogenase, KGDHC: ketoglutarate dehydrogenase complex, GLUD: glutamate dehydrogenase, GAD: glutamate decarboxylase, GABA: γ-aminobutyric acid, EC space: extracellular space, GABA-T: GABA transaminase, SSA: succinate semialdehyde, SSADH: succinate semialdehyde dehydrogenase, SSAR: succinate semialdehyde reductase, GHB: γ-hydroxybutyric acid, SDH: succinate dehydrogenase; SCOT: succinyl-CoA:3-ketoacid coenzyme A transferase 1, SR: succinate receptor.

Oncometabolites tend to be assumed to be increased in malignancy and their accumulation initiates or sustains tumour growth and metastasis (Wishart, 2016). However, it could be argued that a sustained increase in one metabolite will feedback to another causing a concomitant decrease, therefore oncometabolites could be increased or decreased in malignancy depending on the function of the metabolite. An example of this is the decrease in tryptophan seen in EOC in this and other studies (Hilvo *et al.*, 2016, Ke *et al.*, 2014, Zhang *et al.*, 2012, Bachmayr-Heyda *et al.*, 2017). Tryptophan was also associated with overall survival in this study. The likely mechanism for this is the catabolism of tryptophan to kynurenine via the enzyme indoleamine-2,3-dioxygenase (IDO) expressed in tumour cells. This up regulation of IDO has been associated with a poor prognosis in serous ovarian carcinoma (Takao *et al.*, 2007) which is supported by the negative association of tryptophan and survival seen in the present study, where lower tryptophan levels are associated with poorer survival. Clinical trials have already been initiated investigating IDO inhibitors in combination with PDL-1 Inhibitors in melanoma (ClinicalTrials.gov Identifier NCT02073123). IDO like PDL-1 are pathways used by tumours to evade T cell detection. However, it doesn't appear that levels of tryptophan or kyneurine are measured before treatment. It would be interesting to determine the difference in response to the drug between the two cohorts stratified by tryptophan concentration as the work presented here has shown a difference in overall survival between patients with tryptophan concentrations above and below the median concentration as this may have an impact on the overall results from the trial. If the hypothesis that patients with lower tryptophan levels pre-treatment do better than patients with higher levels, this would impact on who received the treatment and move towards personalising treatment based on the tumour biology.

### *6.2.1 Limitations and future work*

To be truly a systems biology approach, results from the metabonomics study would be correlated to data from other 'omics'. For example as with the study of dihydroxybutyric acid metabolites and SSADH/ALDH5A1 and EMT, data from transcriptomics of tumour tissue could be related to the metabolic profile to try and prove this hypothesis. At ICHNT tissue bank a number of frozen tissue samples relating to patients in this study would have been stored for future use and could be used to further elucidate the functional role of these metabolites using a hypothesis-driven approach.

A relatively large proportion of the samples from patients undergoing gynaecological surgery at Hammersmith Hospital were from patients with another gynaecological malignancy such as cancer arising from the endometrium. Although this decreased the overall number of samples included in the primary analysis, this could be used as an alternative 'control' group to assess whether the changes in the metabolites were related to malignancy or EOC specifically. As this was outside the

original hypothesis, limited clinical data was available on these patients. However, in the future clinical data on survival of these patients will be identified in a clinical database set up within the ECMC and the findings in the EOC cohort can be validated in the 'Malignant' cohort.

Although STOCSY analysis positively identified the ketone metabolites: beta-hydroxybutyrate, acetone and acetoacetate, further spike-in experiments of pure standards to confirm these identities could have been performed. In addition, to validate the findings of the NMR and Biocrates study, another cohort, or a larger cohort split into a training and validation set could have given more confidence to the findings of this study. However, another collaborator within the Division of Cancer has performed untargeted metabolic profiling using $^{1}$H NMR on an independent sample collection from another centre in another country. Preliminary analysis of this dataset, not shown here, identifies the same discriminatory metabolites that were found in this study further supporting the association of these metabolites with EOC.

Although the number of samples in this pilot study was small, it was sufficient to identify trends and associations between the groups and further validation studies should use multivariate power calculations to identify the number of samples required for analysis to achieve statistically significant results. Furthermore, the benign and healthy groups within the study were combined to reflect the clinical setting as not all patients without disease will have a benign pathology. Both multivariate and univariate analysis of metabolic data suggest the healthy group demonstrate larger variation in metabolites compared to the EOC group, than the benign, and this may introduce bias into the 'Normal' population when compared to the EOC group. Unfortunately due to availability of volunteers the 'healthy' group were younger with an average age of 30 compared to the benign and EOC groups where the average age was 50-60 respectively. Future studies should focus on collecting samples from an appropriate control population where study participants are matched on age and menopausal status, have not had a hysterectomy and are deemed at higher risk e.g. low parity.

### 6.3 Biomarker study limitations

Correct study design, effective and robust bio banking and appropriate controls are important in studies identifying and validating biomarkers of disease. In this study, both DNA methylation and metabonomics heavily relied upon effective bio banking through either double-spinning samples or timely pre-analytical centrifugation and storage. Post-analytical analysis of the clinical data identified a large proportion of samples that were deemed to be inappropriate, due to delayed processing, histological subtype or they were not taken pre-operatively. The histology cannot be confirmed pre-analysis but tighter control on sample bio banking is imperative to avoid analytical variation as shown in the NMR and Biocrates studies in Chapters 4 and 5. The metabolome is particularly

susceptible to changes due to their labile and perishable nature that could give rise to false positive/negative results that are secondary to degradation after collection.

There is also a need to compare the candidate biomarkers with the current gold standard in this case CA 125. Unfortunately, limited CA 125 data was available due to the fact that ICHNT is a tertiary referral centre, a number of patients had previously had CA 125 measured at their local hospitals and therefore this data was not available at the time of the study. Ideally CA 125 needs to be assayed at the same time point as the candidate biomarkers for direct comparison of its diagnostic and prognostic capabilities. CA 125 in combination with other biomarkers may increase the sensitivity of the markers compared to assessing them in isolation.

Poor study design, lack of reproducibility and incomplete reporting has led to a large number of biomarker studies being published but never reaching the clinic (Harris *et al.*, 2007). This led to guidelines being adopted that recommend studies should include: a clear objective, the patient characteristics, the material and analytical and statistical methods used, termed the REporting of tumour MARKer prognostic studies (REMARK) (Mcshane *et al.*, 2005). Future biomarker studies should be designed to conform to Reporting of tumour MARKer prognostic studies (REMARK) guidelines as set out in (Altman *et al.*, 2012).

Many biomarker studies, including this one, lack a causal association or link between the biomarker and the underpinning tumour biology past just correlation. Although the primary aim of the study was to identify biomarkers, follow up validation studies should aim to identify the mechanism of increase, or decrease in the biomarkers. Identifying the functional role of the biomarkers in cancer initiation, progression, metastasis and other aspects will help to increase confidence in use of the marker, identify novel therapeutic targets and help speed up the validation and translation to the clinic.

### 6.4 Concluding remarks

More sensitive biomarkers for diagnosis of ovarian cancer are a clinical priority. This study has identified a number of potential biomarkers of EOC including ctDNA methylation and metabolite concentrations in plasma that are not only diagnostic but are also associated with survival. In an era where we are striving towards personalised medicine, we can build on studies such as this to further validate these markers in larger cohorts and combine markers to form a 'multiomic' liquid biopsy consisting of DNA alterations: methylation, mutations, copy number variation, microRNAs, proteins and metabolites. Panels of biomarkers with more sensitive detection methods will lead to a better understanding of disease pathogenesis which should in turn identify targets for drug development and combined with early detection will ultimately lead to better outcome for patients.

# Chapter 7. References


Abbosh, C., Birkbak, N. J., Wilson, G. A., Jamal-Hanjani, M., Constantin, T., Salari, R., Le Quesne, J., Moore, D. A., Veeriah, S., Rosenthal, R., Marafioti, T., Kirkizlar, E., Watkins, T. B. K., Mcgranahan, N., Ward, S., Martinson, L., Riley, J., Fraioli, F., Al Bakir, M., Gronroos, E., Zambrana, F., Endozo, R., Bi, W. L., Fennessy, F. M., Sponer, N., Johnson, D., Laycock, J., Shafi, S., Czyzewska-Khan, J., Rowan, A., Chambers, T., Matthews, N., Turajlic, S., Hiley, C., Lee, S. M., Forster, M. D., Ahmad, T., Falzon, M., Borg, E., Lawrence, D., Hayward, M., Kolvekar, S., Panagiotopoulos, N., Janes, S. M., Thakrar, R., Ahmed, A., Blackhall, F., Summers, Y., Hafez, D., Naik, A., Ganguly, A., Kareht, S., Shah, R., Joseph, L., Marie Quinn, A., Crosbie, P. A., Naidu, B., Middleton, G., Langman, G., Trotter, S., Nicolson, M., Remmen, H., Kerr, K., Chetty, M., Gomersall, L., Fennell, D. A., Nakas, A., Rathinam, S., Anand, G., Khan, S., Russell, P., Ezhil, V., Ismail, B., Irvin-Sellers, M., Prakash, V., Lester, J. F., Kornaszewska, M., Attanoos, R., Adams, H., Davies, H., Oukrif, D., Akarca, A. U., Hartley, J. A., Lowe, H. L., Lock, S., Iles, N., Bell, H., Ngai, Y., Elgar, G., Szallasi, Z., Schwarz, R. F., Herrero, J., Stewart, A., Quezada, S. A., Peggs, K. S., Van Loo, P., Dive, C., Lin, C. J., Rabinowitz, M., Aerts, H.*, et al.* (2017). Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature,* **545**(7655): 446-451.

Ahmad, M. S., Alsaleh, M., Kimhofer, T., Ahmad, S., Jamal, W., Wali, S. O., Nicholson, J. K., Damanhouri, Z. A. & Holmes, E. (2017). Metabolic Phenotype of Obesity in a Saudi Population. *J Proteome Res,* **16**(2): 635-644.

Ahmed, A. A., Etemadmoghadam, D., Temple, J., Lynch, A. G., Riad, M., Sharma, R., Stewart, C., Fereday, S., Caldas, C., Defazio, A., Bowtell, D. & Brenton, J. D. (2010). Driver mutations in TP53 are ubiquitous in high grade serous carcinoma of the ovary. *J Pathol,* **221**(1): 49-56.

Allis, C. D. & Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. *Nat Rev Genet,* **17**(8): 487-500.

Altman, D. G., Mcshane, L. M., Sauerbrei, W. & Taube, S. E. (2012). Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): explanation and elaboration. *PLoS Med,* **9**(5): e1001216.

Andreopoulos, B. & Anastassiou, D. (2012). Integrated Analysis Reveals hsa-miR-142 as a Representative of a Lymphocyte-Specific Gene Expression and Methylation Signature. *Cancer Inform,* **11**: 61-75.

Archer, T. C., Fertig, E. J., Gosline, S. J., Hafner, M., Hughes, S. K., Joughin, B. A., Meyer, A. S., Piccolo, S. R. & Shajahan-Haq, A. N. (2016). Systems Approaches to Cancer Biology. *Cancer Res,* **76**(23): 6774-6777.

Aru, V., Lam, C., Khakimov, B., Hoefsloot, H. C. J., Zwanenburg, G., Lind, M. V., Schäfer, H., Van Duynhoven, J., Jacobs, D. M., Smilde, A. K. & Engelsen, S. B. (2017). Quantification of lipoprotein profiles by nuclear magnetic resonance spectroscopy and multivariate data analysis. *TrAC Trends in Analytical Chemistry,* **94**: 210-219.

Bachmayr-Heyda, A., Aust, S., Auer, K., Meier, S. M., Schmetterer, K. G., Dekan, S., Gerner, C. & Pils, D. (2017). Integrative Systemic and Local Metabolomics with Impact on Survival in High-Grade Serous Ovarian Cancer. *Clin Cancer Res,* **23**(8): 2081-2092.

Badgwell, D. & Bast, R. C., Jr. (2007). Early detection of ovarian cancer. *Disease Markers,* **23**: 397-410.

Bahl, J. J. & Bressler, R. (1987). The pharmacology of carnitine. *Ann Rev Pharmacol Toxicol,* **27**: 257-277.

Bast, R. C., Jr., Feeney, M., Lazarus, H., Nadler, L. M., Colvin, R. B. & Knapp, R. C. (1981). Reactivity of a monoclonal antibody with human ovarian carcinoma. *J Clin Invest,* **68**: 1331-1337.

Bast, R. C., Jr., Klug, T. L., St John, E., Jenison, E., Niloff, J. M., Lazarus, H., Berkowitz, R. S., Leavitt, T., Griffiths, C. T., Parker, L., Zurawski, V. R., Jr. & Knapp, R. C. (1983). A radioimmunoassay using a monoclonal antibody to monitor the course of epithelial ovarian cancer. *N Engl J Med,* **309**(15): 883-887.

Bast, R. C., Urban, N., Shridhar, V., Smith, D., Zhang, Z., Skates, S., Lu, K., Liu, J., Fishman, D. & Mills, G. (2002). Early Detection of Ovarian Cancer: Promise and Reality. 61-97.

Beckonert, O., Keun, H. C., Ebbels, T. M., Bundy, J., Holmes, E., Lindon, J. C. & Nicholson, J. K. (2007). Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat Protoc,* **2**(11): 2692-703.

Belinsky, S. A., Nikula, K. J., Palmisano, W. A., Michels, R., Saccomanno, G., Gabrielson, E., Baylin, S. & Herman, J. G. (1998). Aberrant methylation of p16INK4a is an early event in lung cancer and a potential biomarker for early diagnosis. *Proc Natl Acad Sci U S A,* **95**: 11891-11896.

Bettegowda, C., Sausen, M., Leary, R. J., Kinde, I., Wang, Y., Agrawal, N., Bartlett, B. R., Wang, H., Luber, B., Alani, R. M., Antonarakis, E. S., Azad, N. S., Bardelli, A., Brem, H., Cameron, J. L., Lee, C. C., Fecher, L. A., Gallia, G. L., Gibbs, P., Le, D., Giuntoli, R. L., Goggins, M., Hogarty, M. D., Holdhoff, M., Hong, S. M., Jiao, Y., Juhl, H. H., Kim, J. J., Siravegna, G., Laheru, D. A., Lauricella, C., Lim, M., Lipson, E. J., Marie, S. K., Netto, G. J., Oliner, K. S., Olivi, A., Olsson, L., Riggins, G. J., Sartore-Bianchi, A., Schmidt, K., Shih L, M., Oba-Shinjo, S. M., Siena, S., Theodorescu, D., Tie, J., Harkins, T. T., Veronese, S., Wang, T. L., Weingart, J. D., Wolfgang, C. L., Wood, L. D., Xing, D., Hruban, R. H., Wu, J., Allen, P. J., Schmidt, C. M., Choti, M. A., Velculescu, V. E., Kinzler, K. W., Vogelstein, B., Papadopoulos, N. & Diaz, L. A., Jr. (2014). Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med,* **6**(224): 224ra24.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev,* **16**(1): 6-21.

Bird, A. P. (1986). CpG-rich islands and the function of DNA methylation. *Nature,* **321**: 209-213.

Bonito, N. A., Borley, J., Wilhelm-Benartzi, C. S., Ghaem-Maghami, S. & Brown, R. (2016). Epigenetic Regulation of the Homeobox Gene MSX1 Associates with Platinum-Resistant Disease in High-Grade Serous Epithelial Ovarian Cancer. *Clin Cancer Res,* **22**(12): 3097-3104.

Borley, J., Wilhelm-Benartzi, C., Brown, R. & Ghaem-Maghami, S. (2012). Does tumour biology determine surgical success in the treatment of epithelial ovarian cancer? A systematic literature review. *Br J Cancer,* **107**(7): 1069-74.

Boroughs, L. K. & Deberardinis, R. J. (2015). Metabolic pathways promoting cancer cell survival and growth. *Nat Cell Biol,* **17**(4): 351-9.

Bothwell, J. H. & Griffin, J. L. (2011). An introduction to biological nuclear magnetic resonance spectroscopy. *Biol Rev Camb Philos Soc,* **86**(2): 493-510.

Boulesteix, A. L. (2004). PLS dimension reduction for classification with microarray data. *Stat Appl Genet Mol Biol,* **3**: Article33.

Bristow, R. E., Tomacruz, R. S., Armstrong, D. K., Trimble, E. L. & Montz, F. J. (2002). Survival effect of maximal cytoreductive surgery for advanced ovarian carcinoma during the platinum era: a meta-analysis. *J Clin Oncol,* **20**(5): 1248-1259.

Buas, M. F., Gu, H., Djukovic, D., Zhu, J., Drescher, C. W., Urban, N., Raftery, D. & Li, C. I. (2016). Identification of novel candidate plasma metabolite biomarkers for distinguishing serous ovarian carcinoma and benign serous ovarian tumors. *Gynecol Oncol,* **140**(1): 138-44.

Buescher, J. M. & Driggers, E. M. (2016). Integration of omics: more than the sum of its parts. *Cancer Metab,* **4**: 4.

Campan, M., Moffitt, M., Houshdaran, S., Shen, H., Widschwendter, M., Daxenbichler, G., Long, T., Marth, C., Laird-Offringa, I. A., Press, M. F., Dubeau, L., Siegmund, K. D., Wu, A. H., Groshen, S., Chandavarkar, U., Roman, L. D., Berchuck, A., Pearce, C. L. & Laird, P. W. (2011). Genome-scale screen for DNA methylation-based detection markers for ovarian cancer. *PLoS One,* **6**(12): e28141.

Cancer Research Uk. *http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer/incidence#heading-One* [Online]. [Accessed August 2017].

Cancer Research Uk. *http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer/incidence#heading-Three* [Online]. [Accessed August 2017].

Cancer Research Uk. *http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer/survival#heading-Three* [Online]. [Accessed August 2017].

Carmena, R., Duriez, P. & Fruchart, J. C. (2004). Atherogenic lipoprotein particles in atherosclerosis. *Circulation,* **109**(23 Suppl 1): III2-7.

Chan, K. C., Jiang, P., Zheng, Y. W., Liao, G. J., Sun, H., Wong, J., Siu, S. S., Chan, W. C., Chan, S. L., Chan, A. T., Lai, P. B., Chiu, R. W. & Lo, Y. M. (2013). Cancer genome scanning in plasma: detection of tumor-associated copy number aberrations, single-nucleotide variants, and tumoral heterogeneity by massively parallel sequencing. *Clin Chem,* **59**(1): 211-24.

Chen, J., Zhang, X., Cao, R., Lu, X., Zhao, S., Fekete, A., Huang, Q., Schmitt-Kopplin, P., Wang, Y., Xu, Z., Wan, X., Wu, X., Zhao, N., Xu, C. & Xu, G. (2011). Serum 27-nor-5beta-cholestane-3,7,12,24,25 pentol glucuronide discovered by metabolomics as potential diagnostic biomarker for epithelium ovarian cancer. *J Proteome Res,* **10**(5): 2625-32.

Chen, J., Zhou, L., Zhang, X., Lu, X., Cao, R., Xu, C. & Xu, G. (2012). Urinary hydrophilic and hydrophobic metabolic profiling based on liquid chromatography-mass spectrometry methods: Differential metabolite discovery specific to ovarian cancer. *Electrophoresis,* **33**(22): 3361-9.

Chen, V. W., Ruiz, B., Killeen, J. L., Cote, T. R., Wu, X. C. & Correa, C. N. (2003). Pathology and classification of ovarian tumors. *Cancer,* **97**(10 Suppl): 2631-42.

Chen, W. D., Han, Z. J., Skoletsky, J., Olson, J., Sah, J., Myeroff, L., Platzer, P., Lu, S., Dawson, D., Willis, J., Pretlow, T. P., Lutterbaugh, J., Kasturi, L., Willson, J. K., Rao, J. S., Shuber, A. & Markowitz, S.

D. (2005). Detection in fecal DNA of colon cancer-specific methylation of the nonexpressed vimentin gene. *J Natl Cancer Inst,* **97**(15): 1124-32.

Cheng, Y., Li, L., Zhu, B., Liu, F., Wang, Y., Gu, X. & Yan, C. (2016). Expanded metabolomics approach to profiling endogenous carbohydrates in the serum of ovarian cancer patients. *J Sep Sci,* **39**(2): 316-23.

Chi, D. S., Eisenhauer, E. L., Lang, J., Huh, J., Haddad, L., Abu-Rustum, N. R., Sonoda, Y., Levine, D. A., Hensley, M. & Barakat, R. R. (2006). What is the optimal goal of primary cytoreductive surgery for bulky stage IIIC epithelial ovarian carcinoma (EOC)? *Gynecol Oncol,* **103**(2): 559-64.

Cloarec, O., Dumas, M., Craig, A., Barton, R. H., Trygg, J., Hudson, J., Blancher, C., Gauguier, D., Lindon, J. C., Holmes, E. & Nicholson, J. K. (2005a). Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic 1H NMR data sets. *Anal. Chem,* **77**: 1282-1289.

Cloarec, O., Dumas, M., Trygg, J., Craig, A., Barton, R. H., Lindon, J. C., Nicholson, J. K. & Holmes, E. (2005b). Evaluation of the orthogonal projection on latent structure model limitations caused by chemical shift variability and improved visualisation of biomarker changes in 1H NMR spectroscopic metabonomic studies. *Anal. Chem,* **77**: 517-526.

Coleman, R. L., Herzog, T. J., Chan, D. W., Munroe, D. G., Pappas, T. C., Smith, A., Zhang, Z. & Wolf, J. (2016). Validation of a second-generation multivariate index assay for malignancy risk of adnexal masses. *Am J Obstet Gynecol,* **215**(1): 82 e1-82 e11.

Costa-Silva, B., Aiello, N. M., Ocean, A. J., Singh, S., Zhang, H., Thakur, B. K., Becker, A., Hoshino, A., Mark, M. T., Molina, H., Xiang, J., Zhang, T., Theilen, T. M., Garcia-Santos, G., Williams, C., Ararso, Y., Huang, Y., Rodrigues, G., Shen, T. L., Labori, K. J., Lothe, I. M., Kure, E. H., Hernandez, J., Doussot, A., Ebbesen, S. H., Grandgenett, P. M., Hollingsworth, M. A., Jain, M., Mallya, K., Batra, S. K., Jarnagin, W. R., Schwartz, R. E., Matei, I., Peinado, H., Stanger, B. Z., Bromberg, J. & Lyden, D. (2015). Pancreatic cancer exosomes initiate pre-metastatic niche formation in the liver. *Nat Cell Biol,* **17**(6): 816-26.

Dawson, S. J., Tsui, D. W., Murtaza, M., Biggs, H., Rueda, O. M., Chin, S. F., Dunning, M. J., Gale, D., Forshew, T., Mahler-Araujo, B., Rajan, S., Humphray, S., Becq, J., Halsall, D., Wallis, M., Bentley, D., Caldas, C. & Rosenfeld, N. (2013). Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N Engl J Med,* **368**(13): 1199-209.

Deaton, A. M. & Bird, A. (2011). CpG islands and the regulation of transcription. *Genes Dev,* **25**(10): 1010-22.

Deberardinis, R. J. & Chandel, N. S. (2016). Fundamentals of cancer metabolism. *Sci Adv,* **2**: e1600200.

Denkert, C., Budczies, J., Kind, T., Weichert, W., Tablack, P., Sehouli, J., Niesporek, S., Konsgen, D., Dietel, M. & Fiehn, O. (2006). Mass spectrometry-based metabolic profiling reveals different metabolite patterns in invasive ovarian carcinomas and ovarian borderline tumors. *Cancer Res,* **66**(22): 10795-804.

Devos, T., Tetzner, R., Model, F., Weiss, G., Schuster, M., Distler, J., Steiger, K. V., Grutzmann, R., Pilarsky, C., Habermann, J. K., Fleshner, P. R., Oubre, B. M., Day, R., Sledziewski, A. Z. & Lofton-Day, C. (2009). Circulating methylated SEPT9 DNA in plasma is a biomarker for colorectal cancer. *Clin Chem,* **55**(7): 1337-46.

Diaz-Lagares, A., Mendez-Gonzalez, J., Hervas, D., Saigi, M., Pajares, M. J., Garcia, D., Crujerias, A. B., Pio, R., Montuenga, L. M., Zulueta, J., Nadal, E., Rosell, A., Esteller, M. & Sandoval, J. (2016). A Novel Epigenetic Signature for Early Diagnosis in Lung Cancer. *Clin Cancer Res*.

Diaz, L. A., Jr. & Bardelli, A. (2014). Liquid biopsies: genotyping circulating tumor DNA. *J Clin Oncol,* **32**(6): 579-86.

Dibble, C. C. & Manning, B. D. (2013). Signal integration by mTORC1 coordinates nutrient input with biosynthetic output. *Nat Cell Biol,* **15**(6): 555-64.

Diehl, F., Li, M., Dressman, D., He, Y., Shen, D., Szabo, S., Diaz, L. A., Jr., Goodman, S. N., David, K. A., Juhl, H., Kinzler, K. W. & Vogelstein, B. (2005). Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proc Natl Acad Sci U S A,* **102**(45): 16368-73.

Diehl, F., Li, M., He, Y., Kinzler, K. W., Vogelstein, B. & Dressman, D. (2006). BEAMing: single-molecule PCR on microparticles in water-in-oil emulsions. *Nat Methods,* **3**(7): 551-9.

Diehl, F., Schmidt, K., Choti, M. A., Romans, K., Goodman, S., Li, M., Thornton, K., Agrawal, N., Sokoll, L., Szabo, S. A., Kinzler, K. W., Vogelstein, B. & Diaz, L. A., Jr. (2008). Circulating mutant DNA to assess tumor dynamics. *Nat Med,* **14**(9): 985-90.

Dogan, M. V., Shields, B., Cutrona, C., Gao, L., Gibbons, F. X., Simons, R., Monick, M., Brody, G. H., Tan, K., Beach, S. R. H. & Philibert, R. A. (2014). The effect of smoking on DNA methylation of peripheral blood mononuclear cells from African American women. *BMC Genomics,* **15**(151).

Dona, A. C., Jimenez, B., Schafer, H., Humpfer, E., Spraul, M., Lewis, M. R., Pearce, J. T., Holmes, E., Lindon, J. C. & Nicholson, J. K. (2014). Precision high-throughput proton NMR spectroscopy of human urine, serum, and plasma for large-scale metabolic phenotyping. *Anal Chem,* **86**(19): 9887-94.

Drexler, H. G. (1998). Review of alterations of the cyclin-dependent kinase inhibitor INK4 family gene p15, p16, p18 and p19 in human leukaemia-lymphoma cells. *Leukaemia,* **12**: 845-859.

Du, H. & Taylor, H. S. (2004). Molecular regulation of mullerian development by Hox genes. *Ann N Y Acad Sci,* **1034**: 152-65.

Dunn, W. B., Broadhurst, D. I., Atherton, H. J., Goodacre, R. & Griffin, J. L. (2011). Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chem Soc Rev,* **40**(1): 387-426.

Eads, C. A., Danenberg, K. D., Kawakami, K., Saltz, L. B., Blake, C., Shibata, D., Danenberg, P. V. & Laird, P. W. (2000). Methylight: a high-throughput assay to measure dna methylation. *Nucleic Acids Res,* **28**(8).

Egger, G., Liang, G., Aparicio, A. & Jones, P. A. (2004). Epigenetics in human disease and prospects for epigenetic therapy. *Nature,* **429**: 457-463.

Eipel, M., Mayer, F., Arent, T., Ferreira, M. R. P., Birkhofer, C., Gerstenmaier, U., Costa, I. G., Ritz-Timme, S. & Wagner, W. (2016). Epigenetic age predictions based on buccal swabs are more precise in combination with cell type-specific DNA methylation signatures. *Aging,* **8**(5): 1034-1044.

Engle, D. B., Belisle, J. A., Gubbels, J. a. A., Petrie, S. E., Hutson, P. R., Kushner, D. M. & Patankar, M. S. (2009). Effect of acetyl-L-carnitine on ovarian cancer cells' proliferation, nerve growth factor

receptor (Trk-A and p75) expression, and the cytotoxic potential of paclitaxel and carboplatin. *Gynecol Oncol,* **112**(3): 631-636.

Esteller, M. (2008). Epigenetics in cancer. *N Engl J Med,* **358**: 1148-1159.

Esteller, M., Hamilton, S. R., Burger, P. C., Baylin, S. & Herman, J. G. (1999). Inactivation of the DNA repair gene O6-Methylguanine-DNA Methyltransferase by promoter hypermethylation is a common event in primary human neoplasia. *Cancer Res,* **59**: 793-797.

Fan, L., Yin, M., Ke, C., Ge, T., Zhang, G., Zhang, W., Zhou, X., Lou, G. & Li, K. (2016). Use of Plasma Metabolomics to Identify Diagnostic Biomarkers for Early Stage Epithelial Ovarian Cancer. *J Cancer,* **7**(10): 1265-72.

Fan, L., Zhang, W., Yin, M., Zhang, T., Wu, X., Zhang, H., Sun, M., Li, Z., Hou, Y., Zhou, X., Lou, G. & Li, K. (2012). Identification of metabolic biomarkers to diagnose epithelial ovarian cancer using a UPLC/QTOF/MS platform. *Acta Oncol,* **51**(4): 473-9.

Feinberg, A. P. & Vogelstein, B. (1983). Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature,* **301**(89-92).

Fleischhacker, M. & Schmidt, B. (2007). Circulating nucleic acids (CNAs) and cancer--a survey. *Biochim Biophys Acta,* **1775**(1): 181-232.

Fleisher, A. S., Esteller, M., Wang, S., Tamura, G., Suzuki, H., Yin, J., Zou, T., Abraham, J. M., Kong, D., Smolinski, K. N., Shi, Y., Rhyu, M., Powell, S. M., James, S. P., Wilson, K. T., Herman, J. G. & Meltzer, S. J. (1999). Hypermethylation of the hmlh1 gene promoter in human gastric cancers with microsatellite instability. *Cancer Res,* **59**: 1090-1095.

Fong, M. Y., Mcdunn, J. & Kakar, S. S. (2011). Identification of metabolites in the normal ovary and their transformation in primary and metastatic ovarian cancer. *PLoS One,* **6**(5): e19963.

Forshew, T., Murtaza, M., Parkinson, C., Gale, D., Tsui, D. W., Kaper, F., Dawson, S. J., Piskorz, A. M., Jimenez-Linan, M., Bentley, D., Hadfield, J., May, A. P., Caldas, C., Brenton, J. D. & Rosenfeld, N. (2012). Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci Transl Med,* **4**(136): 1-12.

Fridley, B. L., Armasu, S. M., Cicek, M. S., Larson, M. C., Wang, C., Winham, S. J., Kalli, K. R., Koestler, D. C., Rider, D. N., Shridhar, V., Olson, J. E., Cunningham, J. M. & Goode, E. L. (2014). Methylation of leukocyte DNA and ovarian cancer: relationships with disease status and outcome. *BMC Med Genomics,* **7**(21): 1-12.

Galluzzi, L., Pietrocola, F., Bravo-San Pedro, J. M., Amaravadi, R. K., Baehrecke, E. H., Cecconi, F., Codogno, P., Debnath, J., Gewirtz, D. A., Karantza, V., Kimmelman, A., Kumar, S., Levine, B., Maiuri, M. C., Martin, S. J., Penninger, J., Piacentini, M., Rubinsztein, D. C., Simon, H. U., Simonsen, A., Thorburn, A. M., Velasco, G., Ryan, K. M. & Kroemer, G. (2015). Autophagy in malignant transformation and cancer progression. *EMBO J,* **34**(7): 856-80.

Gama-Sosa, M. A., Slagel, V. A., Trewyn, R. W., Oxenhandler, R., Kuo, K. C., Gehrke, C. W. & Ehrlich, M. (1983). The 5-methylcytosine content of DNA from human tumours. *Nucleic Acids Res,* **11**(19): 6883-6894.

Garcia, E., Andrews, C., Hua, J., Kim, H. L., Sukumaran, D. K., Szyperski, T. & Odunsi, K. (2011). Diagnosis of early stage ovarian cancer by 1H NMR metabonomics of serum explored by use of a microflow NMR probe. *J Proteome Res,* **10**(4): 1765-71.

Gaul, D. A., Mezencev, R., Long, T. Q., Jones, C. M., Benigno, B. B., Gray, A., Fernandez, F. M. & Mcdonald, J. F. (2015). Highly-accurate metabolomic detection of early-stage ovarian cancer. *Sci Rep,* **5**: 16351.

Gupta, D. & Lis, C. G. (2009). Role of CA125 in predicting ovarian cancer survival - a review of the epidemiological literature. *J Ovarian Res,* **2**: 13.

Hanahan, D. & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell,* **144**(5): 646-74.

Harris, L., Fritsche, H., Mennel, R., Norton, L., Ravdin, P., Taube, S., Somerfield, M. R., Hayes, D. F., Bast, R. C., Jr. & American Society of Clinical, O. (2007). American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *J Clin Oncol,* **25**(33): 5287-312.

Haukaas, T. H., Euceda, L. R., Giskeodegard, G. F. & Bathen, T. F. (2017). Metabolic Portraits of Breast Cancer by HR MAS MR Spectroscopy of Intact Tissue Samples. *Metabolites,* **7**(2).

Hegi, M. E., Diserens, A., Gorlia, T., Hamou, M., De Tribolet, N., Weller, M., Kros, J. M., Hainfellner, J. A., Mason, W., Mariani, L., Bromberg, J. E. C., Hau, P., Mirimanoff, R. O., Cairncross, J. G., Janzer, R. C. & Stupp, R. (2005). MGMT gene silencing and benefit from temozolomide in glioblastoma. *N Engl J Med,* **352**(10): 997-1003.

Herman, J. G., Graff, J. R., Myohanen, S., Nelkin, B. D. & Baylin, S. (1996). Methylation-specific PCR: a novel PCR for methylation status of CpG islands. *Proc Natl Acad Sci U S A,* **93**: 9821-9826.

Hilvo, M., De Santiago, I., Gopalacharyulu, P., Schmitt, W. D., Budczies, J., Kuhberg, M., Dietel, M., Aittokallio, T., Markowetz, F., Denkert, C., Sehouli, J., Frezza, C., Darb-Esfahani, S. & Braicu, E. I. (2016). Accumulated Metabolites of Hydroxybutyric Acid Serve as Diagnostic and Prognostic Biomarkers of Ovarian High-Grade Serous Carcinomas. *Cancer Res,* **76**(4): 796-804.

Holmes, E., Foxall, P. J. D., Spraul, M., Farrant, R. D., Nicholson, J. K. & Lindon, J. C. (1997). 750 MHz 1H NMR spectroscopy characterisation of the complex metabolic pattern of urine from patients with inborn errors of metabolism: 2 hydroxyglutaric aciduria and maple syrup urine disease. *J Pharm Biomed Anal,* **15**: 1647-1659.

Hoskins, W. J., Mcguire, W. P., Brady, M. F., Homesley, H. D., Creasman, W. T., Berman, M., Ball, H. & Berek, J. S. (1994). The effect of diameter of largest residual disease on survival after primary cytoreductive surgery in patients with suboptimal residual epithelial ovarian carcinoma. *Am J Obstet Gynecol,* **170**(4): 974-979.

Huang, J., Hu, W. & Sood, A. K. (2010). Prognostic biomarkers in ovarian cancer. *Cancer Biomark,* **8**(4-5): 231-51.

Ingham, S. L., Warwick, J., Buchan, I., Sahin, S., O'hara, C., Moran, A., Howell, A. & Evans, D. G. (2013). Ovarian cancer among 8,005 women from a breast cancer family history clinic: no increased risk of invasive ovarian cancer in families testing negative for BRCA1 and BRCA2. *J Med Genet,* **50**(6): 368-72.

Jacobs, I. J., Menon, U., Ryan, A., Gentry-Maharaj, A., Burnell, M., Kalsi, J. K., Amso, N. N., Apostolidou, S., Benjamin, E., Cruickshank, D., Crump, D. N., Davies, S. K., Dawnay, A., Dobbs, S., Fletcher, G., Ford, J., Godfrey, K., Gunu, R., Habib, M., Hallett, R., Herod, J., Jenkins, H., Karpinskyj, C., Leeson, S., Lewis, S. J., Liston, W. R., Lopes, A., Mould, T., Murdoch, J., Oram, D., Rabideau, D. J., Reynolds, K., Scott, I., Seif, M. W., Sharma, A., Singh, N., Taylor, J., Warburton, F., Widschwendter, M., Williamson, K., Woolas, R., Fallowfield, L., Mcguire, A. J., Campbell, S., Parmar, M. & Skates, S. J. (2015). Ovarian cancer screening and mortality in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial. *The Lancet*.

Jen, J. & Wang, Y. C. (2016). Zinc finger proteins in cancer progression. *J Biomed Sci,* **23**(1): 53.

Jeyarajah, E. J., Cromwell, W. C. & Otvos, J. D. (2006). Lipoprotein particle analysis by nuclear magnetic resonance spectroscopy. *Clin Lab Med,* **26**: 847-870.

Jiang, P., Chan, C. W., Chan, K. C., Cheng, S. H., Wong, J., Wong, V. W., Wong, G. L., Chan, S. L., Mok, T. S., Chan, H. L., Lai, P. B., Chiu, R. W. & Lo, Y. M. (2015a). Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. *Proc Natl Acad Sci U S A,* **112**(11): E1317-25.

Jiang, T., Lin, Y., Yin, H., Wang, S., Sun, Q., Zhang, P. & Bi, W. (2015b). Correlation analysis of urine metabolites and clinical staging in patients with ovarian cancer. *Int J Clin Exp Med,* **8**(10): 18165-18171.

Jones, P. A. & Baylin, S. B. (2002). The fundamental role of epigenetic events in cancer. *Nat Rev Genet,* **3**(6): 415-28.

Jorissen, W., Wouters, E., Bogie, J. F., Vanmierlo, T., Noben, J. P., Sviridov, D., Hellings, N., Somers, V., Valcke, R., Vanwijmeersch, B., Stinissen, P., Mulder, M. T., Remaley, A. T. & Hendriks, J. J. (2017). Relapsing-remitting multiple sclerosis patients display an altered lipoprotein profile with dysfunctional HDL. *Sci Rep,* **7**: 43410.

Kaarteenaho, R. (2013). The current position of surgical lung biopsy in the diagnosis of idiopathic pulmonary fibrosis. *Respir Res,* **14**(43).

Kanai, Y. (2010). Genome-wide DNA methylation profiles in precancerous conditions and cancers. *Cancer Sci,* **101**(1): 36-45.

Kang, M. R., Kim, M. S., Oh, J. E., Kim, Y. R., Song, S. Y., Seo, S. I., Lee, J. Y., Yoo, N. J. & Lee, S. H. (2009). Mutational analysis of IDH1 codon 132 in glioblastomas and other common cancers. *Int J Cancer,* **125**(2): 353-5.

Ke, C., Hou, Y., Zhang, H., Fan, L., Ge, T., Guo, B., Zhang, F., Yang, K., Wang, J., Lou, G. & Li, K. (2014). Large-scale profiling of metabolic dysregulation in ovarian cancer. *Int J Cancer*.

Ke, C., Li, A., Hou, Y., Sun, M., Yang, K., Cheng, J., Wang, J., Ge, T., Zhang, F., Li, Q., Li, J., Wu, Y., Lou, G. & Li, K. (2016). Metabolic phenotyping for monitoring ovarian cancer patients. *Sci Rep,* **6**: 23334.

Kelly, Z., Moller-Levet, C., Mcgrath, S., Butler-Manuel, S., Kavitha Madhuri, T., Kierzek, A. M., Pandha, H., Morgan, R. & Michael, A. (2016). The prognostic significance of specific HOX gene expression patterns in ovarian cancer. *Int J Cancer,* **139**(7): 1608-17.

Kim, J., Coffey, D. M., Creighton, C. J., Yu, Z., Hawkins, S. M. & Matzuk, M. M. (2012). High-grade serous ovarian cancer arises from fallopian tube in a mouse model. *Proc Natl Acad Sci U S A,* **109**(10): 3921-6.

Kolwijck, E., Engelke, U. F., Van Der Graaf, M., Heerschap, A., Blom, H. J., Hadfoune, M., Buurman, W. A., Massuger, L. F. & Wevers, R. A. (2009). N-acetyl resonances in in vivo and in vitro NMR spectroscopy of cystic ovarian tumors. *NMR Biomed,* **22**(10): 1093-9.

Kruiswijk, F., Labuschagne, C. F. & Vousden, K. H. (2015). p53 in survival, death and metabolic health: a lifeguard with a licence to kill. *Nat Rev Mol Cell Biol,* **16**(7): 393-405.

Kuhn, T., Floegel, A., Sookthai, D., Johnson, T., Rolle-Kampczyk, U., Otto, W., Von Bergen, M., Boeing, H. & Kaaks, R. (2016). Higher plasma levels of lysophosphatidylcholine 18:0 are related to a lower risk of common cancers in a prospective metabolomics study. *BMC Med,* **14**: 13.

Kulis, M. & Esteller, M. (2010). DNA methylation and cancer. *Adv Genet,* **70**: 27-56.

Kyriakides, M., Rama, N., Sidhu, J., Gabra, H., Keun, H. C. & El-Bahrawy, M. (2016). Metabonomic analysis of ovarian tumour cyst fluid by proton nuclear magnetic resonance spectroscopy. *Oncotarget,* **7**(6): 7216-7226.

Leary, R. J., Sausen, M., Kinde, I., Papadopoulos, N., Carpten, J. D., Craig, D., O'shaughnessy, J., Kinzler, K. W., Parmigiani, G., Vogelstein, B., Diaz, L. A., Jr. & Velculescu, V. E. (2012). Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing. *Sci Transl Med,* **4**(162): 162ra154.

Leon, S. A., Shapiro, B., Sklaroff, D. M. & Yaros, M. J. (1977). Free DNA in the serum of cancer patients and the effect of therapy. *Cancer Res,* **37**: 646-650.

Lim, C. & Savan, R. (2014). The role of the IL-22/IL-22R1 axis in cancer. *Cytokine Growth Factor Rev,* **25**(3): 257-71.

Lo, Y. M. D., Corbetta, N., Chamberlain, P. F., Rai, V., Sargent, I. L., Redman, C. W. G. & Wainscoat, J. S. (1997). Presence of fetal DNA in maternal plasma and serum. *The Lancet,* **350**(9076): 485-487.

Lob, S., Konigsrainer, A., Rammensee, H. G., Opelz, G. & Terness, P. (2009). Inhibitors of indoleamine-2,3-dioxygenase for cancer therapy: can we see the wood for the trees? *Nat Rev Cancer,* **9**(6): 445-52.

Lodi, A., Tiziani, S., Khanim, F. L., Gunther, U. L., Viant, M. R., Morgan, G. J., Bunce, C. M. & Drayson, M. T. (2013). Proton NMR-based metabolite analyses of archived serial paired serum and urine samples from myeloma patients at different stages of disease activity identifies acetylcarnitine as a novel marker of active disease. *PLoS One,* **8**(2): e56422.

Lombardi, G., Corona, G., Bellu, L., Della Puppa, A., Pambuku, A., Fiduccia, P., Bertorelle, R., Gardiman, M. P., D'avella, D., Toffoli, G. & Zagonel, V. (2015). Diagnostic value of plasma and urinary 2-hydroxyglutarate to identify patients with isocitrate dehydrogenase-mutated glioma. *Oncologist,* **20**(5): 562-7.

Lukanova, A. & Kaaks, R. (2005). Endogenous hormones and ovarian cancer: epidemiology and current hypotheses. *Cancer Epidemiol Biomarkers Prev,* **14**(1): 98-107.

Lunt, S. Y. & Vander Heiden, M. G. (2011). Aerobic glycolysis: meeting the metabolic requirements of cell proliferation. *Annu Rev Cell Dev Biol,* **27**: 441-64.

Lynch, H. T., Casey, M. J., Snyder, C. L., Bewtra, C., Lynch, J. F., Butts, M. & Godwin, A. K. (2009). Hereditary ovarian carcinoma: Heterogeneity, molecular genetics, pathology, and management. *Molecular Oncology,* **3**(2): 97-137.

Lynch, H. T., Snyder, C. & Casey, M. J. (2013). Hereditary ovarian and breast cancer: what have we learned? *Ann Oncol,* **24 Suppl 8**: viii83-viii95.

Malpica, A., Deavers, M. T., Lu, K., Bodurka, D. C., Atkinson, E. N., Gershenson, D. M. & Silva, E. G. (2004). Grading ovarian serous carcinoma using a two-tier system. *Am J Surg Pathol,* **28**: 496-504.

Mandel, P. & Metais, P. (1948). Les acides nucleiques du plasma sanguin chez l'homme. *C R Seances Soc Biol Fil,* **142**(3-4): 241-243.

Mardis, E. R., Ding, L., Dooling, D. J., Larson, D. E., Mclellan, M. D., Chen, K., Koboldt, D. C., Fulton, R. S., Delahaunty, B. A., Mcgrath, S. D., Fulton, L. A., Locke, D. P., V.J., M., Abbott, R. M., Vickery, T. L., Reed, J. S., Robinson, J. S., Wylie, T., Smith, S. M., Carmichael, L., Eldred, J. M., Harris, C. C., Walker, J., Peck, J. B., Du, F., Dukes, A. F., Sanderson, G. E., Brummett, A. M., Clark, E., Mcmichael, J. F., Meyer, R. J., Schindler, J. K., Pohl, C. S., Wallis, J. W., Shi, X., Lin, L., Schmidt, H. K., Tang, Y., Haipek, C., Wiechert, M. E., Ivy, J. V., Kalicki, J., Elliott, G., Ries, R. E., Payton, J. E., Westervelt, P., Tomasson, M. H., Watson, M. A., Baty, J., Heath, S., Shannon, W. D., Nagarajan, R., Link, D. C., Walter, M. J., Graubert, T. A., Dipersio, J. F., Wilson, R. K. & Ley, T. J. (2009). Recurring mutations found by sequencing an acute myeloid leukaemia genome. *N Engl J Med,* **361**: 1058-1066.

Margolin, G., Petrykowska, H. M., Jameel, N., Bell, D. W., Young, A. C. & Elnitski, L. (2016). Robust Detection of DNA Hypermethylation of ZNF154 as a Pan-Cancer Locus with in Silico Modeling for Blood-Based Diagnostic Development. *J Mol Diagn,* **18**(2): 283-98.

Mcbride, D. J., Orpana, A. K., Sotiriou, C., Joensuu, H., Stephens, P. J., Mudie, L. J., Hamalainen, E., Stebbings, L. A., Andersson, L. C., Flanagan, A. M., Durbecq, V., Ignatiadis, M., Kallioniemi, O., Heckman, C. A., Alitalo, K., Edgren, H., Futreal, P. A., Stratton, M. R. & Campbell, P. J. (2010). Use of cancer-specific genomic rearrangements to quantify disease burden in plasma from patients with solid tumors. *Genes Chromosomes Cancer,* **49**(11): 1062-9.

Mcshane, L. M., Altman, D. G., Sauerbrei, W., Taube, S. E., Gion, M., Clark, G. M. & Statistics Subcommittee of The, N. C. I. E. W. G. O. C. D. (2005). REporting recommendations for tumour MARKer prognostic studies (REMARK). *Br J Cancer,* **93**(4): 387-91.

Meiboom, S. & Gill, D. (1958). Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. *Review of Scientific Instruments,* **29**(8): 688-691.

Melamed, A., Manning-Geist, B., Bregar, A. J., Diver, E. J., Goodman, A., Del Carmen, M. G., Schorge, J. O. & Rauh-Hain, J. A. (2017). Associations between residual disease and survival in epithelial ovarian cancer by histologic type. *Gynecol Oncol,* **147**(2): 250-256.

Menendez, J. A. & Lupu, R. (2007). Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis. *Nat Rev Cancer,* **7**(10): 763-77.

Menon, U., Mcguire, A. J., Raikou, M., Ryan, A., Davies, S. K., Burnell, M., Gentry-Maharaj, A., Kalsi, J. K., Singh, N., Amso, N. N., Cruickshank, D., Dobbs, S., Godfrey, K., Herod, J., Leeson, S., Mould, T., Murdoch, J., Oram, D., Scott, I., Seif, M. W., Williamson, K., Woolas, R., Fallowfield, L., Campbell, S., Skates, S. J., Parmar, M. & Jacobs, I. J. (2017). The cost-effectiveness of screening for ovarian cancer: results from the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS). *Br J Cancer,* **117**(5): 619-627.

Menon, U., Ryan, A., Kalsi, J., Gentry-Maharaj, A., Dawnay, A., Habib, M., Apostolidou, S., Singh, N., Benjamin, E., Burnell, M., Davies, S., Sharma, A., Gunu, R., Godfrey, K., Lopes, A., Oram, D., Herod, J., Williamson, K., Seif, M. W., Jenkins, H., Mould, T., Woolas, R., Murdoch, J. B., Dobbs, S., Amso, N. N., Leeson, S., Cruickshank, D., Scott, I., Fallowfield, L., Widschwendter, M., Reynolds, K., Mcguire, A., Campbell, S., Parmar, M., Skates, S. J. & Jacobs, I. (2015). Risk Algorithm Using Serial Biomarker Measurements Doubles the Number of Screen-Detected Cancers Compared With a Single-Threshold Rule in the United Kingdom Collaborative Trial of Ovarian Cancer Screening. *J Clin Oncol,* **33**(18): 2062-71.

Mikeska, T., Bock, C., Do, H. & Dobrovic, A. (2012). DNA methylation biomarkers in cancer: progress towards clinical implementation. *Expert Rev Mol Diagn,* **12**(5): 473-87.

Montavon, C., Gloss, B. S., Warton, K., Barton, C. A., Statham, A. L., Scurry, J. P., Tabor, B., Nguyen, T. V., Qu, W., Samimi, G., Hacker, N. F., Sutherland, R. L., Clark, S. J. & O'brien, P. M. (2012). Prognostic and diagnostic significance of DNA methylation patterns in high grade serous ovarian cancer. *Gynecol Oncol,* **124**(3): 582-8.

Moore, R. G., Miller, M. C., Disilvestro, P., Landrum, L. M., Gajewski, W., Ball, J. J. & Skates, S. J. (2011). Evaluation of the diagnostic accuracy of the risk of ovarian malignancy algorithm in women with a pelvic mass. *Obstet Gynecol,* **118**(2 Pt 1): 280-8.

Mukherji, S. K., Schiro, S., Castiloo, M., Kwock, L., Muller, K. E. & Blackstock, W. (1997). Proton MR Spectroscopy of Squamous Cell Carcinoma of the Extracranial head and neck: in vitro and in vivo studies. *Am J Neuroradiol,* **18**: 1057-1072.

Murtaza, M., Dawson, S. J., Tsui, D. W., Gale, D., Forshew, T., Piskorz, A. M., Parkinson, C., Chin, S. F., Kingsbury, Z., Wong, A. S., Marass, F., Humphray, S., Hadfield, J., Bentley, D., Chin, T. M., Brenton, J. D., Caldas, C. & Rosenfeld, N. (2013). Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature,* **497**(7447): 108-12.

Newman, A. M., Bratman, S. V., To, J., Wynne, J. F., Eclov, N. C., Modlin, L. A., Liu, C. L., Neal, J. W., Wakelee, H. A., Merritt, R. E., Shrager, J. B., Loo, B. W., Jr., Alizadeh, A. A. & Diehn, M. (2014). An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat Med,* **20**(5): 548-54.

Nicholls, A. W., Mortishire-Smith, R. J. & Nicholson, J. K. (2003). NMR spectroscopic-based metabonomic studies of urinary metabolic variation in acclimatizing germ-free rats. *Chem Res Toxicol,* **16**(11): 1395-1404.

Nicholson, J. K. & Foxall, P. J. D. (1995). 750 MHz 1H and 1H-13C NMR spectroscopy of human blood plasma. *67,* (793-811).

Nicholson, J. K. & Lindon, J. C. (2008). Systems Biology: Metabonomics. *Nature,* **456**(443): 1054-1056.

Nieman, K. M., Kenny, H. A., Penicka, C. V., Ladanyi, A., Buell-Gutbrod, R., Zillhardt, M. R., Romero, I. L., Carey, M. S., Mills, G. B., Hotamisligil, G. S., Yamada, S. D., Peter, M. E., Gwin, K. & Lengyel, E. (2011). Adipocytes promote ovarian cancer metastasis and provide energy for rapid tumor growth. *Nat Med,* **17**(11): 1498-503.

Norton, M. E., Jacobsson, B., Swamy, G. K., Laurent, L. C., Ranzini, A. C., Brar, H., Tomlinson, M. W., Pereira, L., Spitz, J. L., Hollemon, D., Cuckle, H., Musci, T. J. & Wapner, R. J. (2015). Cell-free DNA analysis for noninvasive examination of trisomy. *N Engl J Med,* **372**(17): 1589-97.

Nygren, A. O., Dean, J., Jensen, T. J., Kruse, S., Kwong, W., Van Den Boom, D. & Ehrich, M. (2010). Quantification of fetal DNA by use of methylation-based DNA discrimination. *Clin Chem,* **56**(10): 1627-35.

Odunsi, K., Wollman, R. M., Ambrosone, C. B., Hutson, A., Mccann, S. E., Tammela, J., Geisler, J. P., Miller, G., Sellers, T., Cliby, W., Qian, F., Keitz, B., Intengan, M., Lele, S. & Alderfer, J. L. (2005). Detection of epithelial ovarian cancer using 1H-NMR-based metabonomics. *Int J Cancer,* **113**(5): 782-8.

Okamoto, T., Yamaguchi, K., Huang, Z., Whitaker, R., Konishi, I., Berchuck, A. & Murphy, S. K. Methylated-mediated repression of ZNF154 in ovarian cancer is associated with poor overall survival. AACR 103rd Annual meeting 2012, 2012 Chicago, IL.

Osl, M., Dreiseitl, S., Pfeifer, B., Weinberger, K., Klocker, H., Bartsch, G., Schafer, G., Tilg, B., Graber, A. & Baumgartner, C. (2008). A new rule-based algorithm for identifying metabolic markers in prostate cancer using tandem mass spectrometry. *Bioinformatics,* **24**(24): 2908-14.

Owen, O. E., Kalhan, S. C. & Hanson, R. W. (2002). The key role of anaplerosis and cataplerosis for citric acid cycle function. *J Biol Chem,* **277**(34): 30409-12.

Page, K., Powles, T., Slade, M. J., Mt, D. E. B., Walker, R. A., Coombes, R. C. & Shaw, J. A. (2006). The importance of careful blood processing in isolation of cell-free DNA. *Ann N Y Acad Sci,* **1075**: 313-7.

Paley, P. J., Staskus, K. A., Gebhard, K., Mohanraj, D., Twiggs, L. B., Carson, L. F. & Ramakrishnan, S. (1997). Vascular endothelial growth factor expression in early stage ovarian carcinoma. *Cancer,* **80**(1): 98-106.

Parkin, D. M., Boyd, L. & Walker, L. C. (2011). 16. The fraction of cancer attributable to lifestyle and environmental factors in the UK in 2010. *Br J Cancer,* **105 Suppl 2**: S77-81.

Pavlova, N. N. & Thompson, C. B. (2016). The Emerging Hallmarks of Cancer Metabolism. *Cell Metab,* **23**(1): 27-47.

Pepin, K., Del Carmen, M. G., Brown, A. & Dizon, D. S. (2014). CA125 and epithelial ovarian cancer: role in screening, diagnosis and surveillance. *Am J Hematol Oncol,* **10**(6): 22-29.

Phallen, J., Sausen, M., Adleff, V., Leal, A., Hruban, C., White, J., Anagnostou, V., Fiksel, J., Cristiano, S., Papp, E., Speir, S., Reinert, T., Orntoft, M. W., Woodward, B. D., Murphy, D., Parpart-Li, S., Riley, D., Nesselbush, M., Sengamalay, N., Georgiadis, A., Ki, Q., Madsen, M. R., Mortensen, F. V., Huiskens, J., Punt, C., Van Grieken, N., Fijneman, R., Meijer, G., Husain, H., Scharpf, R. B., Diaz, L. A., Jr., Jones, S., Angiuoli, S., Orntoft, T., Nielsen, H. J., Anderson, C. L. & Velculescu, V. E. (2017). Direct detection of early-stage cancers using circulating tumour DNA. *Sci Transl Med,* **9**(403).

Phelps, D. L., Borley, J. V., Flower, K. J., Dina, R., Darb-Esfahani, S., Braicu, I., Sehouli, J., Fotopoulou, C., Wilhelm-Benartzi, C. S., Gabra, H., Yazbek, J., Chatterjee, J., Ip, J., Khan, H., Likos-Corbett, M. T., Brown, R. & Ghaem-Maghami, S. (2017). Methylation of MYLK3 gene promoter region: a biomarker to stratify surgical care in ovarian cancer in a multicentre study. *Br J Cancer,* **116**(10): 1287-1293.

Platten, M., Wick, W. & Van Den Eynde, B. J. (2012). Tryptophan catabolism in cancer: beyond IDO and tryptophan depletion. *Cancer Res,* **72**(21): 5435-40.

Portela, A. & Esteller, M. (2010). Epigenetic modifications and human disease. *Nat Biotechnol,* **28**(10): 1057-68.

Prat, J. (2015). FIGO's staging classification for cancer of the ovary, fallopian tube, and peritoneum: abridged republication. *J Gynecol Oncol,* **26**(2): 87-9.

Qiu, Y., Zhou, B., Su, M., Baxter, S., Zheng, X., Zhao, X., Yen, Y. & Jia, W. (2013). Mass spectrometry-based quantitative metabolomics revealed a distinct lipid profile in breast cancer patients. *Int J Mol Sci,* **14**(4): 8047-61.

Rankin, N. J., Preiss, D., Welsh, P., Burgess, K. E., Nelson, S. M., Lawlor, D. A. & Sattar, N. (2014). The emergence of proton nuclear magnetic resonance metabolomics in the cardiovascular arena as viewed from a clinical perspective. *Atherosclerosis,* **237**(1): 287-300.

Reinert, T., Borre, M., Christiansen, A., Hermann, G. G., Orntoft, T. F. & Dyrskjot, L. (2012). Diagnosis of bladder cancer recurrence based on urinary levels of EOMES, HOXA9, POU4F2, TWIST1, VIM, and ZNF154 hypermethylation. *PLoS One,* **7**(10): e46297.

Riedinger, J. M., Wafflart, J., Ricolleau, G., Eche, N., Larbre, H., Basuyau, J. P., Dalifard, I., Hacene, K. & Pichon, M. F. (2006). CA 125 half-life and CA 125 nadir during induction chemotherapy are independent predictors of epithelial ovarian cancer outcome: results of a French multicentric study. *Ann Oncol,* **17**(8): 1234-8.

Rosen, D. G., Zhang, Z., Shan, W. & Liu, J. (2010). Morphological and molecular basis of ovarian serous carcinoma. *J Biomed Res.,* **24**(4): 257-263.

Sanchez-Vega, F., Gotea, V., Petrykowska, H. M., Margolin, G., Krivak, T. C., Deloia, J. A., Bell, D. W. & Elnitski, L. (2013). Recurrent patterns of DNA methylation in the ZNF154, CASP8, and VHL promoters across a wide spectrum of human solid epithelial tumors and cancer cell lines. *Epigenetics,* **8**(12): 1355-72.

Sarojini, S., Tamir, A., Lim, H., Li, S., Zhang, S., Goy, A., Pecora, A. & Suh, K. S. (2012). Early detection biomarkers for ovarian cancer. *J Oncol,* **2012**: 709049.

Schmidt, B., Liebenberg, V., Dietrich, D., Schlegel, T., Kneip, C., Seegebarth, A., Flemming, N., Seemann, S., Distler, J., Lewin, J., Tetzner, R., Weickmann, S., Wille, U., Liloglou, T., Raji, O., Walshaw, M., Fleischhacker, M., Witt, C. & Field, J. K. (2010). SHOX2 DNA methylation is a biomarker for the diagnosis of lung cancer based on bronchial aspirates. *BMC Cancer,* **10**(600).

Sharma, S., Kelly, T. K. & Jones, P. A. (2010). Epigenetics in cancer. *Carcinogenesis,* **31**(1): 27-36.

Shih, K. K. & Chi, D. S. (2010). Maximal cytoreductive effort in epithelial ovarian cancer surgery. *J Gynecol Oncol,* **21**(2): 75-80.

Shirahata, A., Sakata, M., Sakuraba, K., Goto, T., Mizukami, H., Saito, M., Ishibashi, K., Kigawa, G., Nemoto, H., Sanada, Y. & Hibi, K. (2009). Vimentin methylation as a marker for advanced colorectal carcinoma. *Anticancer Res,* **29**: 279-282.

Siliprandi, N., Sartorelli, L., Ciman, M. & Di Lisa, F. (1989). Carnitine: metabolism and clinical chemistry. *Clin Chem Acta,* **183**(1): 3-11.

Simmons, A. R., Clarke, C. H., Badgwell, D. B., Lu, Z., Sokoll, L. J., Lu, K. H., Zhang, Z., Bast, R. C., Jr. & Skates, S. J. (2016). Validation of a Biomarker Panel and Longitudinal Biomarker Performance for Early Detection of Ovarian Cancer. *Int J Gynecol Cancer,* **26**(6): 1070-7.

Singer, G., Stohr, R., Cope, L., Dehari, R., Hartmann, A., Cao, D., Wang, T., Kurman, R. & Shih, L. (2005). Patterns of p53 mutations separate ovarian serous borderline tumors and low- and high-grade carcinomas and provide support for a new model of ovarian carcinogeneis: a mutational analysis with immunohistochemical correlation. *Am J Surg Pathol,* **29**(2): 218-224.

Siskos, A. P., Jain, P., Romisch-Margl, W., Bennett, M., Achaintre, D., Asad, Y., Marney, L., Richardson, L., Koulman, A., Griffin, J. L., Raynaud, F., Scalbert, A., Adamski, J., Prehn, C. & Keun, H. C. (2017). Interlaboratory Reproducibility of a Targeted Metabolomics Platform for Analysis of Human Serum and Plasma. *Anal Chem,* **89**(1): 656-665.

Skates, S. J., Pauler, D. K. & Jacobs, I. J. (2001). Screening based on the risk of cancer calculation from bayesian hierarchical changepoint and mixture models of longitudinal markers. *J Am Stat Assoc,* **96**(454): 429-439.

Slupsky, C. M., Steed, H., Wells, T. H., Dabbs, K., Schepansky, A., Capstick, V., Faught, W. & Sawyer, M. B. (2010). Urine metabolite analysis offers potential early diagnosis of ovarian and breast cancers. *Clin Cancer Res,* **16**(23): 5835-41.

Sorenson, G. D., Pribish, D. M., Valone, F. H., Memoli, V. A., Bzik, D. J. & Yao, S. (1994). Soluble normal and mutated DNA sequences from single-copy genes in human blood. *Cancer Epidemiol Biomarkers Prev,* **3**: 67-71.

Stine, Z. E., Walton, Z. E., Altman, B. J., Hsieh, A. L. & Dang, C. V. (2015). MYC, Metabolism, and Cancer. *Cancer Discov,* **5**(10): 1024-39.

Suzuki, M. M. & Bird, A. (2008). DNA methylation landscape: provocative insights from epigenomics. *Nat Rev Genet,* **9**(6): 465-476.

Takao, M., Okamoto, A., Nikaido, T., Urashima, M., Takakura, S., Saito, M., Saito, M., Okamoto, S., Takikawa, O., Sasaki, H., Yasuda, M., Ochiai, K. & Tanaka, T. (2007). Increased synthesis of indoleamine-2,3-dioxygenase protein is positively associated with impaired survival in patients with serous-type, but not other types of, ovarian cancer. *Oncol Rep,* **17**: 1333-1339.

Taylor, H. S., Vanden Heuvel, G. B. & Igarashi, P. (1997). A conserved Hox axis in the mouse and human female reproductive system: late establishment and persistent adult expression of the hoxa cluster genes. *Biol Reprod,* **57**: 1338-1345.

Thaker, P. H., Deavers, M., Celestino, J., Thornton, A., Fletcher, M. S., Landen, C. N., Kinch, M. S., Kiener, P. A. & Sood, A. K. (2004). EphA2 expression is associated with aggressive features in ovarian carcinoma. *Clin Cancer Res,* **10**: 5145-5150.

Tiziani, S., Lopes, V. & Günther, U. L. (2009). Early Stage Diagnosis of Oral Cancer Using 1H NMR–Based Metabolomics. *Neoplasia,* **11**(3): 269-IN10.

Tretter, L., Patocs, A. & Chinopoulos, C. (2016). Succinate, an intermediate in metabolism, signal transduction, ROS, hypoxia, and tumorigenesis. *Biochim Biophys Acta,* **1857**(8): 1086-1101.

Trygg, J. & Wold, S. (2002). Orthogonal projections to latent structures (O-PLS). *J. Chemometrics,* **16**: 119-128.

Tsilidis, K. K., Allen, N. E., Key, T. J., Dossus, L., Lukanova, A., Bakken, K., Lund, E., Fournier, A., Overvad, K., Hansen, L., Tjonneland, A., Fedirko, V., Rinaldi, S., Romieu, I., Clavel-Chapelon, F., Engel, P., Kaaks, R., Schutze, M., Steffen, A., Bamia, C., Trichopoulou, A., Zylis, D., Masala, G., Pala, V., Galasso, R., Tumino, R., Sacerdote, C., Bueno-De-Mesquita, H. B., Van Duijnhoven, F. J., Braem, M. G., Onland-Moret, N. C., Gram, I. T., Rodriguez, L., Travier, N., Sanchez, M. J., Huerta, J. M., Ardanaz, E., Larranaga, N., Jirstrom, K., Manjer, J., Idahl, A., Ohlson, N., Khaw, K. T., Wareham, N., Mouw, T., Norat, T. & Riboli, E. (2011). Oral contraceptive use and reproductive factors and risk of ovarian cancer in the European Prospective Investigation into Cancer and Nutrition. *Br J Cancer,* **105**(9): 1436-42.

Turkoglu, O., Zeb, A., Graham, S., Szyperski, T., Szender, J. B., Odunsi, K. & Bahado-Singh, R. (2016). Metabolomics of biomarker discovery in ovarian cancer: a systematic review of the current literature. *Metabolomics,* **12**(4): 1-22.

Ueland, F. R. (2017). A Perspective on Ovarian Cancer Biomarkers: Past, Present and Yet-To-Come. *Diagnostics (Basel),* **7**(1): 14.

Van Den Berg, R. A., Hoefsloot, H. C., Westerhuis, J. A., Smilde, A. K. & Van Der Werf, M. J. (2006). Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics,* **7**: 142.

Vasioukhin, V., Anker, P., Maurice, P., Lyautey, J., Lederrey, C. & Stroun, M. (1993). Point mutations of the N-ras gene in the blood plasma DNA of patients with myelodysplastic syndrome or acute myelogenous leukaemia. *Br J Haematol,* **86**: 774-779.

Voelxen, N. F., Walenta, S., Proescholdt, M., Dettmer, K., Pusch, S. & Mueller-Klieser, W. (2016). Quantitative Imaging of D-2-Hydroxyglutarate in Selected Histological Tissue Areas by a Novel Bioluminescence Technique. *Front Oncol,* **6**: 46.

Waddington, C. H. (1942). The epigenotype. *Endeavour,* **1**: 18-20.

Wang, X., Wang, J., Wang, Z., Wang, Q. & Li, H. (2015). Dynamic monitoring of plasma amino acids and carnitine during chemotherapy of patients with alimentary canal malignancies and its clinical value. *Onco Targets Ther,* **8**: 1989-96.

Wang, Y., Holmes, E., Nicholson, J. K., Cloarec, O., Chollet, J., Tanner, M., Singer, B. H. & Utzinger, J. (2004). Metabonomic investigations in mice infected with Schistosoma mansoni: an approach for biomarker identification. *Proc Natl Acad Sci U S A,* **101**(34): 12676-81.

Warburg, O. (1956). On respiratory impairment in cancer cells. *Science,* **124**(3215): 269-270.

Ware Miller, R., Smith, A., Desimone, C. P., Seamon, L., Goodrich, S., Podzielinski, I., Sokoll, L., Van Nagell, J. R., Jr., Zhang, Z. & Ueland, F. R. (2011). Performance of the American College of Obstetricians and Gynecologists' ovarian tumor referral guidelines with a multivariate index assay. *Obstet Gynecol,* **117**(6): 1298-306.

Werner, H. M., Mills, G. B. & Ram, P. T. (2014). Cancer Systems Biology: a peek into the future of patient care? *Nat Rev Clin Oncol,* **11**(3): 167-76.

Widschwendter, M., Apostolidou, S., Jones, A. A., Fourkala, E. O., Arora, R., Pearce, C. L., Frasco, M. A., Ayhan, A., Zikan, M., Cibula, D., Iyibozkurt, C. A., Yavuz, E., Hauser-Kronberger, C., Dubeau, L., Menon, U. & Jacobs, I. J. (2009). HOXA methylation in normal endometrium from premenopausal women is associated with the presence of ovarian cancer: a proof of principle study. *Int J Cancer,* **125**(9): 2214-8.

Wimberger, P., Wehling, M., Lehmann, N., Kimmig, R., Schmalfeldt, B., Burges, A., Harter, P., Pfisterer, J. & Du Bois, A. (2010). Influence of residual tumor on outcome in ovarian cancer patients with FIGO stage IV disease: an exploratory analysis of the AGO-OVAR (Arbeitsgemeinschaft Gynaekologische Onkologie Ovarian Cancer Study Group). *Ann Surg Oncol,* **17**(6): 1642-8.

Wishart, D. S. (2016). Emerging applications of metabolomics in drug discovery and precision medicine. *Nat Rev Drug Discov,* **15**(7): 473-84.

Wold, S., Esbensen, K. & Geladi, P. (1987). Principal component analysis. *Chemom Intell Lab Syst,* **2**: 37-52.

Wold, S., Kettaneh, N. & Tjessem, K. (1996). Hierarchical multiblock pls and pc models for easier model interpretation and as an alternative to variable selection. *J. Chemometrics,* **10**: 463-482.

Wong, I. H. N., Lo, Y. M. D., Zhang, J., Liew, C., Ng, M. H. L., Wong, N., Lai, P. B. S., Lau, W. Y., Hjelm, N. M. & Johnson, P. J. (1999). Detection of aberrant p16 methylation in the plasma and serum of liver cancer patients. *Cancer Res,* **59**: 71-73.

Xie, H., Hou, Y., Cheng, J., Openkova, M. S., Xia, B., Wang, W., Li, A., Yang, K., Li, J., Xu, H., Yang, C., Ma, L., Li, Z., Fan, X., Li, K. & Lou, G. (2017). Metabolic profiling and novel plasma biomarkers for predicting survival in epithelial ovarian cancer. *Oncotarget,* **8**(19): 32134-32146.

Yan, H., Parsons, D. W., Jin, G., Mclendon, R., Rasheed, A., Yuan, W., Kos, I., Batinic-Haberle, I., Jones, S., Riggins, G. J., Friedman, H., Friedman, A., Reardon, D., Herndon, J., Kinzler, K. W., Velculescu, V. E., Vogelstein, B. & Bigner, D. D. (2009). IDH1 and IDH2 mutations in gliomas. *N Engl J Med,* **360**(765-773).

Yang, M., Soga, T. & Pollard, P. J. (2013). Oncometabolites: linking altered metabolism with cancer. *J Clin Invest,* **123**(9): 3652-8.

Yin, P., Lehmann, R. & Xu, G. (2015). Effects of pre-analytical processes on blood samples used in metabolomics studies. *Anal Bioanal Chem,* **407**(17): 4879-92.

Yuan, T. L. & Cantley, L. C. (2008). PI3K pathway alterations in cancer: variations on a theme. *Oncogene,* **27**(41): 5497-510.

Zhang, H., Ge, T., Cui, X., Hou, Y., Ke, C., Yang, M., Yang, K., Wang, J., Guo, B., Zhang, F., Lou, G. & Li, K. (2015). Prediction of advanced ovarian cancer recurrence by plasma metabolic profiling. *Mol Biosyst,* **11**(2): 516-21.

Zhang, T., Wu, X., Ke, C., Yin, M., Li, Z., Fan, L., Zhang, W., Zhang, H., Zhao, F., Zhou, X., Lou, G. & Li, K. (2013). Identification of potential biomarkers for ovarian cancer by urinary metabolomic profiling. *J Proteome Res,* **12**(1): 505-12.

Zhang, T., Wu, X., Yin, M., Fan, L., Zhang, H., Zhao, F., Zhang, W., Ke, C., Zhang, G., Hou, Y., Zhou, X., Lou, G. & Li, K. (2012). Discrimination between malignant and benign ovarian tumors by plasma

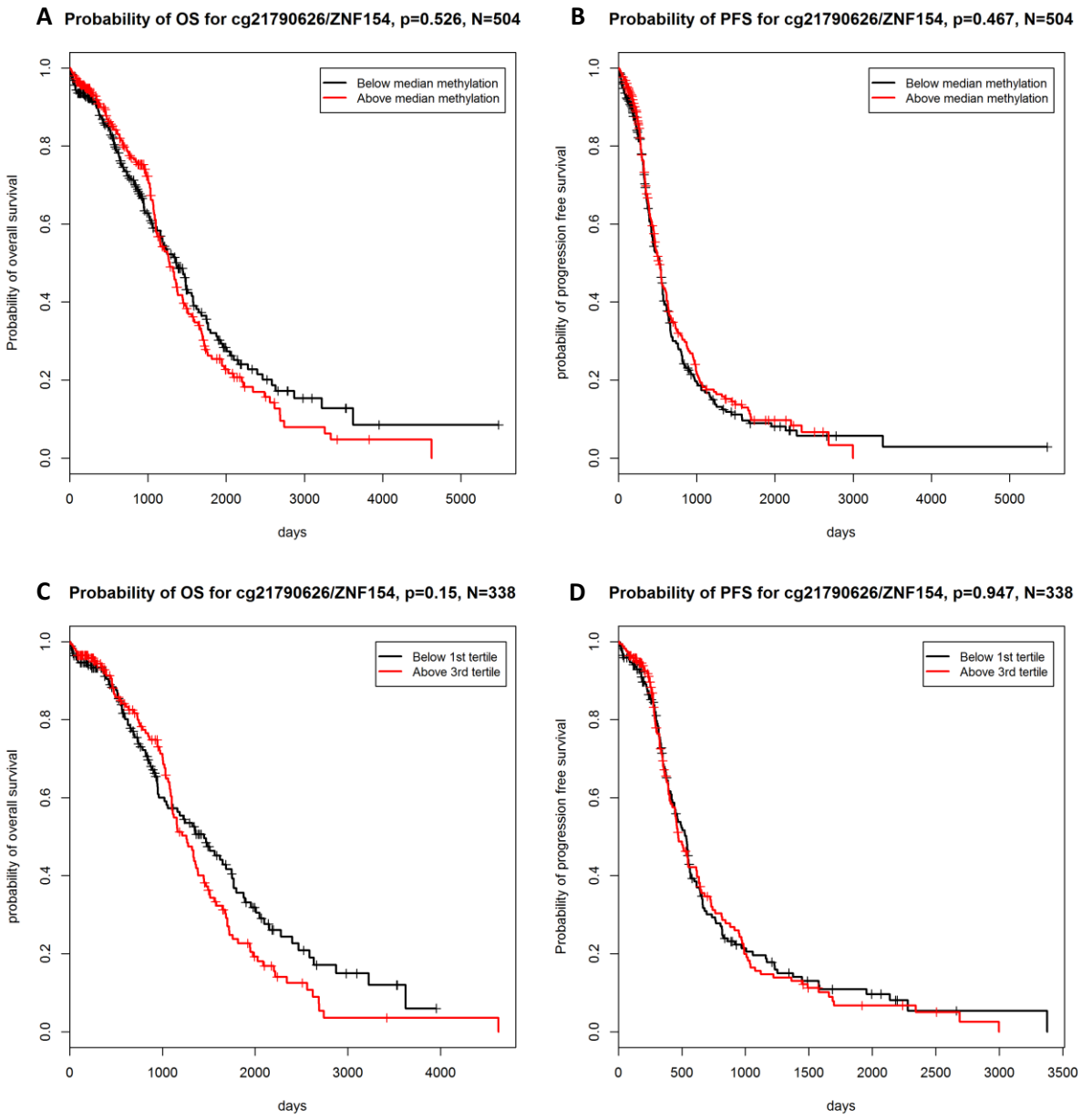
metabolomic profiling using ultra performance liquid chromatography/mass spectrometry. *Clin Chim Acta,* **413**(9-10): 861-8.

Zhang, Y., Liu, Y., Li, L., Wei, J., Xiong, S. & Zhao, Z. (2016). High resolution mass spectrometry coupled with multivariate data analysis revealing plasma lipidomic alteration in ovarian cancer in Asian women. *Talanta,* **150**: 88-96.

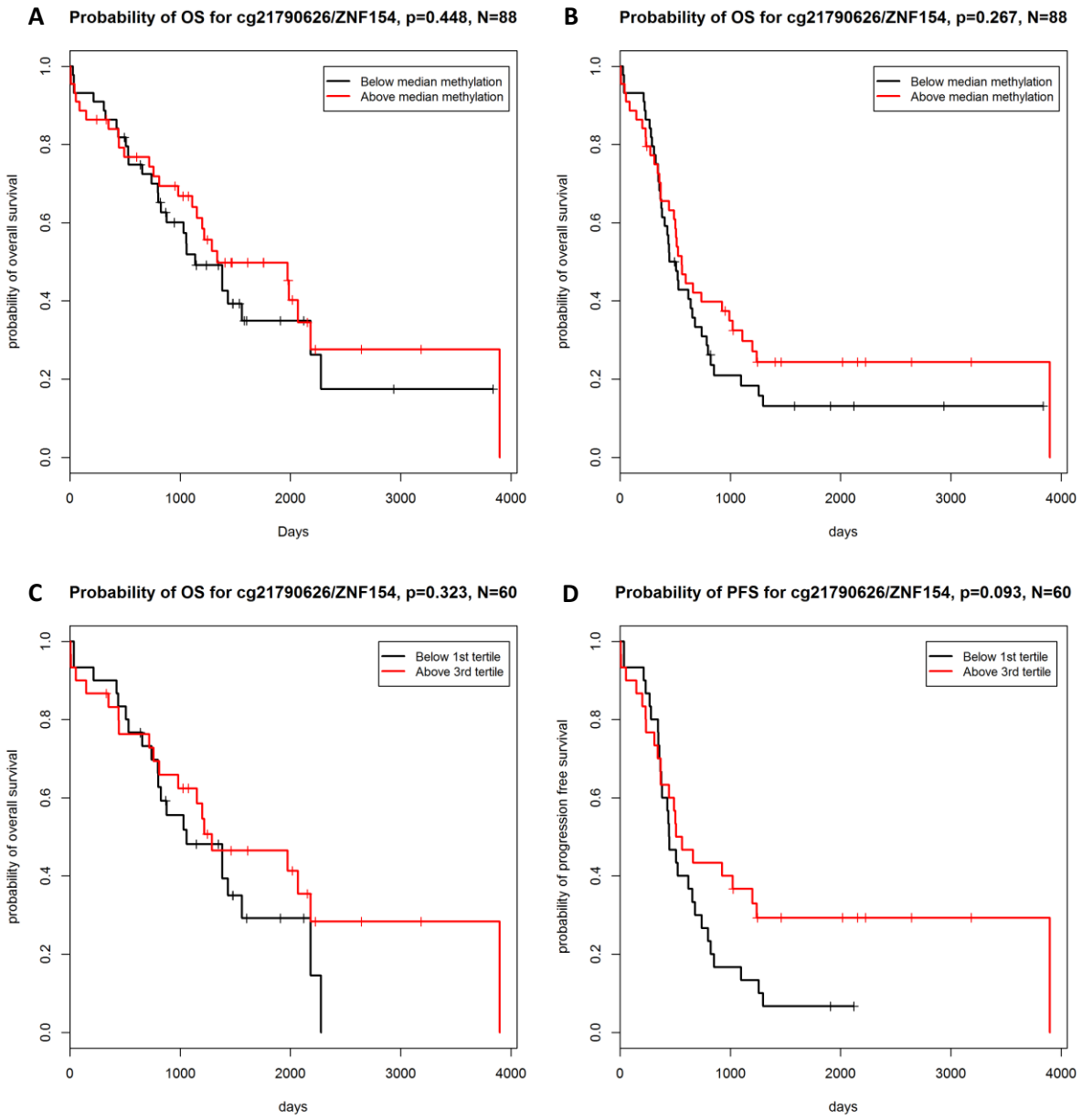
Zhao, Z., Xiao, Y., Elson, P., Tan, H., Plummer, S. J., Berk, M., Aung, P. P., Lavery, I. C., Achkar, J. P., Li, L., Casey, G. & Xu, Y. (2007). Plasma lysophosphatidylcholine levels: potential biomarkers for colorectal cancer. *J Clin Oncol,* **25**(19): 2696-701.

# Chapter 8. Appendix 1

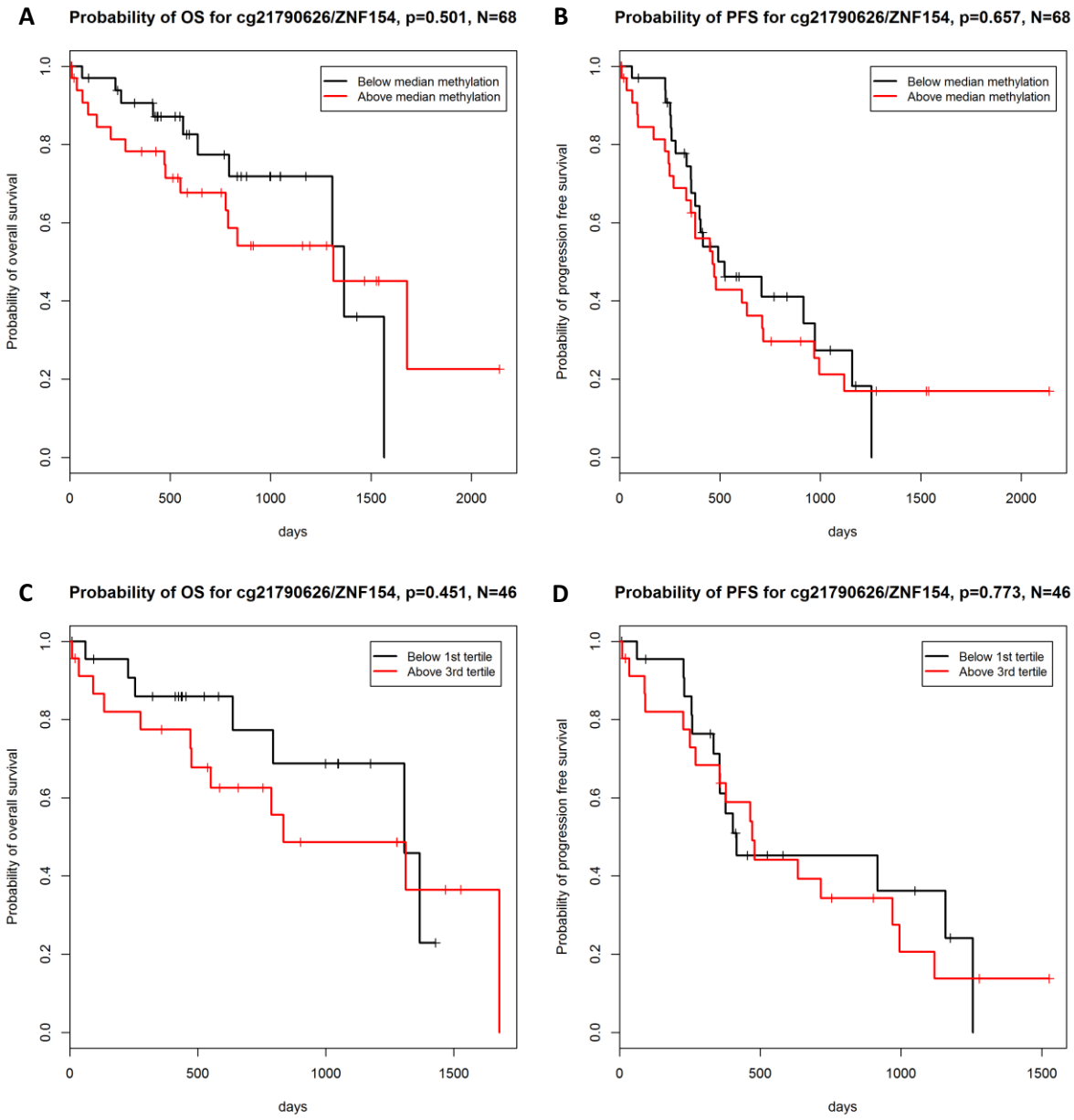
Analysis of the association between methylation of cg21790626 in the *ZNF154* gene and overall and progression free survival in the TCGA, Charité and New HH cohorts.




Supplementary Figure 1.1. Kaplan-Meier curves for OS and PFS in the TCGA dataset. **A**, OS and **B**, PFS dependent on methylation above and below the median. **C**, OS and **D**, PFS dependent on methylation above and below the 1st and 3rd tertiles.




Supplementary Figure 1.2. Kaplan-Meier curves for OS and PFS in the Charité dataset. **A**, OS and **B**, PFS dependent on methylation above and below the median. **C**, OS and **D**, PFS dependent on methylation above and below the 1$^{st}$ and 3$^{rd}$ tertiles.




Supplementary Figure 1.3. Kaplan-Meier curves for OS and PFS in the New HH dataset. **A**, OS and **B**, PFS dependent on methylation above and below the median. **C**, OS and **D**, PFS dependent on methylation above and below the 1$^{st}$ and 3$^{rd}$ tertiles.

# Appendix 2

Supplementary Table 2.1. List of metabolites from the Biocrates Absolute*IDQ*® p150 kit.

| Metabolite Class/Abbreviation | Name |
|---|---|
| **Acylcarnitines** | |
| C0 | Carnitine |
| C2 | Acetylcarnitine |
| C3 | Propionylcarnitine |
| C3-OH | Hydroxypropionylcarnitine |
| C3:1 | Propenoylcarnitine |
| C3-DC (C4-OH) | Hydroxybutyrylcarnitine |
| C4 | Butyrylcarnitine |
| C4:1 | Butenylcarnitine |
| C5 | Valerylcarnitine |
| C5-DC (C6-OH) | Glutarylcarnitine (Hydroxyhexanoylcarnitine) |
| C5-M-DC | Methylglutarylcarnitine |
| C5-OH (C3-DC-M) | Hydroxyvalerylcarnitine (Methylmalonylcarnitine) |
| C5:1 | Tiglylcarnitine |
| C5:1-DC | Glutaconylcarnitine |
| C6 (C4:1-DC) | Hexanoylcarnitine (Fumarylcarnitine) |
| C6:1 | Hexenoylcarnitine |
| C7-DC | Pimelylcarnitine |
| C8 | Octanoylcarnitine |
| C8:1 | Octenoylcarnitine |
| C9 | Nonaylcarnitine |
| C10 | Decanoylcarnitine |
| C10:1 | Decenoylcarnitine |
| C10:2 | Decadienylcarnitine |
| C12 | Dodecanoylcarnitine |
| C12-DC | Dodecanedioylcarnitine |
| C12:1 | Dodecenoylcarnitine |
| C14 | Tetradecanoylcarnitine |
| C14:1 | Tetradecenoylcarnitine |
| C14:1-OH | Hydroxytetradeceonylcarnitine |
| C14:2 | Tetradecadienylcarnitine |
| C14:2-OH | Hydroxytetradecadienylcarnitine |
| C16 | Hexadecanoylcarnitine |
| C16-OH | Hydroxyhexadecanoylcarnitine |
| C16:1 | Hexadecenoylcarnitine |
| C16:1-OH | Hydroxyhexadecenoylcarnitine |
| C16:2 | Hexadecadienylcarnitine |
| C16:2-OH | Hydroxyhexadecadienylcarnitine |

| | |
|---|---|
| C18 | Octadecanoylcarnitine |
| C18:1 | Octadecenoylcarnitine |
| C18:1-OH | Hydroxyoctadecenoylcarnitine |
| C18:2 | Octadecadienylcarnitine |
| **Amino acids** | |
| Arg | Arginine |
| Gln | Glutamine |
| Gly | Glycine |
| His | Histidine |
| Met | Methionine |
| Orn | Ornithine |
| Phe | Phenylalanine |
| Pro | Proline |
| Ser | Serine |
| Thr | Threonine |
| Trp | Tryptophan |
| Tyr | Tyrosine |
| Val | Valine |
| xLeu | Leucine/Isoleucine |
| **Glycerophospholipids** | |
| lysoPCaC14:0 | Lysophosphatidylcholine acyl C14:0 |
| lysoPCaC16:0 | Lysophosphatidylcholine acyl C16:0 |
| lysoPCaC16:1 | Lysophosphatidylcholine acyl C16:1 |
| lysoPCaC17:0 | Lysophosphatidylcholine acyl C17:0 |
| lysoPCaC18:0 | Lysophosphatidylcholine acyl C18:0 |
| lysoPCaC18:1 | Lysophosphatidylcholine acyl C18:1 |
| lysoPCaC18:2 | Lysophosphatidylcholine acyl C18:2 |
| lysoPCaC20:3 | Lysophosphatidylcholine acyl C20:3 |
| lysoPCaC20:4 | Lysophosphatidylcholine acyl C20:4 |
| lysoPCaC24:0 | Lysophosphatidylcholine acyl C24:0 |
| lysoPCaC26:0 | Lysophosphatidylcholine acyl C26:0 |
| lysoPCaC26:1 | Lysophosphatidylcholine acyl C26:1 |
| lysoPCaC28:0 | Lysophosphatidylcholine acyl C28:0 |
| lysoPCaC28:1 | Lysophosphatidylcholine acyl C28:1 |
| PCaaC24:0 | Phosphatidylcholine diacyl C24:0 |
| PCaaC26:0 | Phosphatidylcholine diacyl C26:0 |
| PCaaC28:1 | Phosphatidylcholine diacyl C28:1 |
| PCaaC30:0 | Phosphatidylcholine diacyl C30:0 |
| PCaaC32:0 | Phosphatidylcholine diacyl C32:0 |
| PCaaC32:1 | Phosphatidylcholine diacyl C32:1 |
| PCaaC32:2 | Phosphatidylcholine diacyl C32:2 |
| PCaaC32:3 | Phosphatidylcholine diacyl C32:3 |
| PCaaC34:1 | Phosphatidylcholine diacyl C34:1 |

| | |
|---|---|
| PCaaC34:2 | Phosphatidylcholine diacyl C34:2 |
| PCaaC34:3 | Phosphatidylcholine diacyl C34:3 |
| PCaaC34:4 | Phosphatidylcholine diacyl C34:4 |
| PCaaC36:0 | Phosphatidylcholine diacyl C36:0 |
| PCaaC36:1 | Phosphatidylcholine diacyl C36:1 |
| PCaaC36:2 | Phosphatidylcholine diacyl C36:2 |
| PCaaC36:3 | Phosphatidylcholine diacyl C36:3 |
| PCaaC36:4 | Phosphatidylcholine diacyl C36:4 |
| PCaaC36:5 | Phosphatidylcholine diacyl C36:5 |
| PCaaC36:6 | Phosphatidylcholine diacyl C36:6 |
| PCaaC38:0 | Phosphatidylcholine diacyl C38:0 |
| PCaaC38:3 | Phosphatidylcholine diacyl C38:3 |
| PCaaC38:4 | Phosphatidylcholine diacyl C38:4 |
| PCaaC38:5 | Phosphatidylcholine diacyl C38:5 |
| PCaaC38:6 | Phosphatidylcholine diacyl C38:6 |
| PCaaC40:1 | Phosphatidylcholine diacyl C40:1 |
| PCaaC40:2 | Phosphatidylcholine diacyl C40:2 |
| PCaaC40:3 | Phosphatidylcholine diacyl C40:3 |
| PCaaC40:4 | Phosphatidylcholine diacyl C40:4 |
| PCaaC40:5 | Phosphatidylcholine diacyl C40:5 |
| PCaaC40:6 | Phosphatidylcholine diacyl C40:6 |
| PCaaC42:0 | Phosphatidylcholine diacyl C42:0 |
| PCaaC42:1 | Phosphatidylcholine diacyl C42:1 |
| PCaaC42:2 | Phosphatidylcholine diacyl C42:4 |
| PCaaC42:4 | Phosphatidylcholine diacyl C42:4 |
| PCaaC42:5 | Phosphatidylcholine diacyl C42:5 |
| PCaaC42:6 | Phosphatidylcholine diacyl C42:6 |
| PCaeC30:0 | Phosphatidylcholine acyl-akyl C30:0 |
| PCaeC30:1 | Phosphatidylcholine acyl-akyl C30:1 |
| PCaeC30:2 | Phosphatidylcholine acyl-akyl C30:2 |
| PCaeC32:1 | Phosphatidylcholine acyl-akyl C32:1 |
| PCaeC32:2 | Phosphatidylcholine acyl-akyl C32:2 |
| PCaeC34:0 | Phosphatidylcholine acyl-akyl C34:0 |
| PCaeC34:1 | Phosphatidylcholine acyl-akyl C34:1 |
| PCaeC34:2 | Phosphatidylcholine acyl-akyl C34:2 |
| PCaeC34:3 | Phosphatidylcholine acyl-akyl C34:3 |
| PCaeC36:0 | Phosphatidylcholine acyl-akyl C36:0 |
| PCaeC36:1 | Phosphatidylcholine acyl-akyl C36:1 |
| PCaeC36:2 | Phosphatidylcholine acyl-akyl C36:2 |
| PCaeC36:3 | Phosphatidylcholine acyl-akyl C36:3 |
| PCaeC36:4 | Phosphatidylcholine acyl-akyl C36:4 |
| PCaeC36:5 | Phosphatidylcholine acyl-akyl C36:5 |
| PCaeC38:0 | Phosphatidylcholine acyl-akyl C38:0 |

| | |
|---|---|
| PCaeC38:1 | Phosphatidylcholine acyl-akyl C38:1 |
| PCaeC38:2 | Phosphatidylcholine acyl-akyl C38:2 |
| PCaeC38:3 | Phosphatidylcholine acyl-akyl C38:3 |
| PCaeC38:4 | Phosphatidylcholine acyl-akyl C38:4 |
| PCaeC38:5 | Phosphatidylcholine acyl-akyl C38:5 |
| PCaeC38:6 | Phosphatidylcholine acyl-akyl C38:6 |
| PCaeC40:1 | Phosphatidylcholine acyl-akyl C40:1 |
| PCaeC40:2 | Phosphatidylcholine acyl-akyl C40:2 |
| PCaeC40:3 | Phosphatidylcholine acyl-akyl C40:3 |
| PCaeC40:4 | Phosphatidylcholine acyl-akyl C40:4 |
| PCaeC40:5 | Phosphatidylcholine acyl-akyl C40:5 |
| PCaeC40:6 | Phosphatidylcholine acyl-akyl C40:6 |
| PCaeC42:0 | Phosphatidylcholine acyl-akyl C42:0 |
| PCaeC42:1 | Phosphatidylcholine acyl-akyl C42:1 |
| PCaeC42:2 | Phosphatidylcholine acyl-akyl C42:2 |
| PCaeC42:3 | Phosphatidylcholine acyl-akyl C42:3 |
| PCaeC42:4 | Phosphatidylcholine acyl-akyl C42:4 |
| PCaeC42:5 | Phosphatidylcholine acyl-akyl C42:5 |
| PCaeC44:3 | Phosphatidylcholine acyl-akyl C44:3 |
| PCaeC44:4 | Phosphatidylcholine acyl-akyl C44:4 |
| PCaeC44:5 | Phosphatidylcholine acyl-akyl C44:5 |
| PCaeC44:6 | Phosphatidylcholine acyl-akyl C44:6 |
| **Sphingolipids** | |
| SM(OH)C14:1 | Hydroxysphingomyeline C14:1 |
| SM(OH)C16:1 | Hydroxysphingomyeline C16:1 |
| SM(OH)C22:1 | Hydroxysphingomyeline C22:1 |
| SM(OH)C22:2 | Hydroxysphingomyeline C22:2 |
| SM(OH)C24:1 | Hydroxysphingomyeline C24:1 |
| SMC16:0 | Sphingomyeline C16:0 |
| SMC16:1 | Sphingomyeline C16:1 |
| SMC18:0 | Sphingomyeline C18:0 |
| SMC18:1 | Sphingomyeline C18:1 |
| SMC20:2 | Sphingomyeline C20:2 |
| SMC24:0 | Sphingomyeline C24:0 |
| SMC24:1 | Sphingomyeline C24:1 |
| SMC26:0 | Sphingomyeline C26:0 |
| SMC26:1 | Sphingomyeline C26:1 |
| **Monosaccharides** | |
| H1 | Sum of hexoses |

Supplementary Table 2. List of lipoprotein subfractions from $^{1}$H NMR analysis of plasma samples.

| Subclass abbreviation | Matrix | Analyte |
|---|---|---|
| TPTG | Total plasma | Triglycerides |
| TPCH | Total plasma | Cholesterol |
| TPFC | Total plasma | Free cholesterol |
| TPA1 | Total plasma | Apo-A1 |
| TPA2 | Total plasma | Apo-A2 |
| TPAB | Total plasma | Apo-B |
| VLTG | VLDL | Triglycerides |
| VLCH | VLDL | Cholesterol |
| VLFC | VLDL | Free cholesterol |
| VLPL | VLDL | Phospholipids |
| VLAB | VLDL | Apo-B |
| IDTG | IDL | Triglycerides |
| IDCH | IDL | Cholesterol |
| IDFC | IDL | Free cholesterol |
| IDPL | IDL | Phospholipids |
| IDAB | IDL | Apo-B |
| LDTG | LDL | Triglycerides |
| LDCH | LDL | Cholesterol |
| LDFC | LDL | Free cholesterol |
| LDPL | LDL | Phospholipids |
| LDAB | LDL | Apo-B |
| HDTG | HDL | Triglycerides |
| HDCH | HDL | Cholesterol |
| HDFC | HDL | Free cholesterol |
| HDPL | HDL | Phospholipids |
| HDA1 | HDL | Apo-A1 |
| HDA2 | HDL | Apo-A2 |
| V1TG | VLDL-1 | Triglycerides |
| V1CH | VLDL-1 | Cholesterol |
| V1FC | VLDL-1 | Free cholesterol |
| V1PL | VLDL-1 | Phospholipids |
| V2TG | VLDL-2 | Triglycerides |
| V2CH | VLDL-2 | Cholesterol |
| V2FC | VLDL-2 | Free cholesterol |
| V2PL | VLDL-2 | Phospholipids |
| V3TG | VLDL-3 | Triglycerides |
| V3CH | VLDL-3 | Cholesterol |
| V3FC | VLDL-3 | Free cholesterol |
| V3PL | VLDL-3 | Phospholipids |
| V4TG | VLDL-4 | Triglycerides |

| | | |
|---|---|---|
| V4CH | VLDL-4 | Cholesterol |
| V4FC | VLDL-4 | Free cholesterol |
| V4PL | VLDL-4 | Phospholipids |
| V5TG | VLDL-5 | Triglycerides |
| V5CH | VLDL-5 | Cholesterol |
| V5FC | VLDL-5 | Free cholesterol |
| V5PL | VLDL-5 | Phospholipids |
| V6TG | VLDL-6 | Triglycerides |
| V6CH | VLDL-6 | Cholesterol |
| V6FC | VLDL-6 | Free cholesterol |
| V6PL | VLDL-6 | Phospholipids |
| L1TG | LDL-1 | Triglycerides |
| L1CH | LDL-1 | Cholesterol |
| L1FC | LDL-1 | Free cholesterol |
| L1PL | LDL-1 | Phospholipids |
| L1AB | LDL-1 | Apo-B |
| L2TG | LDL-2 | Triglycerides |
| L2CH | LDL-2 | Cholesterol |
| L2FC | LDL-2 | Free cholesterol |
| L2PL | LDL-2 | Phospholipids |
| L2AB | LDL-2 | Apo-B |
| L3TG | LDL-3 | Triglycerides |
| L3CH | LDL-3 | Cholesterol |
| L3FC | LDL-3 | Free cholesterol |
| L3PL | LDL-3 | Phospholipids |
| L3AB | LDL-3 | Apo-B |
| L4TG | LDL-4 | Triglycerides |
| L4CH | LDL-4 | Cholesterol |
| L4FC | LDL-4 | Free cholesterol |
| L4PL | LDL-4 | Phospholipids |
| L4AB | LDL-4 | Apo-B |
| L5TG | LDL-5 | Triglycerides |
| L5CH | LDL-5 | Cholesterol |
| L5FC | LDL-5 | Free cholesterol |
| L5PL | LDL-5 | Phospholipids |
| L5AB | LDL-5 | Apo-B |
| L6TG | LDL-6 | Triglycerides |
| L6CH | LDL-6 | Cholesterol |
| L6FC | LDL-6 | Free cholesterol |
| L6PL | LDL-6 | Phospholipids |
| L6AB | LDL-6 | Apo-B |
| H1TG | HDL-1 | Triglycerides |
| H1CH | HDL-1 | Cholesterol |

| | | |
|---|---|---|
| H1FC | HDL-1 | Free cholesterol |
| H1PL | HDL-1 | Phospholipids |
| H1A1 | HDL-1 | Apo-A1 |
| H1A2 | HDL-1 | Apo-A2 |
| H2TG | HDL-2 | Triglycerides |
| H2CH | HDL-2 | Cholesterol |
| H2FC | HDL-2 | Free cholesterol |
| H2PL | HDL-2 | Phospholipids |
| H2A1 | HDL-2 | Apo-A1 |
| H2A2 | HDL-2 | Apo-A2 |
| H3TG | HDL-3 | Triglycerides |
| H3CH | HDL-3 | Cholesterol |
| H3FC | HDL-3 | Free cholesterol |
| H3PL | HDL-3 | Phospholipids |
| H3A1 | HDL-3 | Apo-A1 |
| H3A2 | HDL-3 | Apo-A2 |
| H4TG | HDL-4 | Triglycerides |
| H4CH | HDL-4 | Cholesterol |
| H4FC | HDL-4 | Free cholesterol |
| H4PL | HDL-4 | Phospholipids |
| H4A1 | HDL-4 | Apo-A1 |
| H4A2 | HDL-4 | Apo-A2 |