

# Exploring metabolomic biomarkers using mass spectrometry techniques

Kathryn Trudy Coates

Submitted to Swansea University in fulfilment of the requirements for the Degree of Doctor of Philosophy

Swansea University

Swansea University Medical School

Supervised by Dr E. Dudley and Professor C. Thornton

Submission: 2023

## Summary

Mass spectrometry based metabolomic approaches prove an excellent source for biomarker discovery through their ability to seek out robust and unique disease signatures in a fast, accurate and high throughput manner whilst delivering a great deal of sensitivity and biological detail from a broad range of sample source. This study has identified a number of unique and novel biological signatures to improve diagnosis/ metabolic pathway understanding for a broad range of conditions.

**Chapter 2-** Pancreatic cancer is a disease that occludes diagnosis through its lack of unique symptomatic presentation in patients leading to late-stage diagnosis, and low survival rates. This chapter successfully identifies novel metabolites capable of distinguishing pancreatic cancer from healthy and pancreatitis conditions using both internally held metabolites and externally excreted modified nucleosides for both diagnosis and screening potential. Further experimental analysis using SILAC to trace glucose and glutamine use by cancerous and healthy cells has been used to elucidate differences in energy resource utilisation.

**Chapter 3-** Looks at the development of a metabolomic analysis capable of distinguishing various serum-based conditions from each other as well as analysing metabolic profile changes between males, females (pregnant and non-pregnant) and cord samples to establish key profile changes that occur in pregnancy as well as establishing how these changes may be conveyed in breast-milk production to further support baby development after birth.

**Chapter 4-** Captures faecal markers capable of distinguishing successfully from unsuccessfully reproducing captive elephant, seeking diagnostic and explanatory, indicating potential upstream mis-regulations that could be causative of the ailment, markers aiding to potential treatments or future preventions to aid reproduction.

**Chapter 5-** Evaluates drosophila fly as a suitable model for future neurological studies including Alzheimer's disease. This study has successfully identified numerous lipids, including plasmalogens, in fly heads that make them a suitable candidate in future studies as well as showing the lipidomic changes resultant of disrupted metabolic pathways associated with Alzheimer's disease.

**Chapter 6-** Establishes the relationship between taurine consumption and exercise performance relating to thermoregulation, through detecting and analysing taurine's presence in plasma samples taken before and after exercise periods.

## Declarations and statements

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed .....Kathryn Coates.....

Date .....

This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by footnotes giving explicit references.

Signed .....Kathryn Coates.....

Date .....

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed .....Kathryn Coates.....

Date .....

The university's ethical procedures have been followed and, where appropriate, that ethical approval has been granted

Signed .....Kathryn Coates.....

Date .....

## Contents

Exploring metabolomic biomarkers using mass spectrometry techniques.....	1
Summary.....	2
Acknowledgements .....	8
List of Figures.....	9
List of Tables and illustrations .....	24
Definitions and Abbreviations .....	26
Chapter 1 -Introduction .....	27
1.1 Metabolomics.....	27
1.2 Biomarkers.....	28
1.6.2- Biomarker critique using ROC curves .....	29
1.3 Targeted vs Untargeted.....	30
1.4 Sample sources.....	31
1.5 Mass Spectrometry and spectroscopy techniques for use as an analytical platform.....	40
1.5.1- Gas chromatography coupled to mass spectrometry- GC-MS.....	41
1.5.2- Liquid Chromatography mass spectrometry- LC-MS .....	42
1.5.3- Matrix Assisted Laser Desorption/Ionization- MALDI .....	44
1.5.4- Orbitrap .....	46
1.6 Data analysis.....	47
1.6.1- Statistical evaluations.....	48
1.6.3- Identification provision.....	49
1.7- Ion suppression .....	51
1.8- Thesis structure .....	51
1.9- COVID Impact statement.....	52
1.10- References.....	53
Chapter 2 – Metabolomic analysis of potential cellular biomarkers and the study of altered metabolism in Pancreatic and Lung cancer.....	55
2.1- Introduction.....	55
2.1.1. Chapter aims and objectives .....	59
2.2 Methodology .....	60
2.2.1- Cell lines.....	60

2.2.2. Urine modified nucleoside analysis .....	66
2.2.3. Biomarker evaluation .....	67
2.3 Results .....	68
2.3 1. GCMS.....	68
2.3.3. MALDI Analysis.....	94
2.3.4. Urinary modified nucleoside analysis .....	114
2.3.5 SILAC metabolic pathway analysis.....	129
2.4. Discussion .....	148
GCMS .....	149
2.5- References.....	172
Chapter 3 – Human fertility and maternity and metabolic dynamics .....	183
3.1- Introduction.....	183
3.1.1- Serum biomarkers profile in distinct demographics .....	185
3.1.2- Female serum fertility exploration .....	185
3.1.3- Lipidomic evaluation of pregnancy dynamics and neonatal requirements .....	187
3.1.4- Chapter aims and objectives .....	189
3.1.5- The objectives of this chapter are as follows: .....	190
3.2. Methodology .....	191
3.2.1 Samples.....	191
3.2.2 Metabolite extraction and analysis .....	191
3.2.3- Statistical analysis.....	193
3.3- Results .....	194
3.3.1- Comparisons of males, non-pregnant females, pregnant females and cord serum .....	194
3.3.2- Foetal lipidomic demands Breast-milk .....	231
3.3.2.2- Negative mode lipidomic analysis .....	232
3.3.3- Amino/ organic acid biomarker exploration PCOS and Endometriosis .....	236
3.4. Discussion .....	241
3.4.1- Discussion of serum markers that can distinguish male, pregnant and non-pregnant female and cord candidates. ....	242
3.4.2. Exploring metabolite requirements of foetus and neonates. ....	248
3.4.3.- Seeking novel fertility biomarkers for endometriosis and PCOS diagnosis. .....	251

3.5. Conclusion .....	253
3.6. References .....	254
Chapter 4 – Elephant faecal metabolite analysis for the determination of reproductive status .....	259
4.1- Introduction.....	259
4.1.2- Aims .....	262
4.2- Methods and Materials .....	263
4.2.1 Samples.....	263
4.2.2 Metabolite extraction optimisation .....	263
4.2.3 Metabolite analysis by animal group.....	265
4.2.4 Metabolite data processing.....	265
4.3- Results .....	266
4.3.1 Extraction methods and their evaluation.....	266
4.3.2. Elephant reproductive status biomarker analysis .....	271
4.4- Discussion .....	279
4.4.1- Future perspectives .....	282
4.5- References.....	283
Chapter 5 - Seeking plasmalogen biomarkers from Drosophila flies for utilisation as a model for Alzheimer’s disease .....	288
5.1- Introduction.....	288
5.1.1- Lipids and neurological disruptions.....	288
5.1.2- Alzheimer’s disease .....	295
5.1.3- Drosophila.....	296
5.1.4- Aims .....	298
5.2- Methodology .....	299
5.2.1- Samples.....	299
5.2.2- Extraction.....	299
5.2.3- Data analysis.....	300
5.3- Results .....	301
5.3.1- Standard curves .....	301
5.3.2- Biological Extracts.....	303
5.3.1- Plasmalogens .....	313
5.3.3- Mutant flies .....	316

5.3.3- Plasménylethanolamines.....	326
5.3.4- Plasmányl GP .....	327
5.3.5- LPE .....	327
5.3.6- PS and PI presence .....	328
5.4- Discussion .....	333
5.5- Conclusion and future perspectives.....	339
5.5.1- Future work .....	339
5.6- References.....	340
Chapter 6 – Analysing the thermoregulatory properties of taurine under heat stress .....	348
6.1- Introduction.....	348
6.1.2- Taurine intake.....	349
6.1.3- Regulation and uptake .....	350
6.1.5- Roles .....	351
6.1.6- Sport impact .....	352
6.1.7- Role in thermoregulation .....	354
6.1.8- Chapter aims.....	355
6.2- Methodology .....	355
6.2.1- Patient samples .....	355
6.2.2- Quantification using HPLC .....	357
6.3- Results .....	358
6.3.1- Optimization .....	358
6.3.1- Taurine quantification .....	362
6.3.2- Thermoregulation participant measures- analysing sweating capability and impacts.....	366
6.4- Discussion .....	373
6.5- Conclusion .....	376
6.5- References.....	378
Chapter 7 – Conclusions and future perspectives .....	383

## Acknowledgements

Firstly, I would like to express my gratitude to my supervisory team. Without the knowledge, inspiration and guidance provided throughout the course of this project, especially from Dr Ed Dudley, I would have undoubtedly never have made it through my thesis. Dr Dudley's mass spectrometry expertise has been extremely valuable as well as his constant support and rapid feedback to help me see this project through.

I would like to thank Professor Cathy Thornton and Dr April Rees for sample provision as well as advice and guidance on various elements of this project whenever required. Dr Claire Morgan, Lydia Farrell and Dr Ume-Kulsoom Shah for supplying various cancerous cell lines and providing guidance on their maintenance, especially after the loss of my own stocks during the unforgiving pandemic. I would like to thank Mr Tim Brown (vascular surgeon, Morriston Hospital) for providing the patient urine samples with diagnosis.

I would like to thank Roberto Angellini who has lead and inspired the drosophila project and provided vast guidance on lipidomic analysis utilised through-out this thesis.

I would also like to thank the staff at the EPSRC UK National Mass Spectrometry Facility, in particular Dr Ann Hunter and Gareth Llewellyn, for enabling the use of their mass spectrometers and assisting with any and all issues that arose. Dr Ann Hunter's invaluable knowledge and advice on all of the mass spectrometry elements helped me a lot in this project, including providing me with the support and knowledge to maintain and repair equipment after issues arose as well as many supportive tea breaks to help brainstorm new ideas.

Last, but not least, my wonderful wife Ria for supporting me through-out this thesis with encouragement in trying times, humouring my work discussions seeking advice on data presentation and countless cups of coffee provided during write-up times. Also a shout-out to our unborn baby "Flash" (edit: Otis!) for encouraging me to stick to dead-lines and ensuring that I complete this write-up ahead of their due date in November.



## List of Figures

Figure 1-1 Graphical representation of the number of metabolomic journals published per year available to access on NCBI ( <a href="https://www.ncbi.nlm.nih.gov/">https://www.ncbi.nlm.nih.gov/</a> ) using the search terms “Metabolomics”, “Metabolomics” + “Mass spectrometry” and “Metabolomics” + “Pancreatic cancer” correct as of 21/09/2023. ....	29
Figure 1-2- Exemplar ROC curve used to evaluate a biomarkers diagnostic capability in the analysis, which generated a Sensitivity value of 0.867 and 1-specificity value of 0.033 at the most efficacious point of the graph (circled) with an area under the curve scoring of 0.96. ....	30
Figure 1-3- Schematic for the isolation of serum from whole blood involving its clotting for 30 minutes, followed by centrifugation at room temperature and supernatant collection (Kardos, D., et al., 2019) .....	35
Figure 1-4- Schematic showing the acquisition of plasma from whole blood after the addition of EDTA the blood is allowed to clot for 30 minutes prior to centrifugation at room temperature and supernatant collection (Kardos, D., et al., 2019).....	36
Figure 1-5-Illustration of a triple quadrupole, made up of three consecutive quadrupoles for improved mass filtering before detection (Pitt J. J. 2009).....	42
Figure 1-6 Illustration of the function of ESI whereby the sample is pumped through metal capillary (3-5kV) before being nebulised into a fine spray at the tip. Skimmers are used for ion focalisation and heat applied either from nitrogen source (top) or a heated capillary (bottom) (De Hoffman, E., Stroobant, V., 2007). ....	44
Figure 1-7- Schematic showing the principle of MALDI analysis whereby the sample mixed with matrix has been fired with a laser, resulting in energy transfer and a plume of matrix and sample to be released from the surface (De Hoffman, E., Stroobant, V., 2007) .....	45
Figure 1-8-Illustration of an orbitrap showing the injection of ions into the orbitrap analyzer via the c-trap. They then oscilate around the central electrode and between the outer electrodes maintained by an altering electric field being applied. After separation ions are then amplified and detected (Zubarev, R. A., & Makarov, A., 2013). ....	47
Figure 1-9- Example spectral identification made using NIST libraries couple to AMDIS whereby the A) peak recorded at retention time 20.8058 has generated an identity match with B) Hexadecenoic acid with a confidence scoring of weighted 92% and reverse 97%. A comparison is also made between the samples spectra C) and the NIST library spectra D). ....	50
Figure 2-1 PCA showing the relationship of all of the different cell line groups when considering the most significantly different metabolites, comparing all cell lines (A) and just the lung cancer cell lines (B). ....	69
Figure 2-2. PCA chart showing the relationship between the pancreatic groups significantly different markers including HDPE (healthy), Mia PaCa-2 and Panc-2 (cancerous).....	70

Figure 2-3- Significantly different metabolites that were successfully identified that differentiate A549 cancerous cell lines from BEAS-2B non-cancerous cell lines whilst achieving better than “good” diagnostic scorings. All p. values obtained were less than 0.007. Error bars represent standard error, n= 9 for each cell line. “Methane” represents Methane phosphophonic acid. ....	71
Figure 2-4 Graphical representation of the most significantly different metabolite markers that distinguish HDPE from Panc-1 pancreatic cells. These all achieved p. values of less than 0.005 and better than “good” diagnostic values relating to the AUC and LHR. ....	74
Figure 2-5 Significantly different metabolite markers that distinguish MiaPaca-2 cancerous models from HDPE healthy pancreatic models using internal metabolites that achieved better then “good” diagnostic scorings and provided sensible identifications via NIST. All p.values were below 0.008. ....	76
Figure 2-6- Heat map analysis of the variable metabolites identified in the analysis comparing cancerous and non-cancerous cell lines .....	80
Figure 2-7 Heatmap showing the metabolite profile changes and relationship between the different lung cell lines A549 and Beas-2B.....	82
Figure 2-8. Heatmap showing the variance in the significantly different markers between the different pancreatic groups and the impact this has one the inter group relationship .....	83
Figure 2-9 Amino acids that were found to be significantly upregulated in more than one of the groups comparison. The standard error bars are plotted, for n=9, to show the variation across the group. ....	84
Figure 2-10 Non-amino acid metabolites that were found to be significantly up-regulated in more than one of the groups comparisons. The standard error bars, where n=9, are plotted to illustrate the group variation.....	86
Figure 2-11 Example standard curves generated for amino acid standards a) threonine, b) aspartic acid, c) histidine and d) cysteine with R <sup>2</sup> values and equations of a straight line used for latter calculations shown on the graph that were used for evaluation and further calculations. Standard deviations are also shown for n=3... ..	88
Figure 2-12 PCA evaluation of amino acid relationships between different cell lines .....	92
Figure 2-13 Amino acids in lung cancer and pancreatic cancer cell lines arranged according to whether they are essential A) or non-essential B) in nature. Significant changes were made between each mutant group and the healthy groups values with bars noted with *>0.05 **>0.025*** > 0.005 .....	93
Figure 2-14 Percentage change of quantified amino acids in cancerous cell lines when compared to their healthy equivelent cell line. ....	94
Figure 2-19- Example spectra of Pancreatic cells analysed in negative mode using MALDI analysis with 9aa matrix. A) HDPE (Healthy) dark blue , B) Panc-1 (cancerous) purple, C) Mia PaCa-2 royal blue lower. ....	95
Figure 2-20- PCA analysis of the 5 different cell line groups analysed in negative MALDI analysis when assessed using the 20 common significantly different	

metabolite markers. A) shows the pancreatic cell lines, B) The lung cell lines and C) all cell lines in the analysis. ....	96
Figure 2-21- Heatmaps illustrating the variation and relationships between all cell lines groups understudy for altered lipidomic presence .....	97
Figure 2-22- Significantly different markers that distinguished Panc-1 and HDPE cell lines analysed in MALDI using negative mode analysis. ....	99
Figure 2-23- Significantly different markers that were found to be better than good diagnostically that distinguish MIA PaCa-2 from HDPE in MALDI operated in negative mode .....	100
Figure 2-24- Significantly different markers that distinguish Beas-2b and A549 cell lines from each other with a better than excellent biomarker rating. ....	101
Figure 2-25- Common significantly different metabolites found to distinguish all of the groups of interest from each other. ....	102
Figure 2-26- PCA plot showing the separation of the 5 different groups under analysis when considering the shared significantly different metabolite marker profiles .....	104
Figure 2-27- PCA plots generated to show the separation of a) Lung cancer and healthy lung cells b) healthy pancreatic cell lines and Panc-1/MiaPaCa-2 with confidence intervals shown with ovals. ....	105
Figure 2-28- Heatmaps generated using shared significantly different metabolites to analyse profile changes and group relationships between the 5 different groups under study. ....	106
Figure 2-29- Significantly different markers found to distinguish Panc-1 cell lines from HDPE healthy cell lines whilst maintaining a better than "good" biomarker quality under ROC analysis.....	107
Figure 2-30- Example spectra for HDPE and Mia PaCa-2 cell lines respectively with the three "good" diagnostic markers of interest labelled. ....	108
Figure 2-31- Significantly different metabolite marker found to distinguish MiaPaCa-2 cancerous cell lines from HDPE healthy cell lines whilst achieving a better than "good" ROC scoring. ....	109
Figure 2-32- Significantly different profile markers obtained using MALDI analysis in positive mode that distinguish cancerous and non-cancerous lung cells from each other whilst achieving better than "good" ROC scoring. Standard error bars are shown where n=9. here multiple identifications are possible these shown. Asterisks for significance are shown, with * $<0.05$ ** $<0.025$ , *** $<0.01$ . ....	110
Figure 2-33- Mean intensity comparison of lipid differences found to significantly discriminate lung cancer and healthy cell lines from each other whilst maintaining better than "good" ROC scoring as individual markers. ....	111
Figure 2-34 A) - shared significantly different fatty acids found to distinguish cancerous and non-cancerous cell lines from each other. B) Significantly different ceramide and DG lipids found to be significantly different between healthy and cancerous lines across more than one different cell line comparison. ....	111

Figure 2-35- Significantly different lysophospholipids found to discriminate multiple different cancerous cell lines from their healthy equivalent.....	113
Figure 2-36- Significantly different Sphingolipids/ Triacylglycerol markers found to differentiate multiple different group comparisons when looking at Beas-2B, A549, HDPE, Panc-1 and MIAPaCa-2 cell lines relative abundances. ....	113
Figure 2-37- Significantly different peptide markers found to distinguish multiple different cancerous groups from healthy equivalent cell lines .....	114
Figure 2-38- The average intensities of all of the modified nucleosides of interest comparing all of the different groups of interest; Pancreatic ductal adenocarcinoma, Pancreatitis, other cancer, healthy and other pancreatic conditions .....	123
Figure 2-39- Significantly upregulated modified nucleosides in Pancreatic ductal adenocarcinoma vs all other candidate groups combined. Significant differences have been signified with asterisks with * denoting a p. value of less than 0.05, ** less than 0.025 and *** less than 0.005. ....	124
Figure 2-40- Significantly different modified nucleoside intensities between healthy and all other group candidates combined. Significant differences have been signified with asterisks with one denoting a p. value of less than 0.05, two less than 0.025 and three less than 0.005.....	125
Figure 2-41- the significantly different modified nucleosides between pancreatic ductal adenocarcinoma and healthy cohorts. Significant differences have been signified with asterisks with one denoting a p. value of less than 0.05, two less than 0.025 and three less than 0.005.....	126
Figure 2-42- Alterations in nucleosides found to be significantly different between healthy and pancreatic ductal adenocarcinoma cohorts, with additional comparison of pancreatitis candidate levels. Significant differences have been signified with asterisks with one denoting a p. value of less than 0.05, two less than 0.025 and three less than 0.005.....	127
Figure 2-43 The average variation of methylguanosine levels between pancreatic ductal adenocarcinoma and all other groups under study.....	128
Figure 2-44- The metabolic pathway under evaluation in the study showing the carbon presence from glucose labelling (red) and glutamine (blue). Carbons are shown by the circles with coloured circles indicating a labelled carbon and a hollow circle no labelling upon the carbon.....	131
Figure 2-45- Identifying the fate of pyruvate produced during glucose metabolism A) Alterations in key metabolites for lung cell lines B) Alterations in key metabolites from pancreatic cell extracts C) Schematic showing the possible pathways pyruvate can enter and how these can be identified; I) can be measured using the generation of M+3 lactate II) Glutamine mediated amination into alanine, this can be measured utilising the presence of M+3 alanine III) Decarboxylation of pyruvate into acetyl CoA, this can be measured looking at the presence of M+2 citrate. IV) The carboxylation of pyruvate into oxaloacetate This can be measured looking at M+3 labelled citrate. Significant variance between the cancerous cell lines and healthy equivalent is illustrated with asterisk. ....	132

Figure 2-46- The TCA related compounds average relative abundance of labelled marker compared to unlabelled through incorporation of glucose (U <sup>13</sup> ) for A) Lung cell lines and B) Pancreatic cell lines with standard error plotted and significant values <0.05 indicated with asterisk. ....	135
Figure 2-47- A) Illustrates the pathway for Glutamine production that can produce detected isotopologues from the carbons incorporated from Glucose <sup>13</sup> B) Illustrates the proportion of each labelled form of glucose in comparison to unlabelled in Lung cell lines. B) Illustrates the proportion of each labelled form of glucose in comparison to unlabelled in Pancreatic cell lines.....	136
Figure 2-48- Potential isotopologue formation possible through the incorporation of uniformly labelled glutamine C <sup>13</sup> , upon entry into the TCA cycle via transamination. The black labelled metabolites show the initial route of entry into the cycle, the blue show the possible down-stream isotopologues. ....	138
Figure 2-49- comparative labelled abundances of glutamic acid isotopologues compared to unlabelled marker presence in A) Lung cell lines and B) Pancreatic cell lines .....	139
Figure 2-50- Average proportion of labelled isotopologue to unlabelled for glutamic acid in A) Lung cells and B) pancreatic cells with the absence of glucose from the media.....	140
Figure 2-51-Illustrates the proportion of M+2 glucose labelled and M+5 glutamine labelled glutamic acid present during glucose and glutamine tracing experiments respectively compared to the sum of all other labelled and unlabelled markers present. Illustrating the main initial influx of each marker for the metabolite of interest suggesting whether the glutamine has come from glucose or glutamine initially. ....	140
Figure 2-52- The proportional change in the presence of isotopologues of measurable TCA intermediates in A) lung cells and B) Pancreatic cells with metabolic pathways described in C) detailing the “traditional” entry of glutamine into the TCA cycle generating citrate isotopologue (M+4) B) The pyruvate salvage pathway leading from glutamine that could explain the presence of pyruvate (M+3) and citrate (M+2). E) The alternative pathway of glutamine to enable to formation of ((M+5) isotopologue of citrate.....	142
Figure 2-53 The average proportionate changes of labelled isotopologues of metabolites proportionate to unlabelled for A) Lung cell lines and B) Pancreatic cell lines. ....	144
Figure 2-54- Alterations in the proportion of labelled isotopologues and unlabelled for different metabolites detected that relate to the TCA cycle in A) Lung cells supplemented with glucose and labelled glutamine 2) Pancreatic cells supplemented with glucose and labelled glutamine 3) Lung cells supplemented with labelled glutamine but no glucose 4) Pancreatic cells supplemented with labelled glutamine but no glucose.....	145

Figure 2-55- The variation in labelled to unlabelled ratio of isotopologues of serine and glycine generated from labelled glucose metabolism in A) lung and B) Pancreatic cell lines .....	147
Figure 2-56- Average proportion of labelled to unlabelled isotopologue of amino acids after introduction of glucose labelled media in A) Lung cells and B) Pancreatic cells with standard error shown and significant variance to healthy equivalent cells shown with asterisk. ....	148
Figure 3-1 Example GCMS chromatogram for the four analysed groups: pregnant females (blue), Males (red), non-pregnant females (yellow) and cord serum samples (green). Figure 3.1B shows an extract of the chromatogram in more detail between 19 and 26 minutes.....	195
Figure 3-2 The average alterations in the significantly different metabolite profiles comparing non-pregnant females, males, pregnant females and cord serum. Relative abundance is shown with high expression in red and low expression in dark blue.....	197
Figure 3-3 Dendrogram illustrating the relationship between the different samples and groups depicting red labels for non-pregnant females, green for males, Blue for pregnant females and red for cord serum samples.....	197
Figure 3-4- PCA chart showing the differentiation of males and non-pregnant women's serum samples from each other using GCMS analysis using biomarkers found to be the most diagnostically capable of distinguishing the groups.....	198
Figure 3-5- The average normalised relative abundance of the 22 most significantly different metabolites identified between males (red) and non-pregnant females (blue), in order of most to least significantly different, with standard error plotted. Bars represent normalised average abundance, with error bars representing the standard error (n=30 for both groups). X- denotes an unidentified metabolite. Significance is shown with asterisk with *>0.05 **>0.025 *** > 0.005 .....	199
Figure 3-6-PCA chart showing the separation of non-pregnant and pregnant females serum metabolites analysed using GCMS analysis. ....	202
Figure 3-7- The 21 most significantly different metabolites identified between non-pregnant and pregnant females ordered from most significant on the left. Bars represent normalised average abundance, with error bars representing the standard deviation (n=30 for both groups) .....	203
Figure 3-8- The most diagnostically successful metabolites capable of distinguishing pregnant (blue) and cord (red) serum samples from each other in order of significance. The average relative abundance for the groups is illustrated on a broken scale with standard error bars representing variation in the groups (n=30 for both groups).....	206
Figure 3-9- Graph illustrating the average abundance of the metabolites indicated as being significantly different between the four different groups with error bars shown (n=30). ....	208

Figure 3-10 Example MALDI spectra for Non-pregnant females (dark blue), Males (red), cord (pink) and pregnant females serum samples (royal blue) obtained using 9-AA matrix in negative reflectron mode. ....	210
Figure 3-11- PCA chart showing the differentiation of cord, males, pregnant and non-pregnant women's serum samples from each other utilising MALDI-TOF analysis in negative mode using biomarkers found to be the most diagnostically capable at distinguishing multiple groups of interest. ....	211
Figure 3-12- Heatmap illustrating the average intensity comparison of the shared markers of interest between the 4 different demographical groups of interest; males, non-pregnant females, pregnant females and cord serum. ....	212
Figure 3-13- Overlaid Non-pregnant (blue) and male (red) MALDI spectra showing two of the significantly different markers identified. One was highlighted as “excellent” in the analysis, 497.5. With a further two 498.5 and 499.5 which were identified as “good” which actually show isotopic variants of the principle 497.5. This is clear from the approximate third decrease in intensity seen between each peak as it increases in $m/z$ by 1 unit. These can therefore all be assigned as being the same marker. The other is at 511.5.....	213
Figure 3-14- Average peak intensities for the significantly different metabolites found to differentiate non-pregnant females from male serum cohorts that were rated better than good diagnostically. Error bars represent standard deviation. N=30 for both cohorts.....	214
Figure 3-15 Overlaid non-pregnant (blue) and male (red) MALDI spectra showing two of the significantly different markers identified. Both of which were highlighted as diagnostically “excellent” in the analysis. The first of these is sat $m/z$ 564.5, followed by a further 2 isotopic variants at 565.5 and 566.5. The other marker is $m/z$ of 581.5 and is found to be significantly up-regulated in males when being compared to non-pregnant females.....	216
Figure 3-16 Average relative abundance of the metabolites which were major sources of variation between non-pregnant females from pregnant females. Error bars represent standard deviation. N=30 for both cohorts.....	217
Figure 3-17 The most diagnostically capable biomarkers, rated as “good” found to distinguish Pregnant women and cord serum from each other using MALDI-TOF in negative mode. Error bars represent standard deviation. N=30 for both cohorts. The asterisks illustrate markers rated as better than excellent diagnostically. ....	218
Figure 3-18- The common most biologically capable biomarkers found to be able to distinguish more than one of the demographical groups, male, non-pregnant females, pregnant females and cord serum from each other.....	220
Figure 3-19. Example MALDI spectra of lipid extraction from a non-pregnant females serum ran in positive ion mode; a) displays a full scan focused on the area of most interest $m/z$ 's between 500 and 860 b) focused on the LPC content under analysis C) Areas with SM and LPC .....	222

Figure 3-20- PCA plot showing the division of the different groups utilising the most significantly different metabolites identified that appeared to distinguish multiple groups.....	223
Figure 3-21- Heatmap showing the average metabolite intensity change of the most distinguishing markers used to generate the PCA plot including an illustration of the these impact the way in which the groups relate to each other.....	223
Figure 3-22- Average intensities of the lipids capable of distinguishing non-pregnant females from pregnant females with corresponding standard error bars shown. N=30 for both.....	225
Figure 3-23 Average intensities of the most significantly different <i>m/z</i> diagnostically capable of distinguishing pregnant females (red) from cord (orange) serum samples with corresponding standard error bars shown (N=30). ....	227
Figure 3-24 Average intensities of the most significantly different <i>m/z</i> diagnostically capable of distinguishing all of the different groups under analysis with corresponding standard error bars shown (N=30 per group, 120 total). Significance is shown with symbols whereby a comparison has been made between the bar denoted and n being non- pregnant females, m- males, p- pregnant females and c being cord. Red symbol shows a significance of less than 0.5, orange symbolises less than 0.01 and green less than 0.001.....	228
Figure 3-25- PCA plot showing the division of the different groups utilising the most significantly different metabolites identified that appeared to distinguish multiple groups.....	229
Figure 3-26- Heatmap showing the average metabolite intensity change of the most distinguishing markers used to generate the PCA plot including an illustration of the these impact the way in which the groups relate to each other.....	229
Figure 3-27- Average total presence of phosphatidylcholine (PC) and lysophosphatidylcholine (LPC) in each different demographical group.....	229
Figure 3-28- Graphical representation of the average variation of phosphatidylcholine (PC)/ Lysophosphatidylcholine (LPC) in each demographical group .....	230
Figure 3-29. Significantly different biomarkers found to differentiate breastmilk collected at 2 and 6 months post-partum analysed in positive mode analysis. All of the markers achieved better than “good” ROC scoring showing high quality and capability of the metabolites as individual markers. All markers achieved a significance score p. value < 0.001.....	231
Figure 3-30. Markers found to be significantly up-regulated in pregnant serum samples when compared to cord serum with additional notation of alterations in breastmilk samples between 2 and 6 months. Significant variance between pregnant/ cord and 2 and 6 month breast-milk is illustrated with asterisks whereby *<0.05, **< 0.01 & *** < 0.001. Tentative ID’s are shown where possible, where these could not be generated they have been denoted with an x.....	233



Figure 3-31. Comparison of the most distinguishing biomarkers found to differentiate pregnant and cord serum, to those that were also seen in breastmilk samples comparing 2 and 6 months post-partum. See text for details. ....	235
Figure 3-32 PCA analysis comparing the amino acid profiles of healthy, endometriosis and polycystic ovary syndrome (PCOS) candidate's serum. A) Compares the profiles when considering all amino acids under evaluation. There is a division between healthy and PCOS candidates with the former largely being congregated to the left of the chart. The confidence intervals do overlap (shown by the ovals) B) Considers only the significantly different metabolites. This shows a slightly better differentiation between healthy and endometriosis samples where the healthy candidates are found towards the top right-hand corner of the chart and the endometriosis candidates towards the bottom left. PCOS samples in this instance are congregated towards the top left of the chart, with overlapping confidence intervals. ....	236
Figure 3-33 heatmaps showing the intensity variation of the significantly different amino acid markers within and between different groups of interest. Shows the individual groups fluctuation with 83% of the healthy group branching off together at the first major arm with one polycystic ovary syndrome (PCOS) candidate group. the other major branch consists of two minor arms, one consisting of one of the endometriosis groups and the other further sub-dividing into one arm consisting of all of the PCOS members and the other the remaining two endometriosis samples. ....	238
Figure 3-34. PLSDA analysis comparing the significantly different amino acids relationship between healthy and A) endometriosis B) Polycystic ovary syndrome (PCOS) candidate serums. Both charts generate a very good group separation with all group candidates clearly defined from each other and minimal confidence interval overlap depicting good group separation using these markers. ....	239
Figure 3-35 Amino acid concentrations observed in serum samples of healthy (blue), polycystic ovary syndrome (PCOS) (red) and endometriosis (green) candidates with statistical significance between healthy and PCOS/endometriosis indicated with asterisks whereby $* < 0.05$ , $** < 0.01$ , $*** < 0.001$ .....	240
Figure 3-36 Average total relative intensity of all amino acids present in serum of healthy, polycystic ovary syndrome (PCOS) and endometriosis candidates.....	241
Figure 3-37- Average total intensity of all amino acids measured using GCMS analysis in the four different study groups serum; non-pregnant (blue), male (red), pregnant (green) and cord (purple). ....	246
Figure 4-1- Comparison of raw GCMS spectra from the same pooled elephant sample extracted using acetone, methanol, acetonitrile, dichloromethane, chloroform: methanol, dichloromethane: methanol, Ethyl acetate and water with a 10kDa filter respectively. ....	267
Figure 4-2- The numbers of metabolites captured by the three different extraction methods used.....	269

Figure 4-3- Heat map showing the intensities of the 25 most distinguishing features between the three different extraction methods used- illustrating the majority of these to be generated by methanol extraction .....	270
Figure 4-4- Metabolites with differences between A) non cycling and cycling elephants, B) cycling and pregnant elephants and C) non cycling and pregnant elephants. D) metabolites shown to be elevated in non-cycling elephants when compared to the other two groups. Error bars= SD. Relevant statistical P values summarised in Table 4-1.....	275
Figure 4-5- Heat map detailing the differentiation of the elephants within the three groups based upon normalised levels of the metabolites identified as having discriminatory power. ....	278
Figure 5-1- Lipid nomenclature with focus on Glycerophospholipid group glycerophosphoethanolamine (PE) .....	290
Figure 5-2- Biosynthesis and degradation pathways of plasmalogens within cells. The biosynthesis is catalysed by the enzymes <b>1) GNPAT 2) FAR1 3) AGPS</b> 4)AADHAP in the peroxisome and 5) AAG3P-AT 6) PAP-1 7) EPT and <b>8)PEDS1</b> in the endoplasmic reticulum. 9) alkyl glycerol kinase phosphorylates AG which then crosses the membrane entering the endoplasmic reticulum (Bozelli J.C.JR and Epand R.M., 2021). ....	294
Figure 5-3- A) Plasmalogen (PE-P) and B) phosphatidylethanolamine (PE) standard curves showing linearity and good reproducibility for use in quantification of lipid classes. Standards have been normalised to the ISTD and have been extracted in the same way as samples were prepared. The PE-P standards were checked with the separate preparation of a quality control samples (QC) at a concentration of 6µM to ensure that this could be correctly quantified from the graphs. ....	301
Figure 5-4- Plot showing a reproducibility assessment for the quantification of A) plasmalogen and B) phosphatidylethanolamine using the standard curves generated in Figure 5-7. This was conducted through taking the values of the first replicate set and dividing by the second or third replicate set and plotting the values generated were then divided from the values of replicate 1 sample set. The strong values of R <sup>2</sup> generated the good level of reproducibility achievable using this range and methodology. ....	302
Figure 5-5- Box and whisker plot demonstrating the linear range of each mutant extract vs the dynamic range of the standard PE(16:1/16:1) with the majority of the detected phosphatidylethanolamine lipids falling within the working range of the standard curve but some dropping below, namely PE(O-16:0/18:3), PE(O-16:0/18:2), PE(18:0/18:1), PE(16:0/20:0) and PE(19:0/20:5) .....	303
Figure 5-6- Fragmentation patterns obtained for ion 716.5 illustrating the key fragmentation patterns associated with lipid PE(16:0/18:1). A) Illustrates the full MS spectra obtained for a healthy drosophila head extract with the peak of interest highlighted. B) MS2 structural fragmentation seen for PE(16:0/18:1); red 478.33 resultant of the loss of the sn-1 C16:0 acyl chain as ketene. 452.17 is shown in green which is resultant of the loss of C18:1 sn2 as ketene. 434.25 was the loss of sn2	

RCOOH group shown. 281.25 is reflecting of C18:1 shown in yellow, 255.25 is C16:0 shown in blue and finally 253.2 is reflecting of C16:1 similarly shown in blue but with one less H present. MS2 spectra obtained from peak 716.5 generated with corresponding peaks to those expected in B. C) MS-2 fragmentation pattern anticipated for  $m/z$  716.5, 452.5. E) MS-3 fragmentation seen for peak 716.5 > 452.5..... 305

Figure 5-7- Fragmentation pattern of ion 726.5 A) Shows the full spectra with  $m/z$ / peak, 726.5, of interest highlighted. B) Shows MS2 fragmentation pattern and structural details seen 466.3 seen in green with the loss of sn2 acyl chain, shown in red is  $m/z$  448.3 and 277 due to loss of sn2 RCOOH and sn2 RCOO ion respectively. C) Shows the MS3 fragmentation structure and spectra obtained from  $m/z$  726.5>466.5 generating identifying fragmentation patterns seen at and 140 loss of ethanolamine phosphate, 196 and 270 and 153 which shows the fragmentation product Glycerol-3-phosphate ion with loss of water. .... 307

Figure 5-8- A) Shows the fragmentation pattern obtained from the peak at 726.5 with the PE-P(18:1/18:1) plasmalogen fragmentation pattern shown structurally with peaks seen at 476 and 464. This is further evidenced in the B) MS-3 fragmentation through its presence of ions at 403.3 (further fragmentation from the phosphate group), 267 (loss of the SN chain at the oxygen ion) 196 and 140 (indicative of fragmentation at sn-1 1-o-alk-1'-enyl residue as an alcohol). 464 & 446 loss of 18:1 group as acid and ketene respectively. .... 309

Figure 5-9- Seeking proof of plasmalogen presence using peak 750.5. A) Shows the peak at 749.5, which generates fragmentation patterns of 534, 479, 282 and 253. This peak has an isotope that is also present at 750.5. B) Shows the full MS-2 fragmentation pattern generated at peak 750.5. The blue circle highlights a principle plasmalogen related fragment ion 464, the yellow circle highlights the isotopic peak generated from the previously discussed 749.5 with principal fragment ion 535 generated. C) Shows another example whereby the ratio of the two fragment ions has changed with the 749.5 isotope now being the more prominent. D) Illustrates the correct ms-3 fragmentation of the 750.5/464.5 plasmalogen lipid present confirming the presence of plasmalogen PE-P 18:0/18:1. This fragmentation patter is completely unlike that of the isotopes. .... 314

Figure 5-10- MS-2 spectra generated for ion peak 750.5 with focus on the region between 450 and 470 for samples A) Normal drosophila brain extract showing the presence of plasmalogen through the presence of fragment ion 464.5 B) Mutant extract AGPS C) GNAP D) FAR and E) PEDS1 with all mutant strains (B-E) showing complete loss of plasmalogen with no detectable 464.5 fragment ion. .... 315

Figure 5-11- example lipidomic chromatogram obtained from A) healthy control, B) FAR1, C) AGPS, D) TMEM189 and E) GNAP mutant drosophila heads over the region of  $m/z$  700-800 focusing on the region which is indicative of the presence of PE, PE-O and PE-P EtnGPL. .... 317

Figure 5-15- heatmap showing the identified lipids relative abundance changes between each of the different mutant strains samples and overall how these then

relate to each other in terms of pattern similarity. A) demonstrates the individual samples alterations in lipidomic profiles whereas B) shows the groups average changes. All of the groups, with exception of one replicate in AGPS mutation are seen to hold very similar profile changes, noted by there close proximity to each other. The TMEM-189 (pink) and the control group (green) had the most similar profiles to each other coming from the same primary branch, whereas the other three mutant strains differed more from the control group with FAR-1 (dark blue) and AGPS (red) baring more similarity than GNAP (turquoise). ..... 318

Figure 5-13- PCA analysis of all groups considering all three lipidomic subclasses identified in the study. FAR-1 showed the greatest variation to the control group, followed by TMEM-189, GNAP and AGPS. .... 320

Figure 5-17 A) Shows the variation of phosphatidylethanolamine (PE) and alkyl-phosphatidyl-ethanolamines (PE-O) species in individual mutant samples B) shows the mutant group average associations. Looking at the variations in PE and PE-O TMEM-189 holds the most similarity to the control group, illustrated through these both coming off the same major branch. There is a good differentiation of these groups through PE/PE-O profiles as all samples sub-branched in their group clusters. GNAP and AGPS showed the greatest similarity in PE/PE-O profiles of the remaining mutant strands. .... 321

Figure 5-15- Average total abundance of each different class of lipids identified with standard error shown for n=3 samples/ mutant group. .... 322

Figure 5-16- Graphical representation of the number of times each different length of carbon chain with specific level of saturation was seen in samples ..... 323

Figure 5-17- graphical representation to show the most common degree of saturation identified in all lipid species under study, with 3 double bonds being the most prevalent. .... 324

Figure 5-18- PCA analysis showing the differentiation of the mutant/ control strains from each other. A) shows the greatest distance achieved between Control and FAR-1 which was 57% accountable of PC1 and 39% PC2. B) Shows all of the groups separation using PCA analysis with PC1 accounting for 52% of the variation and PC2 25.5%. The confidence intervals around the groups is plotted with the ovals. .... 325

Figure 5-19- Fluctuations shown between different plasmenylethanolamines 's with an overall trend of upregulation seen in mutant strains. T-test variance for significant changes were made between each mutant group and the control groups values with bars noted with \* > 0.05 \*\* > 0.01 \*\*\* > 0.001 ..... 326

Figure 5-20- The relative alterations in plasmalogen abundance for PE(O-16:0/18:3), PE(O-16:0/18:2), PE(O-18:0/18:3) and PE(O-18:0/18:2) plasmanyglycerophospholipids. Significant variance between each of the mutant groups and the control group is denoted with p values as per \* > 0.05 \*\* > 0.01 \*\*\* > 0.001 ..... 327

Figure 5-21- LPE(P-16:0) was shown to decrease in all mutant strains with a significant decrease due to total absence reported in FAR-1 and GNAP. Significant variance between each of the mutant groups and the control group is denoted with p values as per \* > 0.05 \*\* > 0.01 \*\*\* > 0.001. .... 328

Figure 5-22- Total relative abundance of the sum of all Phosphatidylserine (PS) and Phosphatidylinositol (PI) classes of lipids observed in samples with only GNAP showing a significant variance with p.value of 0.04 for both PS and PI presence when compared to the control group. Significant variance between each of the mutant groups and the control group is denoted with p values as per * $>0.05$ ** $>0.01$ *** $> 0.001$ .....	329
Figure 5-23- Phosphatidylserine (PS) group variance for control and each mutant strain observed. Significant variance between each of the mutant groups and the control group is denoted with p values as per * $>0.05$ ** $>0.01$ *** $> 0.001$ .....	330
Figure 5-24- Subclass phosphatidylinositol (PI)'s relative abundance changes between control and mutant strains. Significant variance between each of the mutant groups and the control group is denoted with p values as per * $>0.05$ ** $>0.01$ *** $> 0.001$ .....	330
Figure 5-25- heatmap showing the identified lipids relative abundance changes between each of the different mutant strains samples and overall how these then relate to each other in terms of pattern similarity. A) demonstrates the individual samples alterations in lipidomic profiles whereas B) shows the groups average changes. All of the groups, with exception of one replicate in AGPS mutation are seen to hold very similar profile changes, noted by there close proximity to each other. The TMEM-189 (pink) and the control group (green) had the most similar profiles to each other coming from the same primary branch, whereas the other three mutant strains differed more from the control group with FAR-1 (dark blue) and AGPS (red) bearing more similarity than GNAP (turquoise).....	331
Figure 5-26 A) Shows the variation of PE and PE-O species in individual mutant samples B) shows the mutant group average associations. Looking at the variations in PE and PE-O TMEM-189 holds the most similarity to the control group, illustrated through these both coming off the same major branch. There is a good differentiation of these groups through PE/PE-O profiles as all samples sub-branched in their group clusters. GNAP and AGPS showed the greatest similarity in PE/PE-O profiles of the remaining mutant strands .....	331
Figure 5-27- PCA analysis showing the differentiation of the mutant/ control strains from each other. A) shows the greatest distance achieved between Control and FAR-1 which was 57% accountable of PC1 and 39% PC2. B) Shows all of the groups separation using PCA analysis with PC1 accounting for 52% of the variation and PC2 25.5%. The confidence intervals around the groups is plotted with the ovals. ....	331
Figure 5-28- heatmap illustrating the fluctuations of average total phosphatidylethanolamine (PE), alkyl-phosphatidyl-ethanolamines (PE-O) and ILPE present in each of the mutant and control sample extracts. ....	331
Figure 5-29- Average total abundance of phosphatidyl-ethanolamines (PE-O) and phosphatidylethanolamine (PE) in each group under study with significant variation from the control group indicated with asterisk whereby * $<0.05$ , ** $< 0.01$ & *** $<0.001$ .....	331

Figure 5-30- Average ratio of phosphatidylethanolamine (PE): phosphatidyl-ethanolamines (PE-O) in each group under study with significant differences between the mutant strain and control group shown with asterisk * $<0.05$ , ** $<0.01$ and *** $<0.001$ .....	332
Figure 6-1-- Structure of taurine .....	348
Figure 6-2- Synthesis of taurine from cysteine .....	349
Figure 6-2- Synthesis of taurine from cysteine .....	350
Figure 6-2- Synthesis of taurine from cysteine .....	350
Figure 6-3- Mode of action of the taurine channels to enable to the uptake of taurine from the plasma into cellular space. ....	351
Figure 6-6- Functions of taurine in the body (De Luca, A., et al., 2015) .....	352
Figure 6-7- The decrease in taurine absorbance after derivatisation over time due to breakdown of fluorescence product illustrating the time critical element of the study between measurements and derivitisation .....	359
Figure 6-8- The change in taurine absorbance measured with changing storage temperatures .....	360
Figure 6-9- Detection of taurine after reconstitution in different buffered solutions including 0.1% formic acid, 0.025M (pH 6) sodium acetate and sodium borate (0.4M, pH 9.5) .....	360
Figure 6-10- overlayed spectra from individually run amino acids showing the different elution times achieved on a c18 column. These included aspartic acid, glutamic acid, valine, asparagine, serine, arginine, histidine, glutamine, glycine, threonine, tyrosine and taurine. Cytidine, proline, methionine, phenyl alanine did not generate peaks. ....	361
Figure 6-11- Standard curve showing the change in fluorescence absorbance of taurine with changes in taurine concentration in $\mu\text{M}$ showing the standard deviation for each data point and a $R^2$ value of 0.9972 .....	362
Figure 6-12- Illustration of the sample's concentration range vs the dynamic range of the standard curve generated with all samples falling within the working range of the graph. ....	363
Figure 6-13- Plasma changes in taurine levels before and after exercise between candidates that took a placebo/ taurine supplement ahead of exercise.....	364
Figure 6-14- Average taurine concentration seen pre and post exercise in placebo and taurine administered participants with standard deviation bars plotted. ....	365
Figure 6-15- Percentage change of each participant's post exercise blood taurine content when compared to the pre-exercise taurine level. ....	366
Figure 6-16- A) Whole body sweat rate (WBSL) and B) local body sweat rate (LSR) observed on participants under fixed and ramped conditions with significant variance between taurine/placebo supplementation noted with asterisks in each group, fixed or ramps, whereby * $<0.05$ , ** $<0.01$ & *** $<0.001$ . Data generated by Jenny Peel.....	367

Figure 6-17- Sweat gland activation shown for fixed and ramped intensity levels with different measures made based on iodine paper size. Data generated by Jenny Peel.....	368
Figure 6-18- A) The gradual changes in core temperature observed over the course of the experiment including during fixed intensity (FI) at different time points and ramped activity b) Shows the average temperature change of each exercise type for both treatment groups. Data generated by Jenny Peel. ....	369
Figure 6-19- Skin temperature change observed during fixed intensity workouts with and without taurine supplementation. Data generated by Jenny Peel.....	370
Figure 6-20- Average perceived thermal comfort reported by candidates during fixed Intensity (A) and Ramped Intensity (B) exercises with and without taurine supplementation. Data generated by Jenny Peel. ....	370
Figure 6-21- Average thermal sensation perceived during A) Fixed Intensity and B) Ramped intensity periods of exercise with and without taurine supplementation. Data generated by Jenny Peel.....	371
Figure 6-22- Average rating of perceived exertion observed during fixed (A) and ramped (B) exercises for placebo and taurine supplemented candidate. Data generated by Jenny Peel. ....	372

## List of Tables and illustrations

Table 2.2-1- Cell line specifications.....	60
Table 2-2.2- Amino acids analysed in experiment with chromatogram ion selected and retention time .....	63
Table 2-3- Details and patient and healthy urine donors .....	66
Table 2-4- The most diagnostically successful metabolites, ranked from highest to lowest intensity, identified between A549 and Beas-2b cell lines identified using GCMS analysis. It should be noted that an $\infty$ symbol is utilised to represent where the value cannot be computed. In this instance LHR+ is calculated by taking the Sensitivity/ 1- Specificity, however in this instance the 1-specificity is zero. “Excellent” markers are shown in green in the table and “good” .....	72
Table 2-5- The most diagnostically successful metabolites identified between HDPE and Panc-1 cell lines identified using GCMS analysis. It should be noted that an $\infty$ symbol is utilised to represent where the value cannot be computed. In this instance LHR+ is calculated by taking the Sensitivity/ 1- Specificity, however in this instance the 1-specificity is zero. “Excellent” markers are shown in green in the table and “good” and “poor” in red.....	74
Table 2-6-The most diagnostically successful metabolites identified between HDPE and MIA PaCa-2 cell lines identified using GCMS analysis. It should be noted that an $\infty$ symbol is utilised to represent where the value cannot be computed. In this instance LHR+ is calculated by taking the Sensitivity/ 1- Specificity, however in this instance the 1-specificity is zero. “Excellent” markers are shown in green in the table and “good” and “poor” in red.....	76
Table 2-7-Table showing the average percentage CV, $R^2$ , Linear equation and dynamic range of each amino acid utilised in this analysis .....	90
Table 2-8-ANOVA results for significantly different amino acid markers with corresponding f. value, p. value, False Discovery rate (FDR) and Post-hoc analysis conducted using Fisher’s LSD. ....	91
Table 2-9- Modified nucleosides of interest identified in urine samples using HPLC analysis; with their corresponding retention times, compound $m/z$ anticipated in positive ionisation mode, fragment ion of interest, in the case of the modified nucleosides this is indicative of a loss of 132, nucleoside identity, formula, Mr and structure with fragmentation indicated in red and blue detail. ....	116
Table 2-10- The diagnostic properties of the combined nucleosides and combined guanosine modifications under study during the urine analysis used to differentiate PDAC from all other groups under analysis. ....	128
Table 3-1 Major ions consistently noted in MALDI spectra and used for manual calibration of spectra .....	193
Table 3-2- Most significantly different and diagnostically capable metabolites identified between non-pregnant females and males ranked in order of significance. Excellent scorings are shown in green, good in yellow and poor in red. In some instances, i.e. heptanedioic acid, the same identification has been generated	



multiple times being the identity assigned to multiple different peaks. This can occur for numerous reasons relating to the limitations of GCMS tentatively assigned identifications. Firstly, they can be associated with different isomers of the same amino acids, or they can in fact represent different metabolites but have similar identifications which couldn't be distinguished. ....	200
Table 3-3 Most significantly different and diagnostically capable metabolites identified between non-pregnant and pregnant females. "Excellent" markers are shown in green in the table and "good" and "poor" in red.....	204
Table 3-4- The 5 most diagnostically successful metabolites identified between pregnant and cord serum samples. It should be noted that an $\infty$ symbol is utilised to represent where the value cannot be computed. In this instance LHR+ is calculated by taking the Sensitivity/ 1- Specificity, however in this instance the 1-specificity is zero. "Excellent" markers are shown in green in the table and "good" and "poor" in red. ....	205
Table 3-5- Acronyms used to denote different lipidomic species during MALDI analysis.....	209
Table 3-6- Diagnostic qualities reported for the significantly different metabolites distinguishing males from non-pregnant female cohorts rated as excellent for all diagnostic qualities. "Excellent" markers are shown in green in the table and "good" and "poor" in red. ....	215
Table 3-7- The diagnostic values associated with the significantly different metabolic markers that are capable of distinguishing pregnant females serum and cord serum from each other. "Excellent" markers are shown in green in the table and "good" and "poor" in red. ....	218
Table 3-8 The most significant differences diagnostically capable of distinguishing non-pregnant females from male serum cohorts using MALDI in positive mode with corresponding AUC, LHR- and LHR+ scores represented in yellow for good and green for excellent with regards to diagnostic potential.....	223
Table 3-9 The most significant differences diagnostically capable of distinguishing pregnant females from cord serum cohorts using MALDI in positive mode with corresponding AUC, LHR- and LHR+ scores represented in yellow for good and green for excellent with regards to diagnostic potential.....	225
Table 4-1- A summary of the metabolites that distinguish different groups of elephants and their comparative levels of distinction ordered by significance. Groups were N= non cycling, C= cycling, P= pregnant. n.c.= likelihood ratios that could not be determined as the metabolite level was never elevated in the group exhibiting a decreased level of the metabolite. ....	276
Table 5-1- Lipids Identified and used during the study with <i>m/z</i> , subclass, assignment and fragmentation pattern seen listed .....	310

## Definitions and Abbreviations

GCMS	Gas chromatography – mass spectrometry
LCMS	Liquid chromatography – mass spectrometry
NMR	Nuclear magnetic resonance
MALDI	Matrix assisted laser desorption
TOF	Time of flight
CP	Chronic Pancreatitis
PDAC	Pancreatic ductal adenocarcinoma
FBS	Fetal Bovine serum
DCM	Dichloromethane
MeOH	Methanol
ACN	Acetonitrile
WBSL	Whole body sweat loss (weight change)
LSR	Local sweat rate = pre/post patch mass change/ (size patch * duration worn)
SGA	Sweat gland activation
OPA	o-phthalaldehyde
RPE	rating of perceiving exertion
VO <sub>2</sub> max	peak oxygen consumption- Criteria for achieving V'O <sub>2</sub> peak was: 1) reaching volitional exhaustion; 2) respiratory exchange ratio (RER) > 1.15; 3) final HR within 10 beats/min of age-predicted maximum; and 4) rating of perceiving exertion (RPE) > 19 (6–20 Borg scale)
TC	Perceived thermal comfort
TS	Skin surface temperature
RPE	Rating of perceiving exertion

## Chapter 1 -Introduction

### 1.1 Metabolomics

Metabolomics involves the analysis of the global metabolite presence within or excreted from an entity at a given point in time, taking a snapshot of the body's status at a singular moment. Quantitative evaluation of the metabolites present within this snapshot can illustrate alterations that can provide further indication of an upstream healthy or aberrant pathway. In the field of disease research these observed imbalances can provide indicative signatures of diseases in the way of a singular metabolite or a combination of metabolites.

The uses of these metabolic observations can be fundamental at different stages of disease prevention and can be used in:

- 1) Screening- Seeking metabolic alterations that occur before the disease symptoms present.
- 2) Diagnostic- when a disease has manifested and there should be a clear difference in biomarkers
- 3) Prognostic- good/poor response to treatment coordinated with biomarker patterns
- 4) Disease progression and response to medication- being able to correspond the return of metabolic profiles to healthy levels after administration of treatment

Metabolic disease signatures can prove useful, not only diagnostically when a disease has manifested and the combination of symptoms and metabolic differences become clear indicators, but additionally as a screening tool through provision of signals prior to disease establishment. This is possible through the fact that metabolic alterations precede disease symptom presentation. In this instance the potential of metabolomics could prove vital to patient outcome for a range of complex and difficult to diagnose disorders, particularly those that have time critical treatment outcomes, such as pancreatic cancer which will be discussed later in **Chapter 2**.

Further to this, in some instances metabolomics can be useful in monitoring a disease' progression or its response to particular treatment. This can be achieved through the monitoring and comparison of metabolic presence and its restoration back to healthy levels and therefore bodily functioning. Another use of metabolomics in disease is the provision of prognostic capability. In such cases certain metabolites of interest can give an indication of how a patient will react to a particular treatment protocol whereby a particular set of metabolites may suggest a good/poor response to a particular treatment protocol.

Metabolomics aims to consider all biological pathways and systems making it a compatible tool for the analysis of the majority of diseases due to the fact that a disease manifestation causes and can be caused by alterations to metabolic processes and therefore impacting on the concentrations of metabolites being present within or excreted by the body. When conducting a metabolomics analysis on a biological system typically a large number of related metabolites are qualitatively and quantitatively observed and analysed commonly using mass spectrometry-based techniques. A comparison can then be drawn to determine any discriminating markers that can be found to significantly differentiate healthy from non-healthy candidates.

## 1.2 Biomarkers

Metabolomics indicates biomarkers that can discriminate between healthy and non-healthy cohorts individually or through panels of the most discriminating biomarkers. Biomarkers themselves include any measurable biological feature that can associate with the presence, prognosis or underpinnings of a disease; as such including any small molecule including metabolites. The most diagnostically valuable biomarkers are sought after during metabolomics evaluation and are selected on their ability to define an indicative signature of a disease. Often one biomarker alone will not have enough capability to be used as a sole diagnostic tool, but instead a panel of the most able biomarkers will be used instead.

Metabolomics is becoming one of the most popular tools utilised for novel disease combating strategies through biomarker identifications in medical ailments. This is

reflected in the numbers of published journals seen to be almost doubling yearly in popularity as depicted in Figure 1-1.

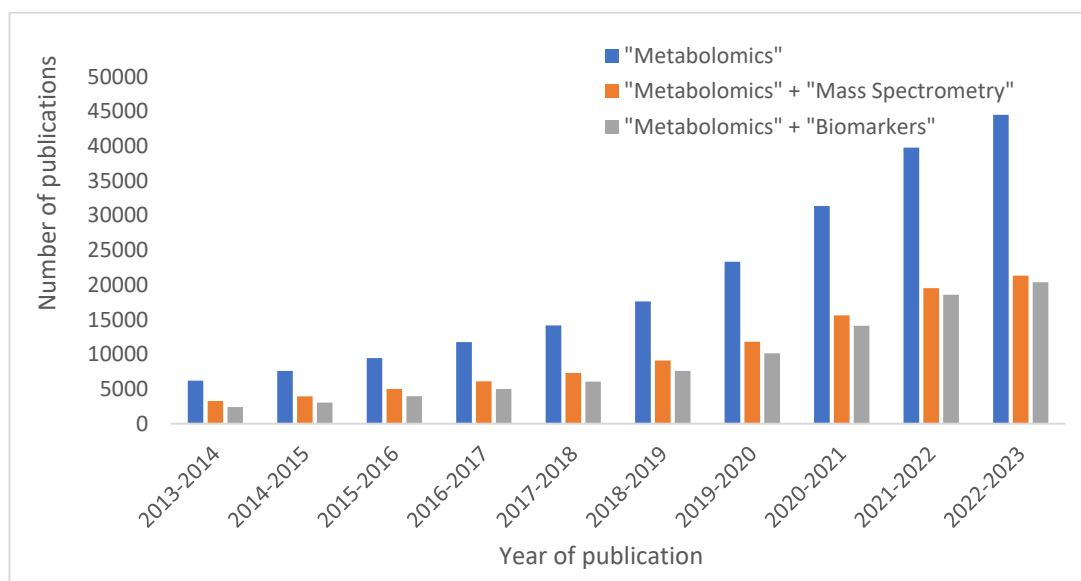


Figure 1-1 Graphical representation of the number of metabolomic journals published per year available to access on NCBI (<https://www.ncbi.nlm.nih.gov/>) using the search terms “Metabolomics”, “Metabolomics” + “Mass spectrometry” and “Metabolomics” + “Pancreatic cancer” correct as of 21/09/2023.

A large reason for this increase has been the great advancement in mass spectrometry and aligned techniques that can now offer faster and more detailed analysis of multiple different metabolites pathways in an automated and simplified protocol.

#### 1.2.1- Biomarker critique using ROC curves

ROC curves were generated and then utilised to critique which biomarkers to capture those that were the most diagnostically capable of distinguishing the two groups of interest from each other. To further critique the capability of the biomarkers identified the ROC curves were used for the determination of area under curve (AUC), specificity and sensitivity scores taking the values from the upper left most corner of the graph, . These in turn can then be used to calculate the positive likelihood ratio (LHR+) through taking the sensitivity value and dividing this by 1-specificity and negative likelihood ratio (LHR-), which was calculated by taking 1-sensitivity value and dividing this by the specificity. These, with the AUC, acted as a biomarker scoring

system, through which only metabolites deemed to have better than “good” diagnostic ratings were selected. These ratings have been previously defined by Ray P., et al., 2010, whereby “excellent” denotes a marker that has an area under the curve of 0.9, LHR+ greater than 10 and an LHR- less than 0.1 and “Good markers” 0.75-0.9, 5-10, 0.1-0.2 respectively. Metaboanalyst was additionally used to provide further visual depiction of the biomarkers capabilities for defining select groups through the generation of heat-maps, which illustrate the change in metabolite profiles between groups referring to their intensity changes, PCA plots that show the maximum variability between groups and PLSA plots in order to illustrate the component decomposition of the data sets and how they attribute to group alterations.

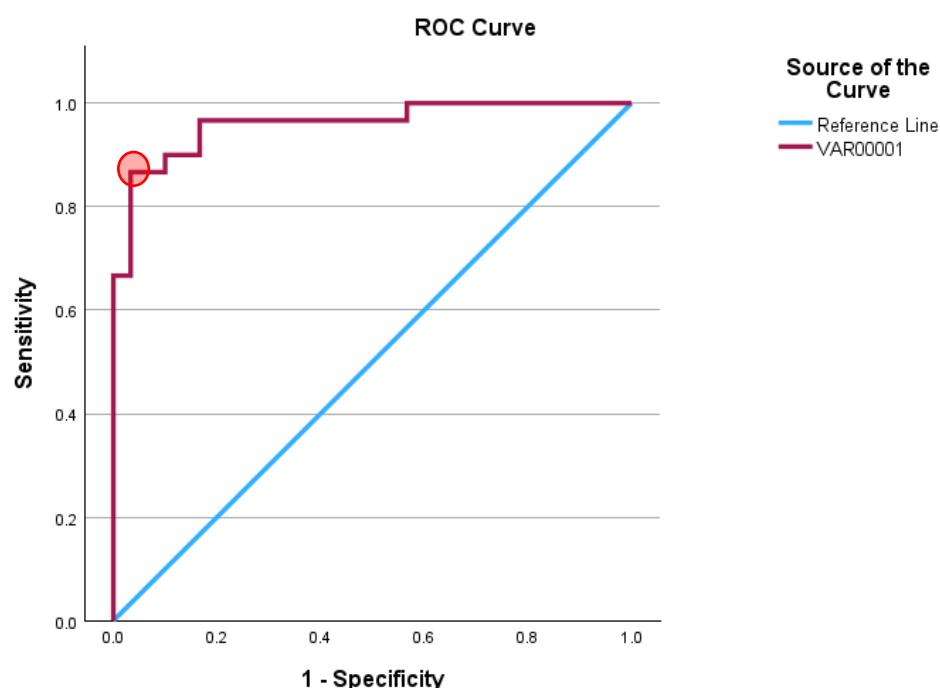


Figure 1-2- Exemplar ROC curve used to evaluate a biomarkers diagnostic capability in the analysis, which generated a Sensitivity value of 0.867 and 1-specificity value of 0.033 at the most efficacious point of the graph (circled) with an area under the curve scoring of 0.96.

### 1.3 Targeted vs Untargeted

Prior to conducting a metabolomics research evaluation, the amount of information about what the metabolites of interest are must be carefully considered and

evaluated to make sure that not only the correct sample source is chosen but also the most efficient extraction technique due to the large number of metabolites present in a living organism.

During instances when the metabolites of interest are known, the researcher will conduct a **targeted** metabolic study usually conducted in order to prove a pre-existing hypothesis. During such, the researcher will exploit the key properties of the metabolites of interest in order to tailor an extraction protocol that will give the purest and most conservative extraction of these metabolites. It will also enable absolute quantifications to be generated through use of known standards.

Conversely if the metabolites that may be of interest are not yet known, the researcher will want to analyse as many metabolites as possible and will conduct an **untargeted** approach. This strategy aims to conserve as many metabolites as possible using the most diverse and unbiased extraction process whilst assuming the poorest metabolite stability in an aim to conserve and detect the entire metabolome. The advantages of such a technique is the vast detection of metabolites and comparisons that may generate more novel and unanticipated metabolic fluctuations. However, this strategy may not preserve as many more sensitive metabolites and due to the sheer number conserved relative quantifications are used rather than absolute. This will also create more analytical challenges due to the vast number of metabolites that will be extracted during this form of analysis often therefore requiring a more intense data processing strategy to be implemented.

#### 1.4 Sample sources

The choice of biological sample must be carefully considered by the researcher in advance of the study being undertaken. In some instances the choice will be made solely on the necessity of a non-invasive and least stressful source of sample acquisition as possible. This was the case in Chapter 4 when analysing fertility markers of captive elephant's. This was conducted on faecal matter in order to avoid any additional undue stress to the animals that could firstly impact metabolic presentations as well as potentially linking to foetal complications. Another consideration is whether multiple samples may be required over a set timeframe as

well as the metabolites clearance rates from different sample sources. In such instances the ease and availability of the sample will need to be considered as well as the best source for detecting presence, for example taurine has quite a high clearance rate from the body and is best detected in the blood for a current picture of circulating levels impacting an exercise study. Biochemically considered however, if the metabolites of potential interest are known the best source of sampling can be chosen based on the metabolic pathway understanding and stability of the metabolite of interest in that sample. Another important consideration is the samples complexity which could mask small or dilute metabolite markers adding also to the amount of processing that will be required for analysis to then be made.

When deciding on the sample source being used for a metabolomics analysis a variety of different factors must be considered including what metabolites are wished to be considered (and if in fact this is known), what is clinically available and with this the ease of sample acquisition, what will provide the greatest source of information, sample stability and sample normalisation strategy and additionally the cost time and resources needed to process the different samples.

As this thesis aims to find new explorative biomarkers a variety are used to exploit as many different metabolic pathways as possible in the least invasive ways to attempt to provide new markers that are clear, cost effective and give a relatively fast analysis strategy.

The samples chosen to be analysed within this thesis include cell culture, urine, plasma, serum, and faecal matter as well as the heads of drosophila flies as a sample representative predominantly of neurological cells.

### *Cell Culture*

Cell culture work is a popular tool used for seeking novel biomarkers for numerous systemic conditions. Specific Primary cell lines are widely available from suppliers, such as ATCC, enabling the simplistic acquisition of specific and multiple cell types required. These cells are grown *in vitro*, artificially controlled and modified



environments with the aim of mirroring the natural environments as closely as possible to ensure a close to true example. This allows for the generation of a large numbers of biological replicates in an easy and cost-efficient way. Whilst cell cultures provide many advantages including ease of acquisition, simplified normalisation, manipulation and near perfect external controls enabling good subject replication there can be disadvantages to their use in metabolomics. For instance, whilst such culturing systems will provide a good initial model close to that of a biological system, they cannot give an exact biological example and so further clinical sample validations are likely to be required. Additionally, there can be detection related issues relating to mass spectrometry due to a large abundance of background signals caused by excessive media supplements. This can lead to suppressed target metabolite signals, especially those in particularly low-level concentrations. This can be corrected for an amount through prior cell washing before extraction in order to remove the external media contents and secreted metabolites as much as possible prior to analysis. However, this comes at the detriment of increased leakage and therefore weaker metabolite measurements as well as the possibility of ion suppression most often found in cases where PBS was utilised for washing (Ser, Z., et al., 2015). For adherent cell lines, such as those used in this study, there is also an additional complexity of metabolic alterations resultant of cellular detachment steps prior to extraction. Cellular detachment is typically performed using trypsin digestion, which whilst maintaining the integrity of the cells, alters the cellular environment prior to extraction and therefore impacting the metabolic profile. This issue can be avoided in part by combining the cellular release, quenching and extraction protocols together where possible, such has been done in Chapter 2. Quenching aims to halt the metabolism as much as possible to preserve the metabolic composition at the exact moment, unhindered by extraction processes. This step is also important for the preservation of metabolites that may additionally otherwise be susceptible to degradation either through stability or enzymatic presence, often whilst preserving cellular integrity. Quenching can be performed using multiple different methods, this study has utilised ice-cold 60% methanol addition, with the cold reducing enzymatic activities. To then avoid metabolite leakage, this study has chosen to scrape the cells

directly into this methanol solution and further utilise the methanol for extraction and thus avoiding metabolite loss that could occur using trypsinisation etc (Kapoor et al., 2017).

### *Blood, Serum and Plasma*

Blood based samples are another popular analyte choice in metabolomic investigations mainly through its presence throughout the body, being influenced by multiple metabolic processes, whilst providing exact status representation and current metabolite presence rather than being a downstream metabolic waste-product illustrating metabolite disease consequences. This is through its role in the provision of material for cell utilisation, cell waste removal and provide factors to help cellular homeostasis maintenance. This enables blood based analysis to hold a potentially large number of metabolic changes influential in disease pathogenesis, homeostatic regulatory factors are more of an exclusive presentation in blood as some of these biomarkers may be eliminated when using urinary analysis as an alternative.

In addition, blood is plentiful and relatively simple to acquire whilst being minimally invasive, opening the option to enable multiple samples to be collected in time-critical or dependent analysis. The metabolite concentration of blood-based products is also very high, and is already in solution form with lower enzymatic presences, limiting the degradation issues already discussed in cells and often require far less complex extraction methodologies.

Blood is already used for these reasons in a large number of different clinical investigations and the usefulness for each is highly dependent on the form of blood that has been extracted, as each different form hold very metabolic profiles. These forms include whole blood, serum and plasma.

Serum is acquired from coagulated blood, losing the large proportion of masking fibrin clots and coagulating factors by leaving the whole blood at room temperature for a duration of time (typically 30-60mins; this in itself has been found to hold significant difference over the metabolic composition of the sample acquired due to

the coagulation cascade (Liu X et al., 2018; Nishiumi S et al., 2018a)) before this is separated using centrifugation techniques, Figure 1-3 (Yu Z., et al., 2011; Liu X et al., 2018). This centrifugation results in various proteins and metabolites being released into the serum, which can then be analysed. Prior to analysis protein depletion via methanol addition and centrifugation is conducted in order to remove any large masking proteins from the sample, most abundantly of which is serum albumin and IgG, which will otherwise conceal and dilute out the small and low abundant metabolite biomarker candidates. Coagulation itself involved the reduction of volume (and therefore concentration of metabolites) in blood through the formation of gel blood clots from the previous liquid state (Nishiumi S et al., 2018). After the depletion has been conducted the sample can then be further extracted or analysed.

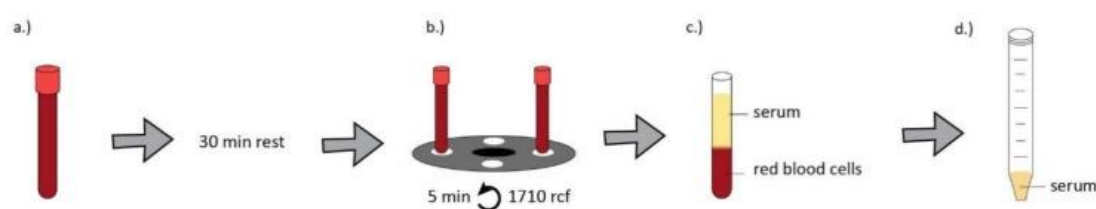


Figure 1-3- Schematic for the isolation of serum from whole blood involving its clotting for 30 minutes, followed by centrifugation at room temperature and supernatant collection (Kardos, D., et al., 2019)

Plasma alternatively is obtained through the addition of anticoagulation factors such as EDTA ensuring that the blood is not allowed to coagulate. The plasma will then be separated from the majority of the blood cells via centrifugation (a large number of platelets remain), which will typically be done as soon after collection as possible, if this cannot be done immediately the sample is kept cold again to prevent any coagulation, Figure 1-4 (Yu Z et al., 2011; Liu X et al., 2018). This step can hold critical implications on metabolic coverage and must be done quickly and consistently during metabolic evaluations to get a true metabolic representation of the sample (Nishiumi S et al., 2018).

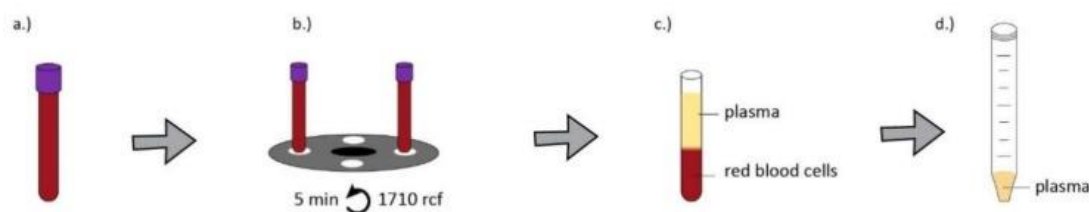


Figure 1-4- Schematic showing the acquisition of plasma from whole blood after the addition of EDTA the blood is allowed to clot for 30 minutes prior to centrifugation at room temperature and supernatant collection (Kardos, D., et al., 2019)

This difference in handling of the samples has reported vast differences on the metabolic profile obtained though both sources are known to provide good reproducibility (plasma more so due to less processing and the previously mentioned coagulation cascaded) Plasma extracted metabolites are also found to be generally more stable to freeze thaw cycling etc than those extracted from serum (Hirayama A., et al., 2015). However, some studies (Nishiumi S et al. 2018) suggest that this clotting process conducted in serum preparation can also cause the loss of some potentially important metabolic biomarkers due to enzymatic breakdown. Both plasma and serum have very good metabolite coverage, with serum holding slightly more metabolites than plasma. In particular serum is believed to hold higher levels of phospholipids which is thought to be due to the platelets releasing them during the coagulation of the whole blood (Nishiumi S et al. 2018).

Extraction methods for both serum and plasma can additionally hold vast differences over the types and range of metabolites that can be seen within the sample. Studies tend to agree that generally the best overall extraction protocol can be achieved using ice cold methanol (cold to stop enzymatic activity via quenching) in a ratio of 1:3 or 1:4 plasma:solvent with the extraction being conducted at 4°C environment. When compared to a number of different conventional solvent based extractions (Methanol: Ethanol [1:1], Methanol-Methyl tert-butyl ether, Liquid-liquid equilibria of Methyl tert-butyl ether and other Solid phase extraction based extraction methodologies (C18, IEX and PEP2) it is found to significantly surpass not only the global coverage of metabolite classes that are extracted but has also been found to support the most reproducible, far succeeding chloroform based extractions due to

the fluctuations that can arise during phase separations (Alshammari TM et al., 2015; Sitnikov DG et al., 2016). The other methods of extraction can provide more specific protein recoveries, such as SPE columns and so these are still employed for **targeted** metabolomics approaches. When methanol-based extraction has been conducted, the vast majority of metabolites of interest are organic solvent soluble and so will need no/little further processing before being run on an MS system of interest. Additionally the majority of different classes of metabolites are found to be stable in this way at 4°C for several months; unless GCMS analysis is wanted in which case sample drying and derivatisation is typically all that is required (Alshammari TM et al., 2015).

### *Urine*

Urine is a commonly utilised biomarker analyte source being readily available and one of the most simplistic and least invasive to collect. The collections themselves are less costly and traumatic to patients as well as providing a vast sample volume for analysis with multiple sample collections being possible. Urine analysis currently utilised in such a way being the source for pregnancy testing, various nephrological disorders (Barratt J., and Topham P., 2007) as well as showing promise in candidate biomarker studies for different cancers (Njoku, K., et al., 2020; Jordeans, S., et al., 2023).

The metabolic composition of urine largely consists of excretion product from glomerular filtration of plasma, excretion of urinary tract epithelial cells, waste epithelial cells and urinary exosomes (Thomas, C. E., et al., 2010) with the latter examples enabling the direct analysis of nephrological dysregulations. The blood waste elements mean that a vast array of water-soluble metabolites, including genetical products such as nucleotides, proteins and peptides, cells and amino acids at their most processed and unregulated stage are captured in urine (Njoku, K., et al., 2020). The analysis of these downstream metabolic products enables the evaluation of potential biomarkers that may be missed from blood evaluations due to homeostatic regulatory mechanisms that remove excessive metabolic products. This is particularly important for catching diseases during early stages of, at a point where metabolic changes may be more minimal and therefore easier to control/regulate.

An issue however with this lack of homeostatic regulation in urine, lies with it having a large variation between samples even if acquired from the same candidate on the same day. Such variations include volumes, concentrations, pH, age, diet and various lifestyle choices. A potential control over this issue is to try and collect the sample in the first instance of the subjects day (it has been proven to have the least variability in protein concentrations; however this can be more prone to loss of proteins through proteolysis degradation even with chemical additions such as boric acid). An alternative solution is to conduct 24hours collections (this can be more troublesome to patients and more time consuming however and can be more prone to bacterial contaminations but avoids the need for normalisation with internal metabolites which in some disease will not be possible as the metabolites used may fluctuate in their concentrations due to the disease manifestation (Bujak R., et al., 2015). The point of urine collection is another factor to control as well regarding whether the whole collections or more commonly mid-stream sampling is done with the aim of achieving the “cleanest” possible sample avoiding bacterial contaminations and normalising proteins to known metabolites within the samples such as creatinine, Albumin and Immunoglobulin G there will still however be flux between samples (Thomas C.E. et al., 2010; Barrat J and Topham P, 2007). The storage of unprocessed samples should be at 4°C then upon processing freezing and not thawed until ready for analysis to avoid degradation of metabolites Thomas C.E. et al., 2010) in order to most effectively retain the metabolites composition.

Urine can be a good source of protein biomarkers, when compared to blood there are far lower concentration of large masking proteins which means clearer more metabolite rich samples are achieved with less processing required alleviating difficulty and potential metabolite loss through protein binding, processing and enzymatic degradation in the time. Additional to this, due to the simplicity of the sample, often the urine can be directly applied to LC based systems with no or very little pre-processing required at all (Gowda GA., et al., 2008). Urea is however a large component of urine and can mask similarly structured metabolites. Salt can be another complexing component of urine in some metabolomics evaluations this will need to be removed in order to simplify the analysis of metabolites. This can be

achieved using dialysis or ultrafiltration, but as with other processing options care must be taken to avoid losing metabolites of interest during the processing (Thomas C.E. et al., 2010).

### *Faecal matter*

Faecal use in metabolomic evaluation is slowly gaining popularity, through its use in disease analysis can be dated far back historically (Zhgun, E. S., & Ilina, E. N. 2020), its full potential as a source of biomarker is only really beginning to be appreciated. Currently faecal samples are used to assay for a number of different conditions, mainly relating to intestinal or gastric ailments such as bowel cancer that is conducted by the NHS UK, 2019 as a home screening tool to seek the presence of blood. Similarly, to urine, faeces offers the waste products of the body as an insight into the goings on of the organism, but unlike urine, it additionally can give a view of the symbiosis with the gut flora which can be key to the presence of or development of a number of different disorders. Faeces can identify alterations and therefore links between disease and the internal metabolic pathways, the impact of diet as well as the interactions of this diet with the gut flora (Loftfield E et al., 2016).

Faeces is a far less “clean” source when compared to blood and even urine and as such the collection, pre-treatment and storage of the sample is far more complex and critical to the validity of the metabolite profile. A large reason for this is the far greater presence of bacterial source, of the 84 to 93% organic matter composition, 25-54% is bacterial biomass (Zhgun, E. S., & Ilina, E. N. 2020). The human microbiome comprises of bacterial, fungal, viral and protozoa based microorganisms and play a vital role in impacting the hosts health through metabolic processing. Alterations to the microbiome have been associated already with numerous different diseases including inflammatory bowel disease, psoriasis, arthritis and diabetes (Singh, R. K., et al., 2017). For this reason the microbiome is an equally important metabolic constituent to be considered when seeking novel biomarkers representative of disease.

As well as the amount of unique data that can be achieved through faecal sample analysis, and the ease of acquisition and availability of a large volume of sample, faecal samples have gained extreme popularity due to the non-invasive nature of this

acquisition. This is particularly important in a number of animal-based investigations whereby urine and blood samples would be too cumbersome to acquire or stressful to the animal. Additionally, when analysing certain metabolites, such as stress hormones, the stress caused to the animal due to sample acquisition could impact upon the data generated in the study and fluctuate the very metabolites that are being analysed.

A large disadvantage in the use of faecal matter for metabolomic evaluation is the vast variation in sample composition which is greatly impacted by diet and gut flora rather than intrinsic alterations, such as faster cell turn-around, representative of disease.

### 1.5 Mass Spectrometry and spectroscopy techniques for use as an analytical platform

Mass spectrometry has pathed the way for metabolomics research to advance into the current era through the provision of a means of analysing small molecular weight molecules in a fast, automated and high through-put manner. The measurements acquired can be used in both quantitative and qualitative manner enabling a broad range of analysis to be conducted.

Mass spectrometry enables the analysis of small complex biological molecules in a variety of different complex samples and biological sources making it an ideal tool for metabolomic analysis. This is aided through its ability to couple the MS to different separation based systems, in order to separate out metabolic presence and reduce the complexity of the sample under study. This also aids with sensitivities through preventing the occlusion by more abundant metabolites in sample mixtures, better enabling quantification. GC and LC are the most commonly utilised mass spectrometry coupled separation techniques and were both utilised in this study.

The most typically used mass spectrometry based metabolite fingerprinting techniques, and the ones that will be used and referred to within this thesis include GC-MS, LC-MS and MALDI. Often more than one of these techniques will be used for



any analysis as they offer different advantages and capabilities over each other which are better suited for different sample complexities.

#### 1.5.1- Gas chromatography coupled to mass spectrometry- GC-MS

GCMS is a very popular choice of analytical technique during metabolomic analysis owing largely to its applicability across a wide range of different sample source types and ability to detect a wide range of different markers making it suitable especially for non-targeted approaches.

GCMS samples first need to be extracted and then fully dried, this is typically dried under vacuum before being derivitised. This is due to the requirement for the samples to be thermally stable and volatile which can be encouraged, most commonly, through silylation. After derivitisation the samples are injected into the column where the basis of the separation is according to molecular vapour pressure, influenced through a gradual increase in temperature during a method. After separation, the metabolites enter the detection phase, typically via electron ionisation (EI). This form of ionisation also encourages fragmentation to occur enabling identifications to be generated based upon this fingerprint like pattern being compared to pre-existing libraries. There are different mass analysers that can be used for the detection including time of flight (TOF), quadrupoles and ion traps (IT). A quadrupole mass separation device was used during this thesis and involves four parallel rods that can separate a narrow band of  $m/z$  along the axis of the rods through varying the voltage being applied, Figure 1-5. This voltage variance with time enables a scan range of  $m/z$  and the generation of spectral detail (Pitt J. J. 2009). Quadrupoles offer a high sensitivity across a good dynamic range to be achieved but do have a lower mass accuracy and slower scan rates when compared to time of flight.

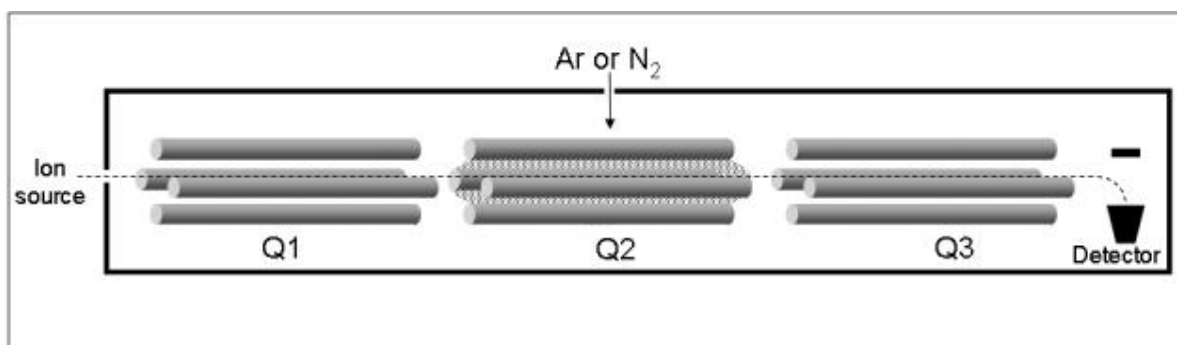


Figure 1-5-Illustration of a triple quadrupole, made up of three consecutive quadrupoles for improved mass filtering before detection (Pitt J. J. 2009)

Overall GCMS offers a high resolution, reproducibility of samples with metabolite identification capabilities and the ability to generate relative or absolute (used in conjunction with standard curves) quantifications. It is generally used for small analyte detections below 4kDa ranges.

#### 1.5.2- Liquid Chromatography mass spectrometry- LC-MS

For very similar reasoning to GC-MS methodologies LC-MS is a hugely popular choice of analytical platform for both targeted and untargeted metabolic studies. LC-MS, unlike GC-MS, does not require sample thermo-stability or volatility whilst still enabling the detection of a broad sample set with detection of a good metabolite range. There is also a very minimal amount of sample preparation required, typically only filtration will be needed, but this does require good solubility in addition to larger sample volumes. Separations in the case of LC-MS methodology are based upon interactions of the metabolites with the column and mobile phase selected, with greater column interactions resulting in greater retention times and conversely a better mobile phase interaction shorter time. Due to the ability to select specific mobile phases and columns, the separation achieved can be tailored to suit the metabolite presence desired's properties. A common form of separation, as used in this study, is reverse phase whereby a porous silica based stationary phase is bonded to hydrophobic hydrocarbons, with varying C8-18 carbons in length. These columns are used with a polar solvent system, leading to separation based on polarity. This is a good choice for proteins, peptides, acids and bases or inorganic ions. The detection method can be achieved using chromatographic detectors such as UV or

fluorescence, as was done in Chapter 6, or for a more sensitive and capable analysis it can be coupled to MS. For this to occur the column eluent is ionised, typically using electrospray ionisation (ESI), which enables the transfer of charged ions from liquid to gaseous phases. For this reason semi-polar and polar compounds are required for this analysis. The liquid sample is pumped through metal capillary (3-5kV) before being nebulised into a fine spray at the tip. This spray product is then evaporated via heating with charges then transferred to analytes. Aperture and focus voltages direct these charged species to the mass spectrometer via a high vacuum system with detections able to be made in positive or negative mode enabling a broad range of analyte detections to be made, Figure 1-6 (Pitt J. J., 2009). Chapter 2, modified nucleotide analysis utilised ESI in conjunction with an ion trap analysis system. This involves three hyperbolic electrodes to trap ions in a 3-dimensional space using different frequency voltages. Ions are then sequentially ejected from the trap according to  $m/z$  generating a spectrum. Alternatively, a specific ion can be isolated in the trap whilst others are ejected. The trapped ions can then be fragmented in order to generate  $MS^n$  details for an enhanced level of molecular detail and identification (Pitt J. J. 2009).

Whilst LC-MS systems provide a good level of sensitivity, quantitative and qualitative analysis the level of resolution is somewhat inferior to that achieved by GC-MS, and typically it is also subject to longer run times though this can be alleviated with UPLC based systems.

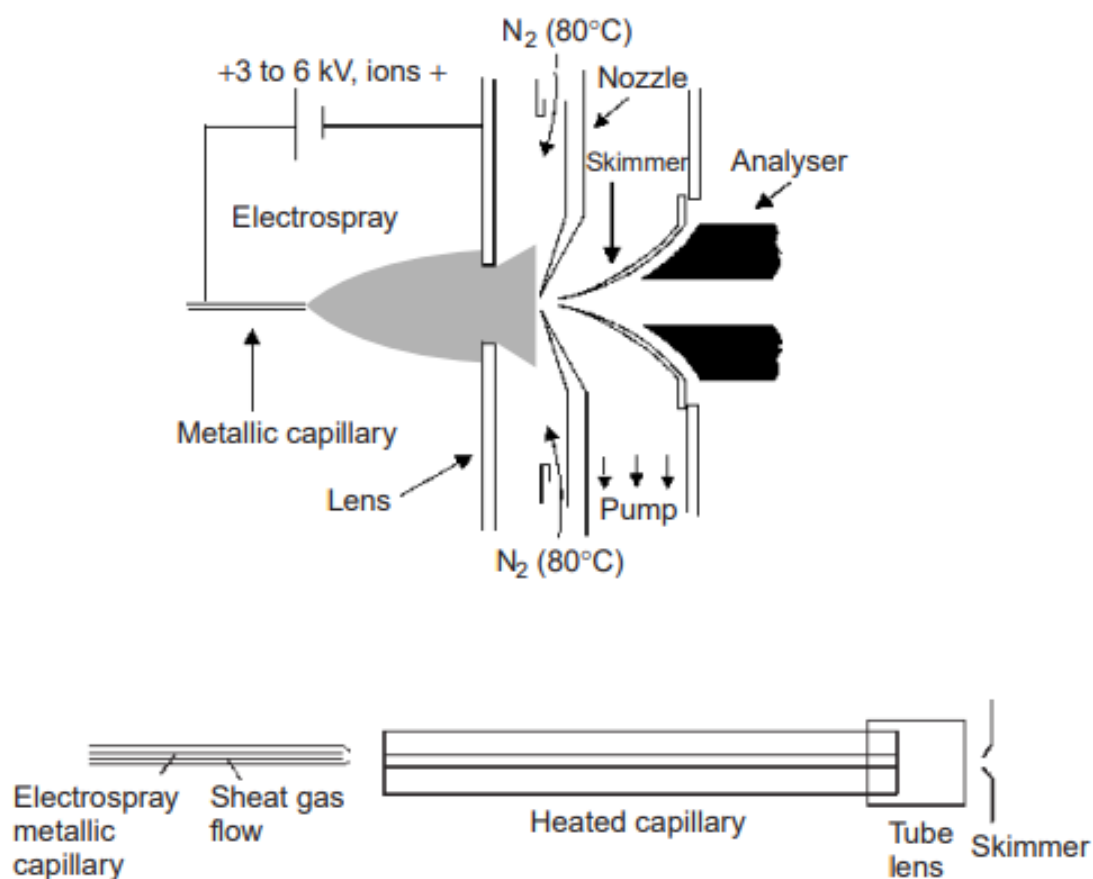


Figure 1-6 Illustration of the function of ESI whereby the sample is pumped through metal capillary (3-5kV) before being nebulised into a fine spray at the tip. Skimmers are used for ion focalisation and heat applied either from nitrogen source (top) or a heated capillary (bottom) (De Hoffman, E., Stroobant, V., 2007).

### 1.5.3- Matrix Assisted Laser Desorption/Ionization- MALDI

MALDI is a slightly less popular choice of MS methodology in the field of metabolomics but is gaining traction for the analysis of large, non-volatile and thermally labile metabolic candidates including proteins, oligonucleotides, peptides, phospholipids and high molecular weight large inorganic compounds metabolites being applicable to metabolites in the 1000's of Da range. MALDI systems require low amounts of sample presence with minimal extraction necessary. The sampling and analysis times achieved using MALDI are rapid and unlike other MS methodology have a high tolerance for salt presence in samples. This makes MALDI a very robust and beneficial methodology for complex sample forms.

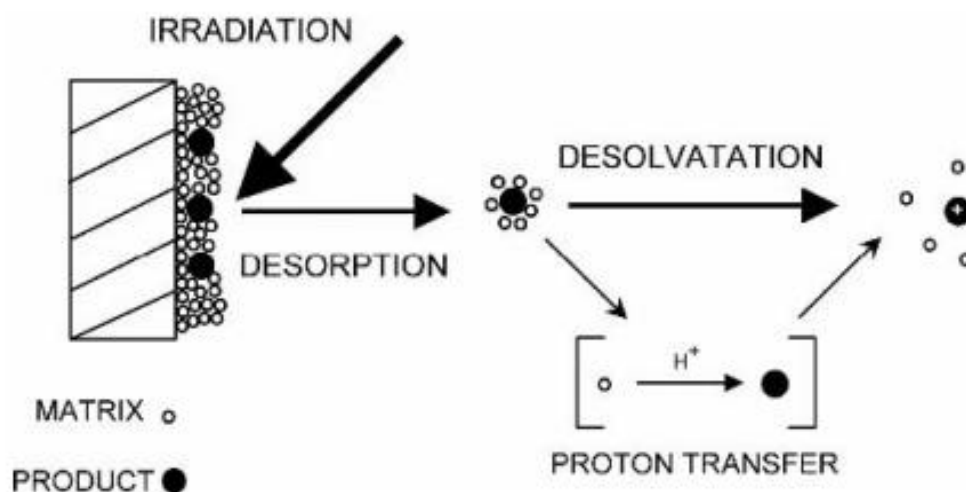


Figure 1-7- Schematic showing the principle of MALDI analysis whereby the sample mixed with matrix has been fired with a laser, resulting in energy transfer and a plume of matrix and sample to be released from the surface (De Hoffman, E., Stroobant, V., 2007)

MALDI involves the application of a specific matrix in excess to the surface or mixed with sample before being spotted and dried onto the MALDI plate. The sample spot is then fired with a laser causing it to absorb the energy, this rapidly heating the crystals, acting as an intermediate to help the ionisation of analytes resulting in a plume of expanding matrix plume and intact analyte, Figure 1-7 (De Hoffman, E., Stroobant, V., 2007). Time of flight (TOF) analysers is then used to separate the generated ions after acceleration by an electric field according to their velocities across the flight tube (De Hoffman, E., Stroobant, V., 2007). The use of a matrix gives MALDI a high sensitivity due to it limiting the damage of a sample during analysis.

The choice of matrix is a very important consideration during the analysis as it will impact the quality of the data generated and have different metabolic class detection capabilities. Another consideration with the matrix selection comes in the fact that the matrix itself can be detected. Usually, the matrix will be between the mass range of 500 and 600 Da, which can complicate an analysis, unless utilised on samples above this range. During this study 9AA (9-aminoacridine) was used for negative mode analysis for the study of free fatty acids and phospholipids, CHCA ( $\alpha$ -Cyano-4-hydroxycinnamic acid) was used for the study of fatty acids, glycerophospholipids and

peptides in positive mode and DHB (2,5-Dihydroxybenzoic acid) was primarily used for phospholipids also in positive mode.

The advantages of MALDI's small sample requirements, speed and sensitivity need to be balanced against its inferior reproducibility (this is due to matrix impacts) leading to quantitative difficulties when considering its utilisation in metabolomic analysis.

#### 1.5.4- Orbitrap

Orbitrap analysis is a newer and more powerful mass spectrometry technique providing a high resolving power, mass accuracy, resolving power as well as selectivity and sensitivity across a broad dynamic range. It works via direct infusion with no prior separation under a short total analysis time, though this will increase if quantitation is desired. The lack of separation means the processing requirement is a lot simpler as no retention time alignment are necessary. The linking with direct infusion also enables the use of shotgun technique for lipidomic analysis utilised in this thesis.

The trap is made up of three electrodes, Figure 1-8. The two outer electrodes are cup shaped and the inner spindle shaped, making this form of MS both an analyser and detector in one. A voltage is applied between the outer and central electrodes, which generates an electric field. The radial component will work to attract ions to the central electrode. Ions are injected between the central and outer electrodes through a deflector. Radial electric fields will bend the ions towards the central electrode and work in conjunction with the tangential velocity creating an opposing centrifugal force. The force is altered in order to keep the ions on a spiral inside the trap. This causes ions to oscillate around the central, but between the outer electrodes; with different ions oscillating at different frequencies, thus leading to the ions separation. The axial electric field pushes ions towards the widest part of the trap whereby the outer electrodes are used as a receiver plate for image current detection of these axial oscillations. Through measuring the oscillation frequencies induced by ions on outer electrodes the mass spectra of the ions are acquired using image current detection, the  $m/z$  is derived from the harmonic ion oscillations along the axis of the

electric field (Kienzl-Wagner, K., & Brandacher G., 2014; Zubarev, R. A., & Makarov A., 2013).

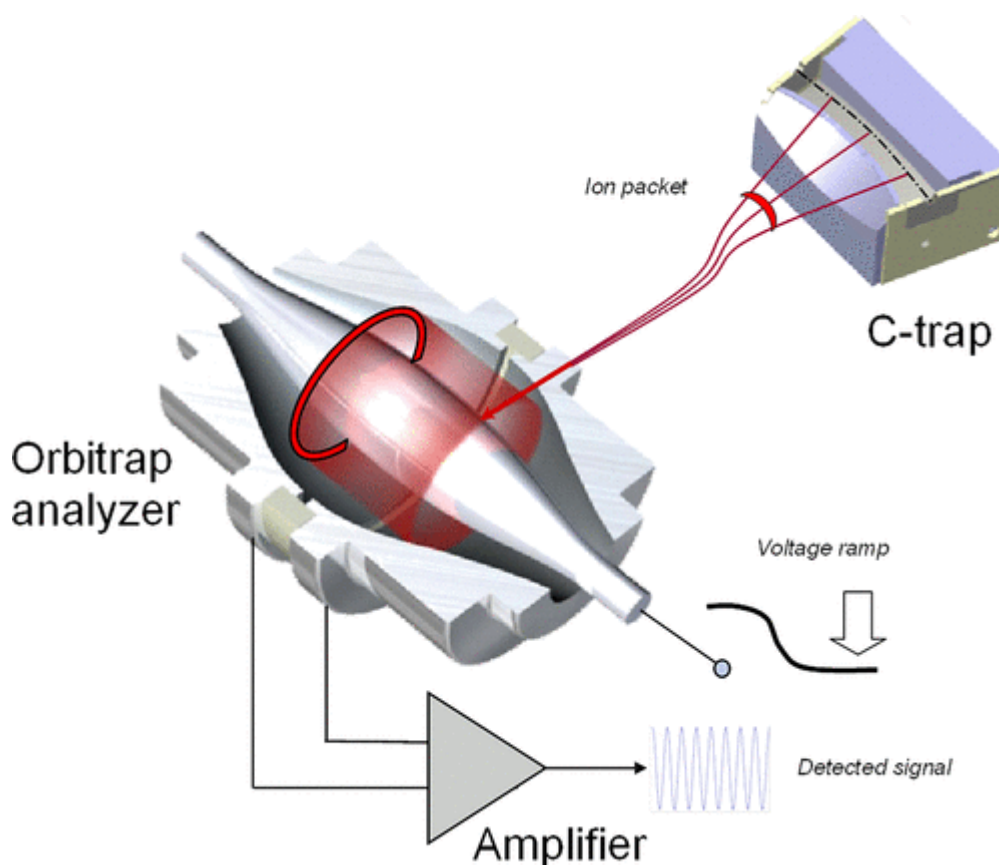


Figure 1-8-Illustration of an orbitrap showing the injection of ions into the orbitrap analyzer via the c-trap. They then oscillate around the central electrode and between the outer electrodes maintained by an altering electric field being applied. After separation ions are then amplified and detected (Zubarev, R. A., & Makarov, A., 2013).

## 1.6 Data analysis

The data acquired using all of these different methodological techniques will first require pre-processing in order to convert the mass spectral or chromatographic file acquired into a usable format. Peak alignment is the first real requirement whereby the desired peaks (if known) or all of the peaks of interest, are matched up to each other in different sample sets enabling the intensities to then be compared.

For targeted GCMS and LC-MS analysis manual peak selection can be utilised in order to extract ion chromatograms peak intensities for specific identifying ions that are

known to be present in the metabolite of interest across the retention range they are known to arise, this is established using standards. For untargeted GCMS analysis, the signals require peak deconvolution and the assignment of each measured ion to a peak. The deconvolution aims to reduce the sample complexities through picking out the most prominent and clear peaks, separating them from others. Once this has been accomplished, the alignment of retention times and identifications of different files can be performed. Additionally, at this point peak identification can be made through comparisons to databases, such as NIST. Library comparisons also generate similarity scores to the peak data fragmentation patterns held in a library to establish the likeliness of the metabolite to the suggested identification. The amalgamated data is then ready for normalisation and statistical evaluation.

MALDI data has prior initial processing conducted in order to calibrate the data sets and apply signal to noise thresholds, smoothing, intensity thresholds, peak width acceptances, and define the peak detection algorithms. The data can then be aligned and amalgamated in the same way before statistical evaluation is applied.

#### 1.6.1- Statistical evaluations

In order to critique the biomarkers of interest statistical evaluation was conducted on the samples. First the normality distribution across the data sets was established using Kolmogorov-Smirnov testing whereby a p. value greater than 0.05 indicated a normally distributed dataset. In these instances statistically differentiation was established using two tailed students T-test's. If however, the distribution was not found to be normal (show through a p. value less than 0.05), statistical evaluation was instead conducted using two-tailed Mann-Whitney evaluation.

In instances where multiple hypothesis testing was required, data was first normalised before being analysed using ANOVA with a Post-hoc pairwise analysis then conducted to determine which groups showed the significant variation.

On particularly large data sets in order to better select the most useful markers any significantly different data were amalgamated, and the Bonferroni correction then applied and used to reduce the number of significant differences as well.



### 1.6.2- Identification provision

After selecting the metabolite features of interest during the evaluations the next stage involved generating identifications for these markers in order to assess their biological involvement. A number of different systems were utilised to provide metabolic identifications in this way.

#### 1.6.2.1- Standards

For targeted analysis including LC-MS and GC-MS standards of the known metabolites were first ran to capture their retention times and any relevant fragmentation information that could be acquired. This could then be matched up to the peaks recorded in the sample sets and utilised as the identification strategy.

#### 1.6.2.2- National Institute for Standards and Technology (NIST) library

For untargeted metabolites captured using GC-MS methodology, where the identities of the metabolites is completely unknown the NIST library was utilised. During this process the fragmentation patterns obtained for each peak are compared to the database for identification, along with the provisions of similarity scores to help assess the quality of the identifications being given, Figure 1-9. Even with a good level of similarity matches being obtained for the markers of interest, the identifications generated are only considered to be tentative at this stage and reported as such.

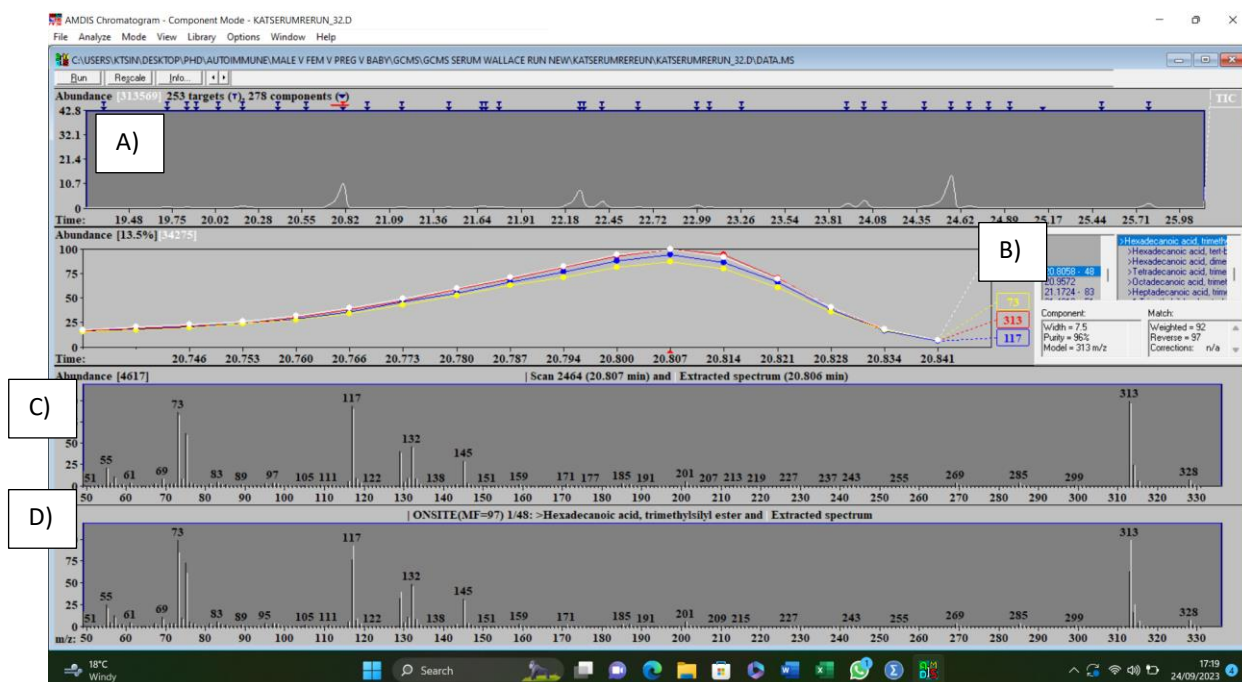


Figure 1-9- Example spectral identification made using NIST libraries couple to AMDIS whereby the A) peak recorded at retention time 20.8058 has generated an identity match with B) Hexadecanoic acid with a confidence scoring of weighted 92% and reverse 97%. A comparison is also made between the samples spectra C) and the NIST library spectra D).

### 1.6.2.3- Lipid maps

Lipid identifications, in this thesis, from MALDI and Orbitrap acquired data have been generated using LIPIDMAPS. The lipids of interest reported  $m/z$  is first entered into the online database, before filtering for what mode of ionisation and considerations of which adducts are likely present being accounted for. The mass tolerance threshold is adjusted according to what methodology has been utilised with accordance to the mass accuracy possible with the MS system before the classes of lipids suspected are filtered for. LIPIDMAPS then generates a list of likely lipid classes with examples of lipid possibilities for each. With MALDI analysis, as no further fragmentation was conducted, comparisons to the potential matches were made with online literature before reporting the tentative identifications. With Orbitrap analysis absolute identifications were possible through the generation of  $MS^2$  and  $MS^3$  data to match up the LIPIDMAPS spectral library to the fragmentation patterns seen in the sample set, examples of this are made in Chapter 5.

### 1.7- Ion suppression

Sample impurities and interfering compounds such as matrices can compete with molecules of interest for ionisation impacting the efficiency of the ionisation and dampening down signal detection. This can effect sensitivity, accuracy and reproducibility in measurements and so must be taken considered during experimental design. This includes considering the clean-up procedure, chromatography changes and internal standardisation selection. Common causative agents include salts, lipids and proteins (Annesley T. M. 2003).

### 1.8- Thesis structure

This thesis has achieved the customization of various mass spectrometry and associated techniques to identify novel biomarkers from diverse sample sources and conditions. The thesis is structured into distinct chapters representing different distinct project evaluations being undertaken, with each chapter focusing on a common theme of a poorly understood or diagnosed metabolic pathway that can be better clarified beyond current practices through biomarker discovery using tailored mass spectrometry techniques. This approach enhances our ability to uncover new metabolic patterns for diagnosis or understand the fundamental underlying pathways.

The methodology section of the first results chapter (Chapter 2) describes the majority of methods utilised across the multiple studies and hence the methodology sections of subsequent chapters describe only the additional / chapter specific methods utilised. Each chapter / project used different cell samples and analytical techniques, specific for the desired evaluation required.

**Chapter 2-** Utilises cell culture samples to find novel biomarkers for Pancreatic cancer diagnosis using both untargeted and targeted methodologies. GCMS, LCMS and MALDI techniques are utilised during the analysis.

**Chapter 3-** Utilised serum and breast-milk based sample sets to looks at the development of a metabolomic analysis capable of diagnosing endometriosis and PCOS conditions. Additionally key profile changes that occur in pregnancy as well as

establishing how these changes may be conveyed in breast-milk production to further support baby development after birth are analysed. GCMS and MALDI are utilised in this chapter using both targeted and untargeted analysis.

**Chapter 4-** Captures faecal markers capable of distinguishing successfully from unsuccessfully reproducing captive elephant using GCMS based untargeted analysis.

**Chapter 5-** Targeted lipidomic analysis was used to evaluate drosophila fly as a model for the neurological study of Alzheimer's disease using Orbitrap analysis.

**Chapter 6-** Targeted HPLC analysis was used to Establishes the relationship between taurine consumption and exercise performance relating to thermoregulation using plasma samples taken before and after exercise periods.

### 1.9- COVID Impact statement

During the pandemic the workflow, sample availability and research outputs were all impacted at a critical point of this thesis. Unfortunately, this resulted in a number of changes being made to the chapters in this thesis and some areas did not manage to achieve as vast a sample set as originally planned.

Chapter 2 was heavily impacted due to the loss of all initial cell-lines due to nitrogen dewars not being maintained during lockdown periods. This happened in the middle of the cellular data acquisition and resulted in alterations to cell types under study and a complete repeat of the data that had already been generated.

Chapter 3 was also heavily impacted due to a lack of sample availability. Serum sample collections were not possible during the pandemic. As a result, the topic of this chapter changed from being one seeking autoimmune biomarkers in neonates being traceable through serum sampling to one on fertility and neonatal nutrition instead as the breast-milk samples were readily available and the endometriosis samples were acquired at a later point but included a much smaller sample set.

These unavoidable issues have resulted in a loss of time, sample repeats and a change of tone to these impacted chapters.

## 1.10- References

- Annesley T. M. (2003). Ion suppression in mass spectrometry. *Clinical chemistry*, 49(7), 1041–1044. <https://doi.org/10.1373/49.7.1041>
- De Hoffman, E., Stroobant, V., (2007) *Mass Spectrometry: Principles and Applications*, 3rd Edition. West Sussex. United Kingdom: John Wiley & Sons, Ltd
- Hirayama, A., Sugimoto, M., Suzuki, A., Hatakeyama, Y., Enomoto, A., Harada, S., Soga, T., Tomita, M., & Takebayashi, T. (2015). Effects of processing and storage conditions on charged metabolomic profiles in blood. *Electrophoresis*, 36(18), 2148–2155. <https://doi.org/10.1002/elps.201400600>
- Jordaens, S., Zwaenepoel, K., Tjalma, W., Deben, C., Beyers, K., Vankerckhoven, V., Pauwels, P., & Vorsters, A. (2023). Urine biomarkers in cancer detection: A systematic review of preanalytical parameters and applied methods. *International journal of cancer*, 152(10), 2186–2205. <https://doi.org/10.1002/ijc.34434>
- Kardos, D., Simon, M., Vác, G., Hinsenkamp, A., Holczer, T., Cseh, D., Sárközi, A., Szenthe, K., Bánáti, F., Szathmary, S., Nehrer, S., Kuten, O., Masteling, M., Lacza, Z., & Hornyák, I. (2019). The Composition of Hyperacute Serum and Platelet-Rich Plasma Is Markedly Different despite the Similar Production Method. *International journal of molecular sciences*, 20(3), 721. <https://doi.org/10.3390/ijms20030721>
- Kienzl-Wagner, K., & Brandacher G., (2014) *Advances in Clinical Chemistry Chapter Six - Proteomics in Transplantation*. USA . Elsevier
- Nishiumi, S., Suzuki, M., Kobayashi, T., & Yoshida, M. (2018). Differences in metabolite profiles caused by pre-analytical blood processing procedures. *Journal of bioscience and bioengineering*, 125(5), 613–618. <https://doi.org/10.1016/j.jbiosc.2017.11.011>
- Njoku, K., Chiasserini, D., Jones, E. R., Barr, C. E., O'Flynn, H., Whetton, A. D., & Crosbie, E. J. (2020). Urinary Biomarkers and Their Potential for the Non-Invasive Detection of Endometrial Cancer. *Frontiers in oncology*, 10, 559016. <https://doi.org/10.3389/fonc.2020.559016>

- Pitt J. J. (2009). Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry. *The Clinical biochemist. Reviews*, 30(1), 19–34.
- Ser, Z., Liu, X., Tang, N. N., & Locasale, J. W. (2015). Extraction parameters for metabolomics from cultured cells. *Analytical biochemistry*, 475, 22–28.  
<https://doi.org/10.1016/j.ab.2015.01.003>
- Singh, R. K., Chang, H. W., Yan, D., Lee, K. M., Ucmak, D., Wong, K., Abrouk, M., Farahnik, B., Nakamura, M., Zhu, T. H., Bhutani, T., & Liao, W. (2017). Influence of diet on the gut microbiome and implications for human health. *Journal of translational medicine*, 15(1), 73. <https://doi.org/10.1186/s12967-017-1175-y>
- Styczynski, M. P., Moxley, J. F., Tong, L. V., Walther, J. L., Jensen, K. L., & Stephanopoulos, G. N. (2007). Systematic identification of conserved metabolites in GC/MS data for metabolomics and biomarker discovery. *Analytical chemistry*, 79(3), 966–973. <https://doi.org/10.1021/ac0614846>
- Thomas, C. E., Sexton, W., Benson, K., Sutphen, R., & Koomen, J. (2010). Urine collection and processing for protein biomarker discovery and quantification. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 19(4), 953–959. <https://doi.org/10.1158/1055-9965.EPI-10-0069>
- Zhgun, E. S., & Ilina, E. N. (2020). Fecal Metabolites As Non-Invasive Biomarkers of Gut Diseases. *Acta naturae*, 12(2), 4–14.  
<https://doi.org/10.32607/actanaturae.10954>
- Zubarev, R. A., & Makarov A. (2013) Orbitrap mass spectrometry. *Analytical chemistry*, 85(11) 5288-5296

## Chapter 2 – Metabolomic analysis of potential cellular biomarkers and the study of altered metabolism in Pancreatic and Lung cancer

### 2.1- Abstract

Pancreatic ductal adenocarcinoma (PDAC) is one of the most lethal malignancies, largely due to its late-stage diagnosis and poor survival rates. There is a critical need for improved diagnostic tools, and metabolomic analysis offers a promising approach to identifying and understanding the metabolic changes associated with pancreatic cancer cells. This chapter employed a range of mass spectrometry based techniques including MALDI, HPLC-MS and GC-MS. GCMS was used to explore global metabolic profile differences between different cancerous cell lines, as well as achieving a targeted analysis quantifying amino acid presence which was accountable for a large amount of the variation seen. A further evaluation looking at stable isotope tracing experiments was used to explore how the different pancreatic cancer cell lines adapted their energy resource, considering glucose and glutamine via the glycolysis and TCA cycle. MALDI analysis was used to explore variations in lipid and phospholipids present. HPLC-MS was used to quantify modified nucleoside levels in the urine of pancreatic cancer patients. Consistent with hypothesis there was a large number of significant differences observed between cancerous and non-cancerous cell-line types with a large number of amino acids, lipids and sugars found to be altered in cancerous cell lines. Stable isotope tracing experiments further examined how pancreatic cancer cell lines adapt their energy metabolism, specifically focusing on glucose and glutamine utilization via glycolysis and the TCA cycle. HPLC-MS quantified modified nucleoside levels in the urine of pancreatic cancer patients. Consistent with the hypothesis, numerous significant differences were observed between cancerous and non-cancerous cell lines, with alterations detected in amino acids, lipids, and sugars. Tracer experiments highlighted substantial differences in energy resource utilization, with upregulation in all cases but distinct preferences for glucose or glutamine as energy sources for ATP generation. As anticipated,

nucleosides were significantly upregulated in the urine of pancreatic cancer patients compared to healthy controls. Overall, this study highlights the complex metabolic adaptations of cancerous cell lines and the heterogeneity in how these cells reprogram their metabolism to support growth and survival.

## 2.2- Introduction

Pancreatic cancer is the 5<sup>th</sup> most deadly cancer diagnosis at present, despite being one of the least common, holding a diagnostic profile of just 3% of all cancers (Cancer research UK, 2017). A large reason for this ominous statistic is the difficulty faced during patient diagnosis; pancreatic cancer is typically a cancer of the pancreas' exocrine cells, effecting enzyme production rather than hormonal alteration. This results in a lack of symptoms and a challenging detection strategy. The most common and specific symptoms of pancreatic cancer include a loss of appetite, jaundice, abdominal pain, and unusual bloating of the upper gastrointestinal tract (Freelove R & Walling A., 2006; Holly, 2004). Typically, these symptoms do not present until the later stages of disease onset.

At the point at which a patient is suspected to have pancreatic cancer, the current diagnosis is based initially on analysing the patient's history, conducting physical examination and analysing the serum bilirubin and alkaline phosphate levels to give an indication of the disease. A CT scan is then typically issued by the GP. If pancreatic cancer is confirmed at this stage the GP typically refers the patient on to the hospital for a significantly more detailed triple phase CT scan to give the staging of the tumour. This then indicates the prognosis of the patient and narrows down the treatment options available. If the tumour presence is uncertain or more detail is needed on its location or stage an Endoscopic ultrasonography (EUS) is issued (Huggett. M.T., and Pereira S.P., 2011; Chari ST., 2007 and Freelove R., and Walling A.D. 2006). The majority of pancreatic cancer diagnosis made at the point of initial presentation are very late staged and advanced, typically stage III/IV (Ballehaninna UK, and



Chamberlain RS. 2012). Resection is the only curative treatment and success depends greatly on early diagnosis. At present only 15-20% of pancreatic cancer cases are diagnosed at a suitable stage for surgery (Ballehaninna, U.K. et al., 2012; Li D., et al., 2004 and Lindahl A., et al., 2007). If the patient is not considered suitable for resection due to metastasis etc. the only option is palliative treatment.

Improved diagnosis and screening strategy implementation could hold the key to dramatically improving the survival, which could possibly be found through metabolomic analysis. Serum tumour associate marker cancer antigen CA 19-9 (Ruckert F., et al., 2010 and Kim J.E., et al 2004) is currently the only recognised biomarker available for assisted pancreatic cancer diagnosis. CA 19-9 can successfully distinguish pancreatic cancer from pancreatitis when applied with IgG4 levels holding specificity and sensitivity of 100% and 84-94% respectively (Ruckert F., et al., 2010 and Kim J.E., et al 2004). Despite the success of CA 19-9 it is not able to be used as an independent diagnostic or screening tool due to specificity and sensitivity levels not yet being good enough. A stronger biomarker candidate could lead to improvement in this field.

A new strategy is needed to tackle the portentous and virtually asymptotic ailment in order to strive for the earliest possible detection, diagnosis and therefore providing the most efficacious patient prognosis possible. The potential of biomarkers in this area is vast, offering fast and reliable indications of disease presence from a variety of easily attainable and non-invasive sources. Assessment of a full metabolomic profile in cancerous cell lines can help to better elucidate novel markers that may be capable of providing this needed early screening and diagnosis. This study will analyse alterations in lipidome and metabolite pathway alterations using multiple techniques to gain a better overview of which pathways are being altered and why as well as directly seeking new biomarkers. Due to the lack of heterogeneity of pancreatic cancer cells two different cell lines will be utilised for the initial analysis. The changes found will also be compared to a very different cell line, lung cancer, to better understand the shared/ different markers. This will help guide the most versatile and robust markers for future analysis.

An additional line of study will seek waste product biomarkers from clinical samples in nucleoside analysis of urine. Modified nucleosides can be considered as a representative of RNA turn-around due to their inability to be recycled by the body. In malignancies the turnover of RNA is accelerated due to alterations in metabolic processes this leads to a larger presence of nucleosides in waste products of the body (Dudley, E., et al., 2003). Under standard healthy conditions the nucleosides can be directly utilised in biological processes or broken down into uric acid and excreted from the body. However, during malignancies such as cancers, the higher demand and turnaround of RNA's and elevated methyltransferase results in a larger presence of modified nucleosides, which, due to the structural alterations cannot be recycled and are instead excreted. Mass spectrometry can offer a fast and high throughput way of identifying and quantifying modified nucleosides in the urine samples. This study uses HPLC combined with mass spectrometry to seek 132 amu fragmentation patterns, indicative of a ribose sugar moiety loss. This loss was detected using ms2 fragmentation patterns whereby the ionized molecular species observed in the full scan (ms1) are separated according to  $m/z$  and then further fragmented before being detected. Ribose sugar is a common fragmentation for most modified nucleosides. The fragmentations can then be further used to identify and summate which modified nucleosides are present (Jora, M., et al., 2019).

The aim of this study was to elucidate quantitative differences in modified nucleosides presence of urine samples between pancreatic ductal adenocarcinoma (PDAC), healthy and other benign hepatic conditions such as pancreatitis and benign pancreatic cysts as well as other hepatic malignancy as other cancers such as duodenal cysts to assess whether this is an alternative route to finding a robust biomarker.

Finally, returning to the use of cell culture models of cancer, this study will seek to use stable isotopically labelled common precursors of energetic and metabolism pathways to study the altered metabolism of pancreatic cancer cells compared to healthy pancreatic cells and healthy and cancerous cells from another tissue (lung).

### 2.2.1. Chapter aims and objectives

The objectives of this chapter are as follows:

1. Identify novel biomarker candidates that can distinguish pancreatic cancer cell lines from healthy cell lines.
2. Identify markers in lung cancer cells using the same methodologies to provide a very different cell line for drawing comparisons.
3. Identify which of these markers are conserved between pancreatic cancer cell line types as well as also determining which are also indicative of different cancers, in this study lung cell lines were used. The overall goal being to therefore identify which are indicative of both pancreatic cancer cell lines, which are unique/ more pronounced in only one particular pancreatic cell line.
4. Use targeted analysis to seek modified nucleoside alterations in urine samples from pancreatic cancer patients.
5. Present a broader understanding of the underlying processes that occur during pancreatic cancer and identify those changes that may be of use in a prognostic capacity in future studies using stable isotope labelling
6. Compare the different biomarker discovery techniques to assess which would give the best diagnostic markers and show the greatest differences in metabolic patterns whilst also considering which would be the most simplistic and available to deploy.

## 2.3 Methodology

### 2.3.1- Cell lines

A variety of different adherent epithelial human cell lines have been incorporated into the analysis of pancreatic cancer including HDPE (representing healthy pancreatic cells), Panc-1 and MIA PaCa-2 (as pancreatic cancer cell lines) in addition to lung cancer cell lines A549 (cancerous) and Beas-2b (healthy) (Gradiz, R., et al., 2016 & Hillyer, P., et al., 2018). The properties of these are displayed in Table 2.2-1- Cell line specification below.

Table 2.2-1- Cell line specifications

Cell type	Donar gender	Donar age/years	Disease	Growth media	Doubling Time/ hours
HDPE(E6c7) Pancreatic			Healthy immortalised with HPV116-E6c7 inhibiting p53 production	RPMI 1640 + 10% FBS	NA
Panc-1 Pancreatic	Male	56	Epithelioid carcinoma	DMEM + 10% FBS	52
MIA PaCa-2 Pancreatic	Male	65	Epithelial cell Carcinoma	DMEM + 10% FBS  Horse serum 2.5%	40

Beas-2B Lung			Healthy Lung bronchus Normal epithelial tissue SV-40 immortalized	DMEM + 10% FBS	NA
A549 Lung	Male	58	Epithelial cell Carcinoma	DMEM + 10% FBS	22

#### *2.3.1.1 Cell culture consumables*

All cell lines were gifted by Dr Claire Morgan, Swansea University having previously been purchased from ATCC. The HDPE cell lines were grown in RPMI (Sigma- R8758) supplemented with 10% Fetal bovine serum (FBS), penicillin and streptomycin, whereas all other cell lines were grown in Gibco DMEM (GIBCO-11995065) supplemented with 10% FBS, penicillin and streptomycin. Cells were maintained in a 5% CO<sub>2</sub> in a humidified incubator at 37°C.

#### *2.3.1.2 Cell extraction, derivatisation and analysis*

##### *2.3.1.2.1 GCMS extraction and derivatisation*

For the evaluation of the broadest set of metabolites of interest a non-targeted methodology was utilised through GCMS analysis to capture as many metabolites for statistical evaluation as possible. Cells were grown in T75 flasks until approximately 80% confluency. The media was then removed, and the cell surface washed twice with 5ml of PBS. The cells were then quenched with and extracted into 800µL of ice-cold methanol using a cell scraper. The cells were sonicated for 30 minutes before the cell suspension was then centrifuged at 10, 000 x g for ten minutes, before the supernatant was collected. Ten µL of the supernatant was taken off for Bradford protein assay, to enable the supernatant to then be normalised in line with the lowest concentration of protein extracted. This was conducted through incubating the 10 µL of sample with 500 µL of Bradford reagent which was shaken and incubated for five minutes before have the absorbance value read at 595nm. The normalised

supernatant was then speed vacuum concentrated using an Edwards vacuum concentrator (Edwards, UK) to dryness before derivatisation for GCMS analysis.

Derivatisation for the total metabolite profile was undertaken through the addition of 30µL of 15mg/mL methoxylamine hydrochloride (in pyridine) to the dried sample. This was then incubated for 60 minutes at 70°C before 50µL of MSTFA + 1% TMCS Silylation Reagent (Thermofisher, UK) was added prior to a further incubation at 40°C for 90 minutes. Samples were then transferred into 300µl total volume glass vials and 10uL of 2mg/mL tetracosane (in hexane) was added for use as an internal standard.

#### 2.3.1.2.2 GCMS analysis

All samples in the analysis were run on an Agilent GC6890 chromatograph using an Agilent 19091S-433 capillary column (30m x 0.25mm x 0.25µm) with a flow rate of 1ml/min. 1µL of sample was injected in split-less mode using a temperature gradient from 180-300°C over the course of 15 minutes, with the run then lasting for a total of 58 minutes. Following analysis, peak areas were extracted from the total ion chromatograms and normalised to the internal standard peak area for inter-sample comparisons and metabolites were tentatively identified by comparison of the mass spectra generated to those of derivatised compounds within the NIST database.

Twelve amino acid were further analysed to make a targeted quantitative analysis of fluctuations in the cell extracts. The amino acids were prepared to the desired concentrations and then extracted in the same manner as the cell culture extracts. These were ran via GCMS using the same specifications as the previous untargeted analysis and generated standard curves to provide quantitative details of amino acid fluctuations between samples.

Data was analysed using MSD Chemstation whereby amino acids ion chromatograms were extracted,

Table 2-2, and the peak areas was taken for evaluation after normalisation to the tetracosane peak. Standard curves were generated with triplicate data points and

taken, and the linearity then used to calculate the concentrations present in the cell cultures.

Table 2-2.2- Amino acids analysed in experiment with chromatogram ion selected and retention time

Amino acid	Retention time/min	Ion extracted/ <i>m/z</i>
Alanine	7.20	116
Valine	8.78	144
Leucine	9.50	158
Isoleucine	9.80	158
Glycine	10.10	248
Threonine	11.11	218
Methionine	12.74	176
Aspartic acid	12.78	232
Cysteine	13.20	220
Glutamate	14.00	246
Arginine	16.93	142
Tyrosine	18.99	218
Tryptophan	24.20	202

#### 2.3.1.2.4 MALDI extraction and analysis

For analysis via MALDI, nine replicate examples of each cell line were grown to ~80% confluency before being trypsin extracted and pelleted with 3 washes in 3mL of PBS. A 100µL aliquot was taken off for protein content normalisation (after a twenty-minute sonication 5µL was mixed with 250µL Bradford reagent) and then the remaining PBS was removed and cells were stored at -80°C. The cells were then resuspended in water and normalised to the lowest protein concentration to ensure the same proportion of cells were in each according to internal protein content. The samples were then speed vacuum concentrated to dryness. These were then analysed using two different matrices. The first was lipid extracted using a 100µL chloroform methanol extraction (2:1) with the addition of 10µL water. After thorough vortexing, the sample was centrifugation at 5,000 rpm in order to assist phase

separation before 50µL of the lower chloroform layer was collected and speed vacuum concentrated to dryness. This was then resuspended in 2µL of the 9-Aminoacridine (9-AA) matrix (30 mg/ml of 9-AA reconstituted in 3:2 Isopropanol:Acetonitrile). 0.5µL of the suspension was spotted onto the ground steel target and dried before being analysed in positive mode on the Bruker MALDI system. Phospholipids, due to their head groups, can hold either a positive or negative charge. For this reason, utilising both positive and negative mode analysis is essential for the most extensive untargeted lipidomic analysis.

The other 30µL cell pellet aliquot was kept dry and directly resuspended in CHCA matrix (5mg CHCA dissolved in 50% ACN and 50% 0.1% TFA ) with 1µL being spot onto the plate and again analysed in positive ionisation mode.

The spotted samples were then ran on a Bruker Ultraflex using a 20kV ion source extraction voltage with random walk set to partial sample and 200 shots being taken over a mass range of 400-2000Da, in an attempt to improve data acquisition. This was calibrated using  $\text{CsI}_3$  before a set of eight samples was read. Centroid peak detection algorithm was used to select peaks for analysis, with a signal to noise ratio of 4 applied to a maximal number of 500 peaks being selected with a peak width of at-least  $0.1m/z$  and a height of 80%. During the data acquisition, TopHat baseline subtraction was applied, with SavitzkyGolay smoothing (with a width of  $0.01 m/z$  over 5 cycles). Additionally, a matrix suppression of 450Da was applied in deflection mode, this limits which ions are deflected in order to boost analyte signal and reduce background peaks as to not suppress the high matrix ions.

All data was analysed using Microsoft excel and SPSS. The data was first compiled into subsets for each group consisting of nine readings for each cell type, listing the  $m/z$  vs relative intensity values obtained. These were then consolidated into  $m/z$ 's and intensity for all with  $m/z$ 's rounded to the nearest half a decimal place. To reduce the dataset to include only the most robust signals,  $m/z$  readings that occurred in less than half of the replica samples were removed from the analysis, before the average relative intensity was consolidated on a new worksheet for each  $m/z$  and cell type.



#### 2.3.1.2.4 Stable Isotope Tracer Analysis

All cell lines were grown to confluency in 10cm Petri dishes with suitable fully supplemented media under standard growth conditions. For cell lines A549, Beas-2B, MIA PaCa-2 and Panc-1 was Dulbecco's Modified Eagle's Medium (DMEM) with 10% Fetal bovine serum (FBS), whereas HDPE was grown in Roswell Park Memorial Institute (RPMI) media also supplemented with 10% FBS.

Labelled media was prepared in advance which comprised of DMEM Stable isotope labelling by amino acids (SILAC) media with no glucose, glutamine or phenol red added (Gibco: A1443001), RPMI SILAC media with no glucose, glutamine, 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid, N-(2-Hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid) (HEPES), phenol red, sodium pyruvate, Arginine or Lysine (these were added at a concentration of 200mg/L and 40mg/L respectively).

Both media types were supplemented with 1mM of sodium pyruvate (Gibco: 11360039) and 10% filtered and dialyzed FBS (GIBCO: A3382001). U- $^{13}\text{C}^6$  labelled tracer metabolites glucose and glutamine were ordered in from CK isotopes (CLM-1396 and CLM-1822-H-0.1 respectively). Three different additions were then made for the tracers whereby the concentration of glucose was 2000mg/L and glutamine was 300mg/L that comprised of a) Labelled glucose  $\text{C}^6$  + unlabelled glutamine b) Labelled glutamine  $\text{C}^6$  + unlabelled glucose c) No Glucose + labelled glutamine  $\text{C}^6$  d) unlabelled glucose + unlabelled glutamine.

At the point of 80% confluency the basic media was removed, and the cell lines were thoroughly washed three times with PBS, ensuring all was removed and no media remained. The specialist media was then put on the cells and the cells were incubated for 24 hours. Each plate was then extracted twice through the addition of 500 $\mu\text{L}$  of ice cold 60% methanol with cell scraping whilst keeping the plate on ice. The extraction was conducted twice to ensure the full transfer of all cells. The 1mL total volume of extract was then sonicated in a water bath for 30 minutes before being speed vacuum concentrated to dryness under cold conditions. Derivatization was then carried out through the addition of 30 $\mu\text{L}$  of methoxylamine hydrochloride (in pyridine) to the dried sample. This was then incubated for 30 minutes at 70°C before

50µL of MTBSTFA (WITH 1% T-BDMCS) Silylation Reagent (Scientific Laboratory Supplies, UK; TS-48927) was added prior to a further incubation at 40°C for 90 minutes. Samples were then transferred into 300µl total volume glass vials and 10uL of 2mg/mL tetracosane (in hexane) was added for use as an internal standard. These were then analysed using GCMS analysis using the same conditions as 2.2.1.2.2 GCMS analysis

### 2.3.2. Urine modified nucleoside analysis

#### 2.3.2.1. Samples and extraction

One hundred and twenty-one previously collected and frozen urine samples (from healthy, pancreatic cancer, pancreatitis, other cancer volunteers- ethic approval : (11/WA/0807- summarised in Table 2-3) were extracted according to Teichert. F., et al. (2011) with some modification. Briefly, the samples were thawed on ice before being centrifuged at 14, 000 x g for 15 minutes at 4°C and having internal standard, tubercidin added at a concentration of 50pmoles. The supernatant was then SPE extracted using Oasis HLB columns (Walters, Elstree, UK) which were pre-conditioned with 1mL of Methanol, followed by 1mL of water before 1mL of samples was loaded and washed with 1mL of water and eluted with 1mL of water:ACN (20:80). The extract was then vacuum concentrated to dryness. The SPE purified urine samples were re-suspended in 200µl of water, centrifuged and 6 fold diluted in mobile phase. 10ul of tubercidin standard was added to 60µl of sample before HPLC-MS analysis.

Table 2-3- Details and patient and healthy urine donors

N	Healthy	Pancreatic ductal adenocarcinoma	Pancreatitis	Other pancreatic conditions	Other Cancer
121 (52 F; 69 M)	32 (15F; 17 M)	60 (25F; 35 M)	15 (6 F; 9 M)	10 (4F; 6M)	4 (2F; 2M)
Mean age	62 (+/- 16.59)	67 (+/- 11.74)	58.6 (+/- 14.38)	58 (+/- 12.78)	63.75 (+/-9.42)

#### 2.3.2.2. HPLC-MS analysis

An ULTIMATE 3000 HPLC system (Dionex, UK), coupled to a LCQ DECA XPplus electrospray mass spectrometer (ThermoFinnigan, UK) was used for analysis. Separation was achieved using a 300um x 25cm C18pepmap column (Dionex, UK) which was eluted with an acetonitrile gradient in 0.15 formic acid (water) at 4ul/minute flowrate. The analysis was conducted over a  $M/Z$  range of 60-545 using  $ms^2$  fragmentation with a capillary temperature of 200°C. ESI was conducted with a positive polarity with a Source Voltage of 3.50 kV, Current of 80.00 uA, Capillary Voltage of 5V and Tube Lens Offset 30V. Additionally the Entrance Lens was set to -56.00V. Data dependent analysis mode using full scan mass spectrum followed by MSMS analysis of the three most abundant ions was used. A constant neutral loss of 132Da was used to identify nucleosides. Data was analysed directly from the Excalibur software package and full scan and MS/MS spectra compared to known standards and the literature for tentative identification of the modified nucleosides.

#### 2.3.3. Biomarker evaluation

The evaluation of the diagnostic potential of any prospective biomarkers identified was determined in line with guidelines as outlined by Ray et al (2010). Generally, following analysis of normality, a T test or Mann Whitney test was carried out to determine statistical significance, and where appropriate, multiple hypothesis testing accounted for using a Bonferroni correction. The prospective biomarkers identified were next used to generate a receiver operator characteristic (ROC) curve from which the area under the curve, sensitivity, specificity (as an inverse of 1-specificity), accuracy and positive and negative likelihood ratios could be determined. These values were compared with the suggested thresholds identified in Ray et al in order to categorise a prospective biomarker as excellent, good or poor as per 1.2.1- Biomarker critique using ROC curves.

## 2.4 Results

### 2.4 1. GCMS

The first analysis undertook the analysis of extracted metabolites from healthy and cancerous cell lines from both the lung and the pancreas in order to identify any prospective biomarkers and the ability of metabolomic analysis to differentiate different states in both tissues.

Six hundred and forty-one different metabolite markers were selected and utilised in the analysis to seek significant variance between cancerous and non-cancerous groups using GCMS analysis. This involved the comparison of nine replicate extracted samples from five different cell lines: HDPE, Panc-1, MIA PaCa-2, Beas-2b and A549. Data processing was conducted on Microsoft Excel and SPSS. The markers observed in the analysis included a range of different metabolite groups. Tentative identifications were then given to the markers of interest based on NIST identifications through the NIST 2005 database, any markers that did not provide an identification or were identified as being due to derivatisation agents or column bleed were eliminated from the study. Relative abundances were used for quantitative comparisons in the study.

#### 2.4.1.1 *Untargeted GCMS analysis*

PCA plots were generated using the metabolite markers detected in the 5 different cell types under evaluation, HDPE, Panc-1, Mia PaCa-2, A549 and Beas-2b's. The plots can be used to determine if a sufficient level of separation was found to suggest that these markers would be capable of distinguishing the different groups from each other and also better appreciate which groups have the most similarity to each other.

The first PCA plot considered all 5 of the cell lines, Figure 2-1a. In the analysis the healthy cell lines HDPE and Beas-2b occupied the top right quadrant, the pancreatic cell lines (both healthy and cancerous) were towards the lower half of the plot whilst the lung more so towards the top. The lung cancer and healthy lung groupings showed a very good degree of separation shown by distinct confidence intervals with little overlap of the groups seen, Figure 2-1b.

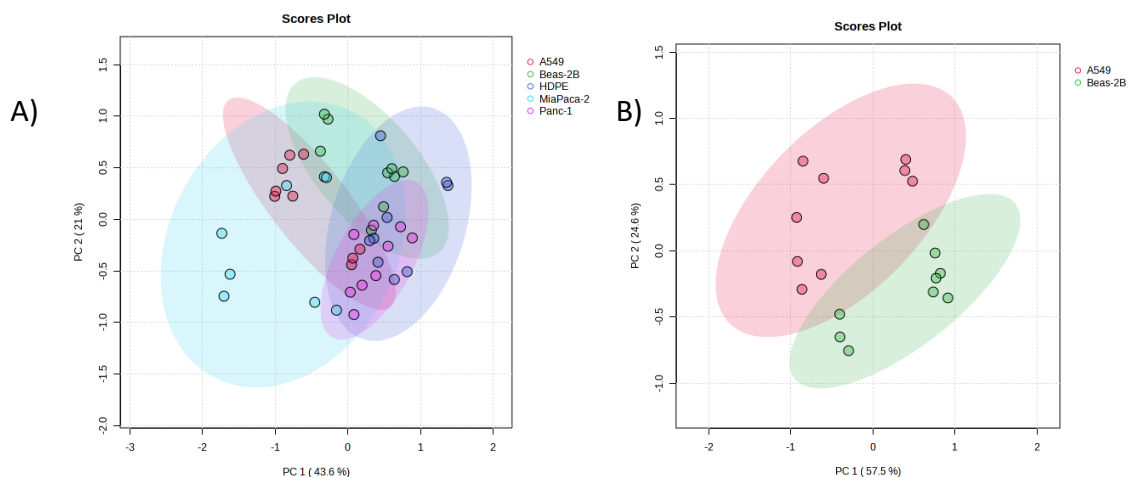


Figure 2-1 PCA showing the relationship of all of the different cell line groups when considering the most significantly different metabolites, comparing all cell lines (A) and just the lung cancer cell lines (B).

When generating the PCA plot specifically for the pancreatic cancer cells the separation between cell types was less clear. In this instance the healthy cell line, HDPE was towards the left half of the plot and overlapped by Panc-1 in the centre and Mia PaCa-2 more towards the right and most sparsely spread out and in this way demonstrates a great difference between the different forms of cancerous cell lines.

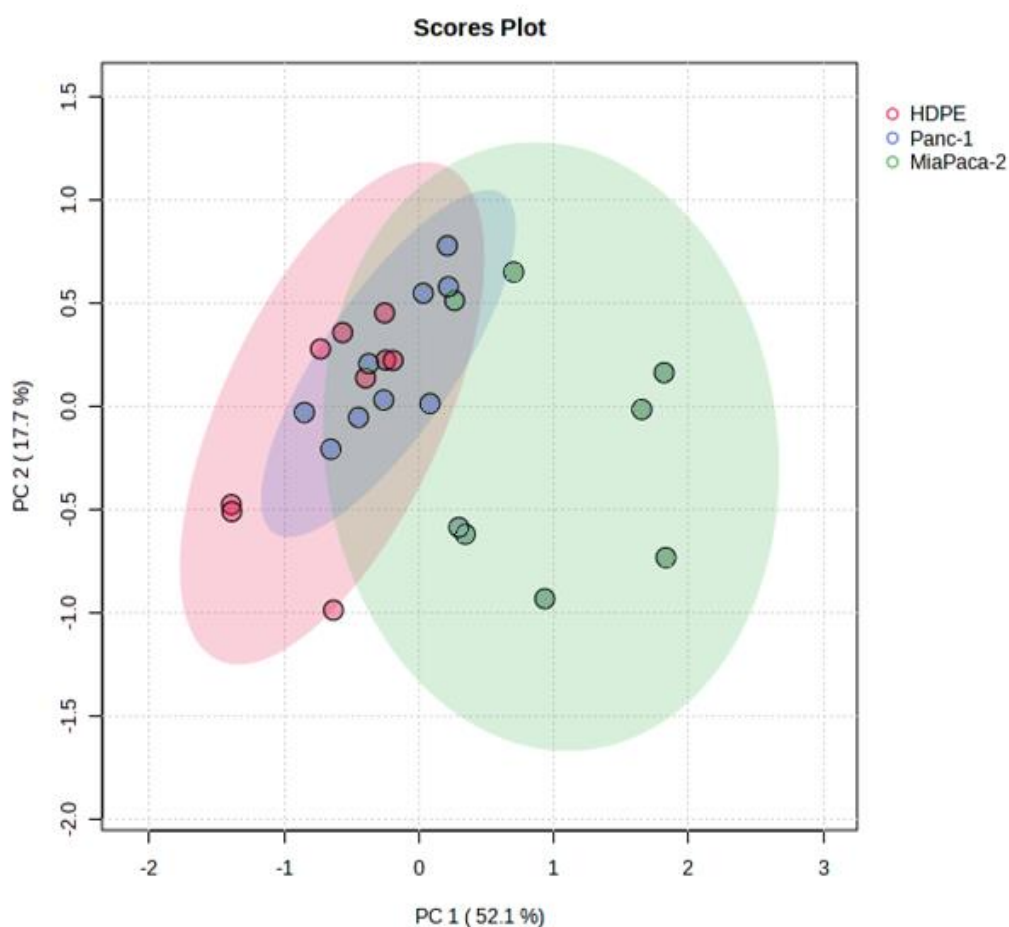


Figure 2-2. PCA chart showing the relationship between the pancreatic groups significantly different markers including HDPE (healthy), Mia PaCa-2 and Panc-2 (cancerous).

After establishing a good degree of separation of the groups, the next stage of the evaluation looked to determine which metabolite markers were responsible for the separations seen.

### Lung cancer cell analysis

When comparing healthy and cancerous lung cancer lines one hundred and twenty-five different markers were found to show significant variance in total, achieving a P value of less than 0.05. To minimise this list of markers to those that bared the most importance for diagnostic utilisation as individual biomarker candidates, firstly a Bonferroni correction was first applied (in order to account for multiple hypothesis testing) followed by curve scoring analysis. Twenty-two markers were selected as achieving better than “good” diagnostic scoring, illustrating their significant impact

as biomarkers for lung cancer. This was narrowed further to nineteen that provided NIST identifications for further evaluation as shown in Figure 2-3.

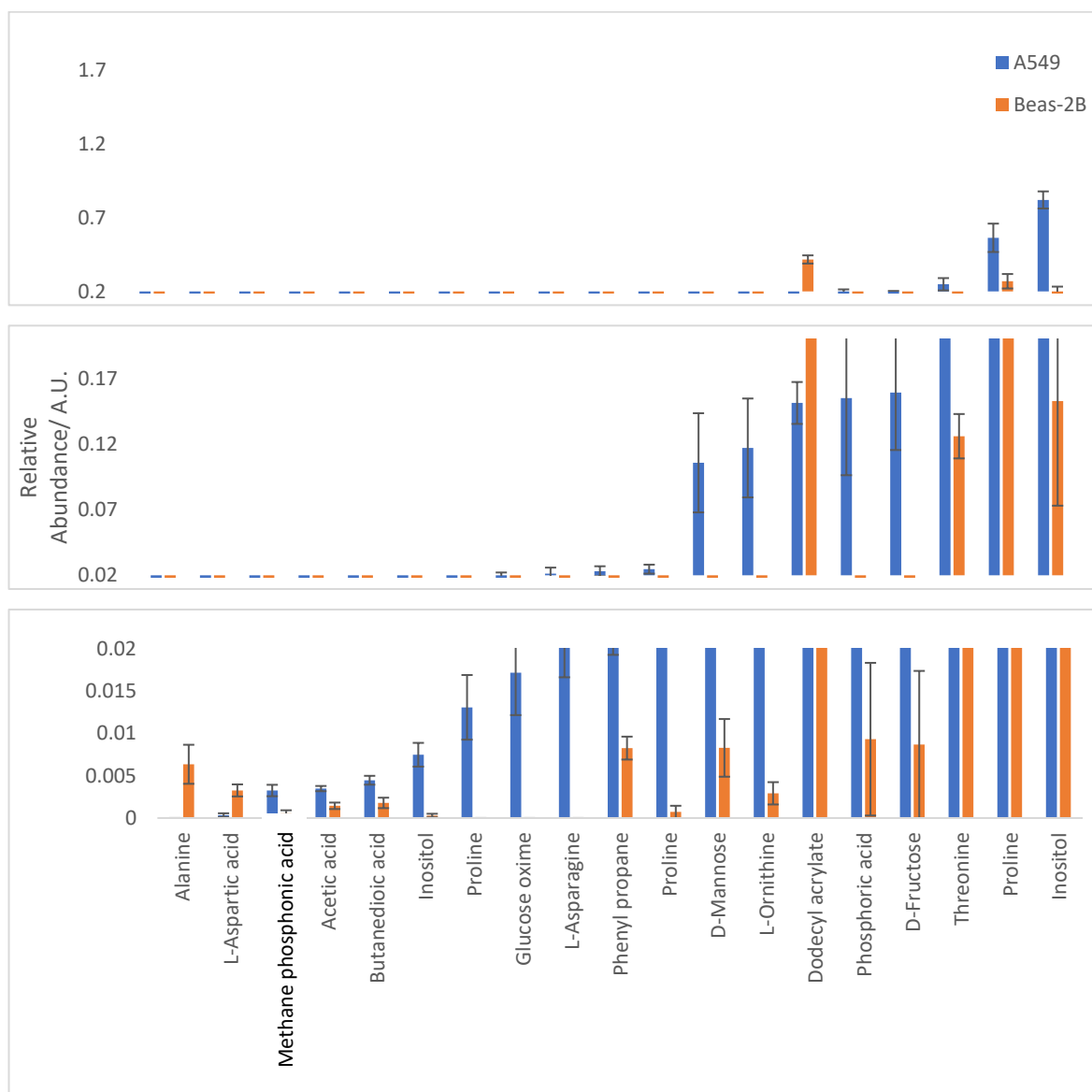


Figure 2-3- Significantly different metabolites that were successfully identified that differentiate A549 cancerous cell lines from BEAS-2B non-cancerous cell lines whilst achieving better than “good” diagnostic scorings. All p. values obtained were less than 0.007. Error bars represent standard error, n= 9 for each cell line. “Methane” represents Methane phosphonic acid.

Of these most significant markers that generated accepted identifications, two markers achieved “excellent” diagnostic scores; asparagine and dodecyl acrylate, illustrating their strength as biomarker candidates. The latter of these markers has been included as despite its unnatural assigned identification the strong biomarker

correlation suggests that it is a true biological candidate and so brings the identification into question. A caveat of using GCMS in this untargeted manner is that there is a limit into the certainty of the identifications assigned. The highest scoring markers diagnostic properties are listed in Table 2-4. Most of the prospective biomarkers demonstrated an increase in levels in the cancerous cell line compared to the healthy cell line with the exception of 5 metabolites tentatively identified as biomarkers. It should be noted, when considering the identifications, that proline is identified multiple times. These multiple identifications of the same biomolecule, may partially arise through isomers of the molecule existing, which cannot be distinguished by GCMS, or metabolites of the biosynthesis or catabolism of the biomolecule having sufficiently similar mass spectra as to not allow differentiation via the NIST spectrum match identification process.

Table 2-4- The most diagnostically successful metabolites, ranked from highest to lowest intensity, identified between A549 and Beas-2b cell lines identified using GCMS analysis. It should be noted that an  $\infty$  symbol is utilised to represent where the value cannot be computed. In this instance LHR+ is calculated by taking the Sensitivity/ 1- Specificity, however in this instance the 1-specificity is zero. “Excellent” markers are shown in green in the table and “good”.

Tentative ID	p. value	AUC	Sensitivity	1-Sensitivity	LHR +	LHR-
Alanine	0.001	0.944	90	10	$\infty$	0.10
L-Aspartic acid	0.001	0.988	100	0	10	0
Methane phosphophonic acid	0.001	0.889	90	10	9	0.11
Acetic acid	0.001	0.914	100	0	10	0
Butanedioic acid	0.001	0.877	90	10	9	0.11
Inositol	0.001	0.932	90	10	$\infty$	0.10
Proline	0.001	0.944	90	10	$\infty$	0.10
Glucose oxime	0.001	0.944	90	10	$\infty$	0.10
L-Asparagine	0.001	1	100	0	$\infty$	0
Phenyl propane	0.004	0.901	90	10	9	0.11
Proline	0.001	0.938	90	10	$\infty$	0.10



D-Mannose	0.001	0.963	90	10	$\infty$	0.10
L-Ornithine	0.002	0.92	90	10	$\infty$	0.10
Dodecyl acrylate	0.001	1	100	0	$\infty$	0
Phosphoric acid	0.003	0.895	90	10	9	0.11
D-Fructose	0.001	0.926	90	10	9	0.11
Threonine	0.004	0.901	100	0	5	0
Proline	0.007	0.877	90	10	9	0.11
Inositol	0.001	0.988	100	0	10	0

### Pancreatic cancer cell line analysis

When comparing the pancreatic cancer cell line Panc-1 against the healthy cell line HDPE eighty-one different markers were found to show significant variance between the two groups. Of which, nine were found to achieve better than good scorings with regards to AUC and LHR scoring and these metabolites levels are shown in Figure 2-4. In this comparison, all the biomarkers were found to be elevated in the cancer cell line compared to the healthy cell. Three of the markers of interest; L-Proline, Phosphoric acid and Tyrosine achieved excellent diagnostic values proving to be the strongest candidates of biological markers.

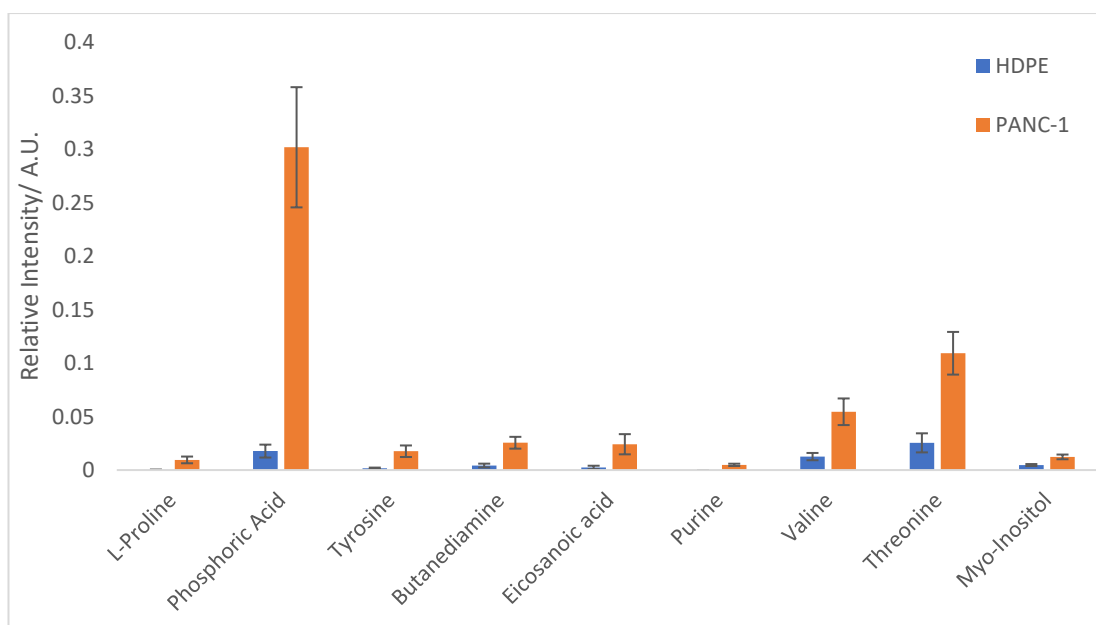


Figure 2-4 Graphical representation of the most significantly different metabolite markers that distinguish HDPE from Panc-1 pancreatic cells. These all achieved p. values of less than 0.005 and better than “good” diagnostic values relating to the AUC and LHR.

The diagnostic properties of these markers as established through ROC analysis are shown in Table 2-5, which illustrate the high capability of these biomarkers for differentiation.

Table 2-5- The most diagnostically successful metabolites identified between HDPE and Panc-1 cell lines identified using GCMS analysis. It should be noted that an  $\infty$  symbol is utilised to represent where the value cannot be computed. In this instance LHR+ is calculated by taking the Sensitivity/ 1-Specificity, however in this instance the 1-specificity is zero. “Excellent” markers are shown in green in the table and “good” and “poor” in red.

Tentative ID	p. value	AUC	Sensitivity	1-Sensitivity	LHR+	LHR-
L-Proline	0.001	1	100	0	$\infty$	0
Phosphoric Acid	0.001	1	100	0	$\infty$	0
Tyrosine	0.001	1	100	0	$\infty$	0
Butanediamine	0.002	0.938	100	0	10	0
Eicosanoic acid	0.003	0.901	100	0	5	0
Purine	0.001	0.944	90	10	$\infty$	0.1
Valine	0.002	0.938	90	10	$\infty$	0.1

Threonine	0.001	0.951	90	10	9	0.11
Myo-Inositol	0.005	0.889	90	10	9	0.11

A further comparison was undertaken between the healthy pancreatic cell line and a second pancreatic cancer cell line MIA PaCa-2. There were one hundred and ninety-six significantly different markers found to distinguish MIA PaCa-2 internal profiles from healthy HDPE cell lines. After Bonferroni corrections and assessment of diagnostic quality, utilising AUC and LHR scoring fifty-nine were found to achieve better than “good” diagnostic quality scoring, ten of which excellent in all diagnostic traits. Forty-nine of these generated suitable tentative identification via NIST as shown in Figure 2-5.

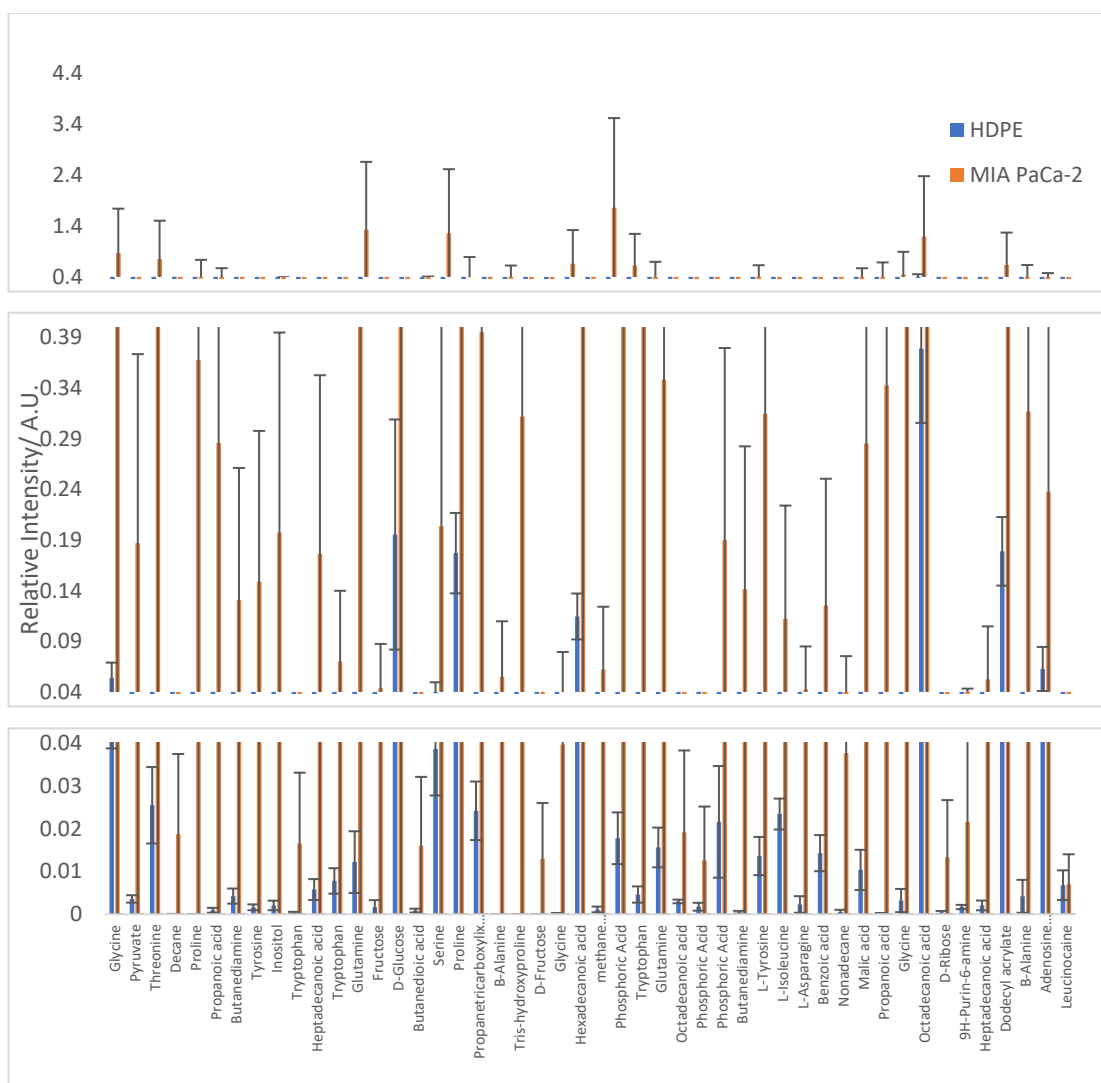


Figure 2-5 Significantly different metabolite markers that distinguish MiaPaca-2 cancerous models from HDPE healthy pancreatic models using internal metabolites that achieved better then “good” diagnostic scorings and provided sensible identifications via NIST. All p.values were below 0.008.

The potential diagnostic properties of these most significantly different metabolites are shown in Table 2-6.

Table 2-6-The most diagnostically successful metabolites identified between HDPE and MIA PaCa-2 cell lines identified using GCMS analysis. It should be noted that an  $\infty$  symbol is utilised to represent where the value cannot be computed. In this instance LHR+ is calculated by taking the Sensitivity/ 1-Specificity, however in this instance the 1-specificity is zero. “Excellent” markers are shown in green in the table and “good” and “poor” in red.

Tentative ID	p. value	AUC	Sensitivity	1-Sensitiity	LHR+	LHR-
Glycine	0.001	1	100	0	$\infty$	0

Pyruvate	0.001	1	100	0	$\infty$	0
Threonine	0.001	1	100	0	$\infty$	0
Decane	0.001	1	100	0	$\infty$	0
Proline	0.001	1	100	0	$\infty$	0
Propanoic acid	0.001	1	100	0	$\infty$	0
Butanediamine	0.001	1	100	0	$\infty$	0
Tyrosine	0.001	1	100	0	$\infty$	0
Inositol	0.001	1	100	0	$\infty$	0
Tryptophan	0.001	1	100	0	$\infty$	0
Heptadecanoic acid	0.001	0.988	100	0	10	0
Tryptophan	0.001	0.988	100	0	10	0
Glutamine	0.001	0.975	100	0	10	0
Fructose	0.001	0.975	100	0	10	0
D-glucose	0.002	0.938	100	0	10	0
Butanedioic acid	0.004	0.895	100	0	5	0
Serine	0.001	0.963	90	10	$\infty$	0.1
Proline	0.001	0.963	90	10	$\infty$	0.1
Propanetricarboxylic acid	0.001	0.963	90	10	$\infty$	0.1
B-alanine	0.001	0.944	90	10	$\infty$	0.1
Tris-hydroxyproline	0.001	0.944	90	10	$\infty$	0.1
D-Fructose	0.001	0.944	90	10	$\infty$	0.1
Glycine	0.001	0.938	90	10	$\infty$	0.1
Hexadecanoic acid	0.002	0.926	90	10	$\infty$	0.1
Methane phosphonic acid	0.002	0.92	90	10	$\infty$	0.1

Phosphoric acid	0.003	0.907	90	10	$\infty$	0.1
Tryptophan	0.003	0.907	90	10	$\infty$	0.1
Glutamine	0.004	0.901	90	10	$\infty$	0.1
Octadecanoic acid	0.005	0.889	90	10	$\infty$	0.1
Phosphoric acid	0.008	0.864	90	10	$\infty$	0.1
Phosphoric acid	0.012	0.84	90	10	$\infty$	0.1
Butanediamine	0.001	0.988	90	10	9	0.11
L-Tyrosine	0.001	0.975	90	10	9	0.11
L-Isoleucine	0.001	0.951	90	10	9	0.11
L-Asparagine	0.001	0.938	90	10	9	0.11
Benzoic acid	0.003	0.914	90	10	9	0.11
Nonadecane	0.002	0.914	90	10	9	0.11
Malic acid	0.003	0.914	90	10	9	0.11
Propanoic acid	0.002	0.914	90	10	9	0.11
Glycine	0.003	0.901	90	10	9	0.11
Octadecanoic acid	0.004	0.901	90	10	9	0.11
D-ribose	0.004	0.883	90	10	9	0.11
9H-Purin-6-amine	0.007	0.877	90	10	9	0.11
Heptadecanoic acid	0.006	0.877	90	10	9	0.11
Dodecyl acrylate	0.009	0.864	90	10	9	0.11
B-Alanine	0.009	0.858	90	10	9	0.11
Adenosine monophosphate	0.012	0.852	90	10	9	0.11

Next, a heatmap was generated to view the overall trend seen in all of the significantly different metabolites variation between groups and identify how these impact the way in which the different groups then relate to each other (). Where

metabolite levels were distinctive of a cell type, but the NIST database comparison could provide no tentative identification for the metabolite, the metabolite has been utilised in the heatmap analysis, but referenced to as x followed by a number to differentiate unidentified metabolites. This process accounted for 8 differentially expressed metabolites.

The first branch of the heatmap separated 78% (7/9 replicate samples) of the MIA PaCa-2 cell line extracts and 2 of the A549 lung cells from the other groups. The next arm subdivided into two further arms with one holding 78% of the Panc-1 samples and two of the MIA PaCa-2 samples. The other branch further subdivided into three with the left most branch holding all of the HDPE and 78% of the Beas-2b lines together, the remaining two Beas-2b samples were on the neighbouring branch with the final branch occupied by all bar two of the A549 samples, two of the Panc-1 samples and one of the MIA PaCa-2 cell lines. In this way it can be seen that a good level of separation was achieved between the cancerous and non-cancerous cell lines through their independent accumulation on one branch.

When considering the trends shown in the heatmaps, and , there are some distinct trends that appear to be unique to different cell types. For example HDPE cell lines show an overall low presence of all of the metabolites under analysis when compared to the others, whereas MIA-PaCa-2 cell lines have a much higher presence in almost all markers.

Looking to the trends shown in the group averages, , this results in Mia PaCa-2 show this as being the most distinct group followed by A549 and Panc-1, with the two healthy lines HDPE and Beas-2b being the most similar.





Considering the general group trends defined in the heatmap, the MIA PaCa pancreatic cancer cell line is identified as the most distinct with a comparative up regulation of a number of biological metabolites, the top 6 metabolites in the heatmap (e.g. alanine) are solely elevated in the healthy lung cancer cell lines, whilst the healthy pancreatic cell line exhibited no singly elevated metabolite, however it did uniquely express a decreased production of adenosine monophosphate. When considering biomarkers of pancreatic cancer, metabolites listed between that annotated as X5 and mannose, appeared to be uniquely elevated in cancerous pancreatic cell lines.

When a heat map was generated using just the variables obtained for lung cell lines alone again a very clear separation was achieved between the cancerous and non-cancerous line with the first branch holding all of the Beas-2b plus three of the A549 samples whereas the second branch held all bar 2 of the A549 cell lines.

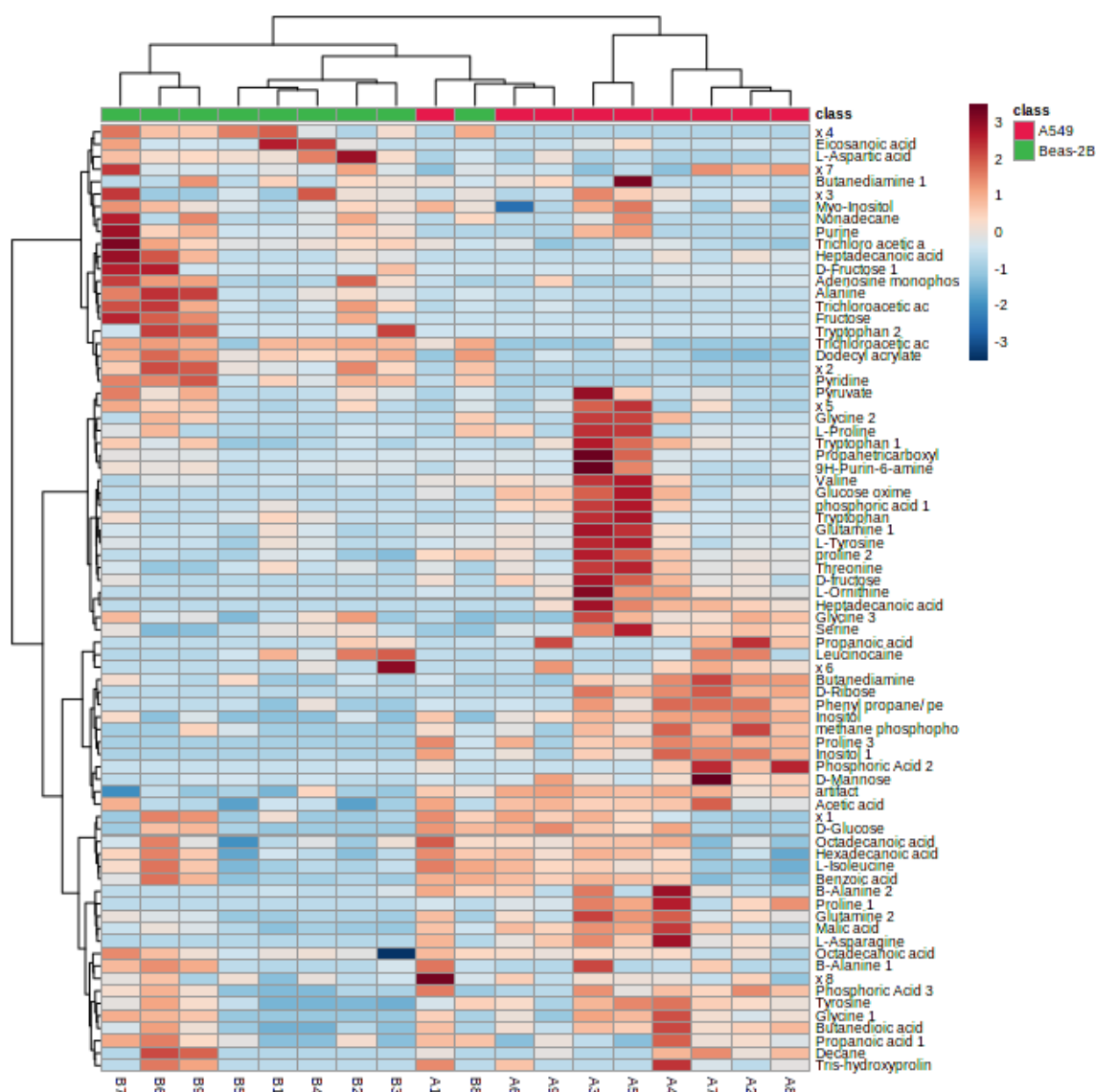


Figure 2-7 Heatmap showing the metabolite profile changes and relationship between the different lung cell lines A549 and Beas-2B.

Again, this trend was reflected in the more specific heat-map including and arranged by pancreatic cell lines only. Though the separation was better than previous there was still more overlapping of cell types. The heatmap divided first into two major branches, the left hand most of which held half of the MIA PaCa-2 cell lines only. The second arm further divided into two, with the left splitting again into two which held, from left, three Panc-1 lines, and right all of the healthy HDPE lines. The final arm then split into three with the first holding three MIA PaCa-2 cell lines, middle the remaining seven Panc-1 lines and final the one outstanding MIA PaCa-2 lines.

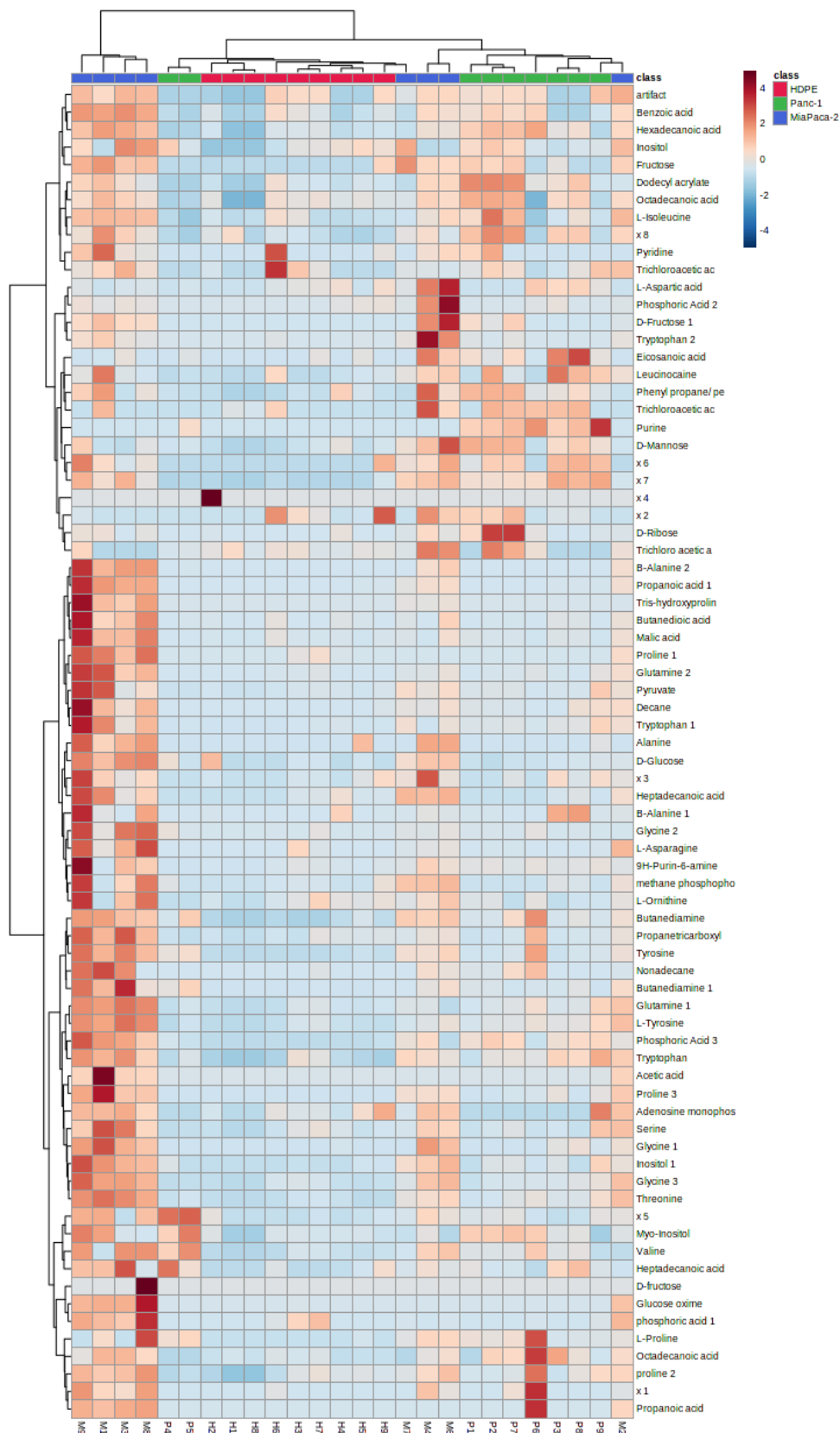


Figure 2-8. Heatmap showing the variance in the significantly different markers between the different pancreatic groups and the impact this has on the inter group relationship

Of the GCMS identified metabolite markers that were identified as having better than “good” ROC scoring and generated tentative biomolecule identifications, forty were found to significantly differentiate more than one of the groups under analysis. Twenty of these were amino acids, Figure 2-9, and the remainder were a mix of metabolite groups including carbohydrates, lipids and organics such as purine, Figure 2-10.

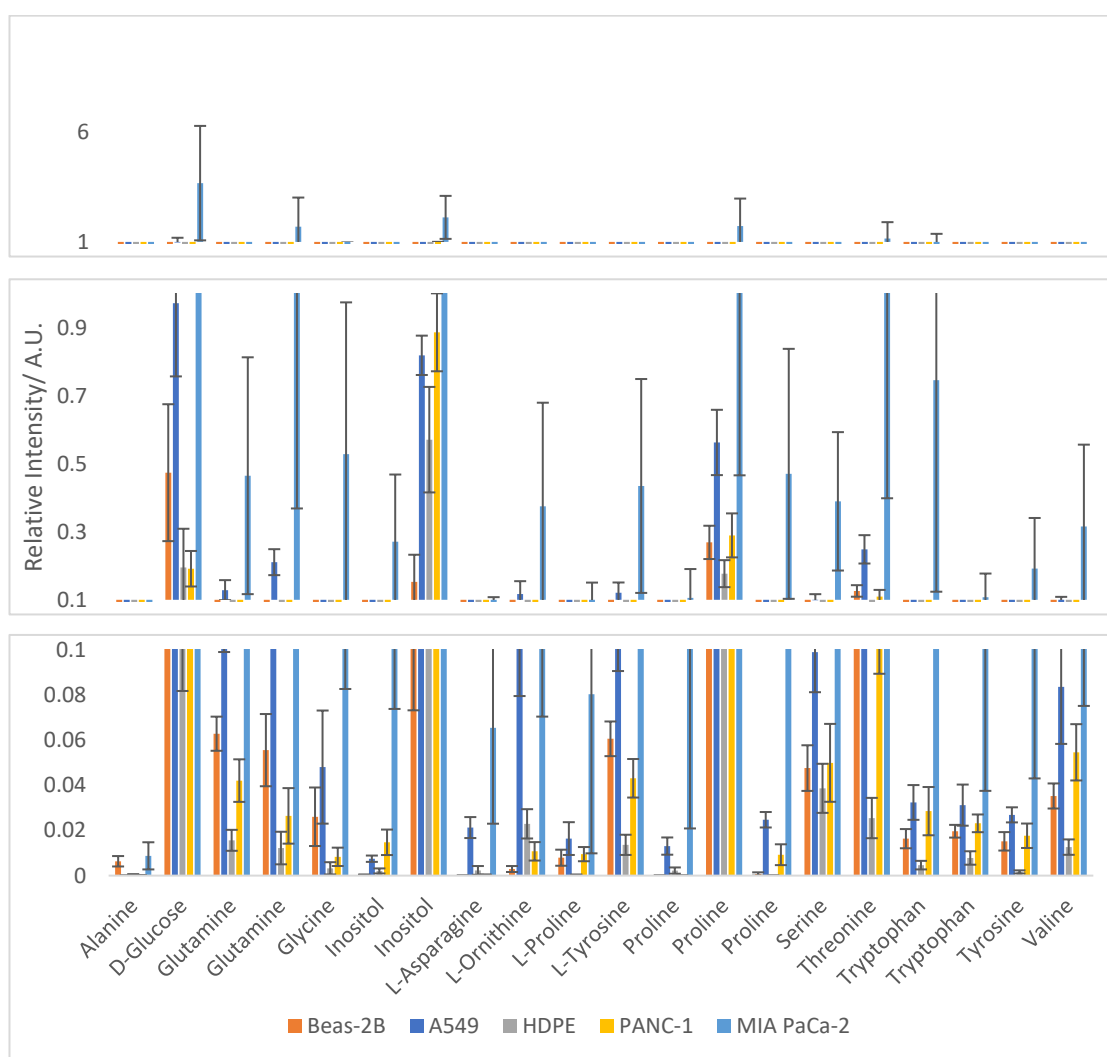


Figure 2-9 Amino acids that were found to be significantly upregulated in more than one of the groups comparison. The standard error bars are plotted, for n=9, to show the variation across the group.

The amino acids generally showed an increase in intensity in the cancerous cell lines in all group comparisons, with the exception of alanine and asparagine in lung cells.

Glucose and glutamine were found to be significantly up-regulated in cancerous cells when compared to their healthy equivalent. As was also the case with the majority of sugar and lipidomic groups analysed. Ribose was found to be significantly increased in both cancerous cell lines.

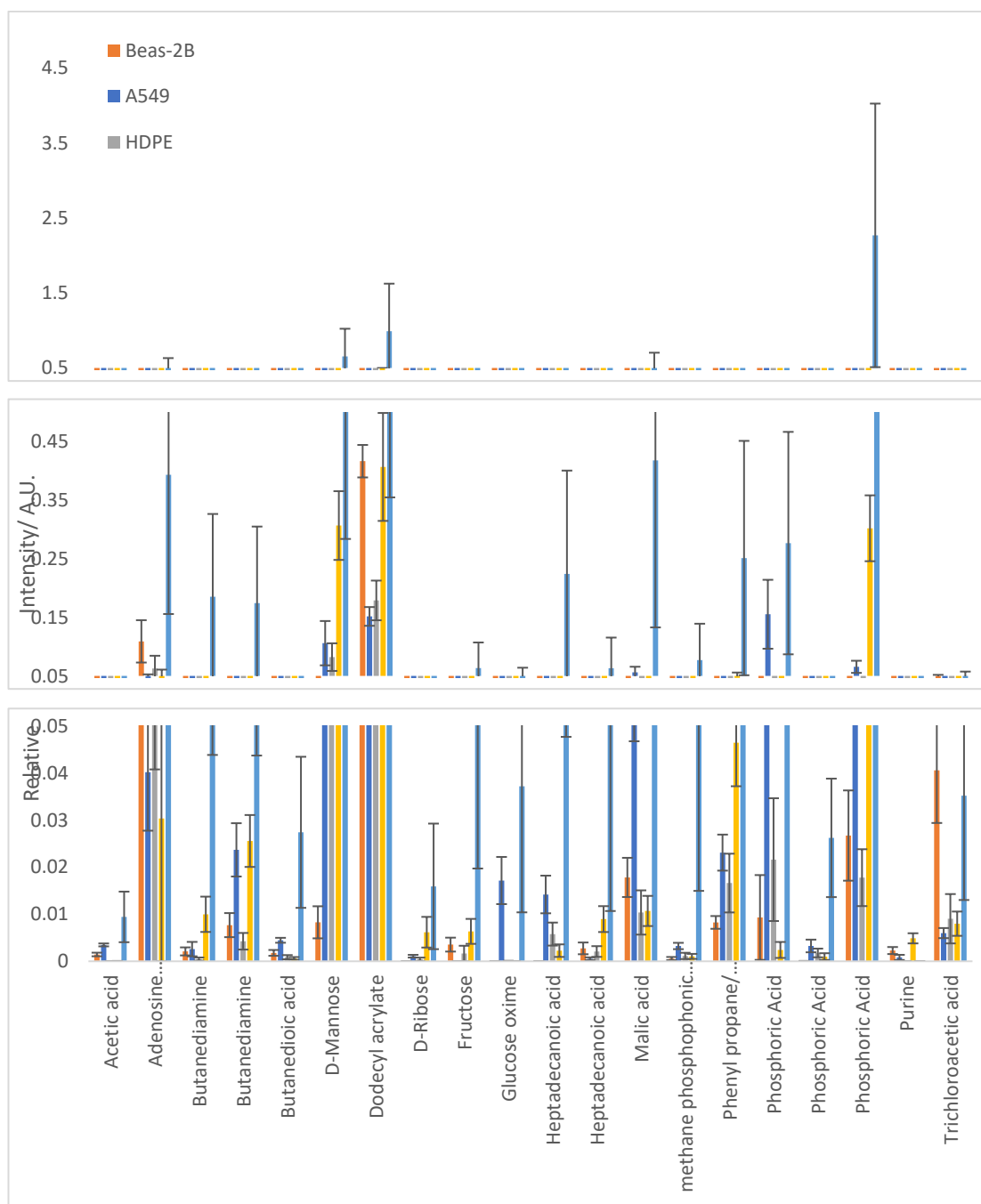


Figure 2-10 Non-amino acid metabolites that were found to be significantly up-regulated in more than one of the groups comparisons. The standard error bars, where n=9, are plotted to illustrate the group variation.

As can be seen in Figure 2-9 and Figure 2-10, the cell line displaying the most common increase in presence of a metabolite is the MIA PaCa-2 cell line, in agreement with its more distinct separation on the heat map generated dendrograms. A similar trend was seen in the “non- amino acid” group of shared metabolites as was seen in the amino acids groups found to be elevated in the cancerous cell lines when compared to their healthy equivalent line. Again, there are a couple of exceptions to this trend. In lung cancer cells the healthy line Beas-2b were found to be up-regulated in adenosine monophosphate, Fructose, heptadecanoic acid, purine and trichloroacetic acid. In the pancreatic lines the healthy line, HDPE, was found to be higher than Panc-1 in adenosine monophosphate, heptadecanoic acid, phosphoric acid and trichloroacetic acid. Mia PaCa-2 always exceeded the healthy cell lines.

#### *2.4.1.1 Targeted amino acid quantitation*

Many of the differences were found to relate to fluctuations in amino acid levels, assessed with tentative identifications. As these reflect only tentative identifications, there is the possibility of misidentifications in the markers. Through next conducting a targeted amino acid analysis, running a series of amino acid standards, it is possible to assess amino acid fluctuations with more certainty as well as the provision of quantitative detail.

Amino acids are important targets for metabolomic studies through their use diverse use as structural protein units and use as energy sources. Alterations during malignancies can illustrate alterations in protein synthesis/ utilisation typically accountable of enhanced proliferation. It is for this reason that alterations in amino acid profiles have been found to be indicative of many other disorders including other cancers (Wei, Z., et al., 2021), Liver disease (Tajiri, K., & Shimizu, Y. 2013) and Alzheimer's disease (Piubelli, L., et al., 2021).

In order to quantify the amino acids present in the samples, standard curves were prepared and derivatised using the same protocol as the metabolite derivatisation and then analysed via GCMS as described in the methodology section. The resulting standard curves for the amino acids analysed are shown in Figure 2-11, with all amino acid standard curves demonstrating a robust linear relationship between normalised peak area and concentration.

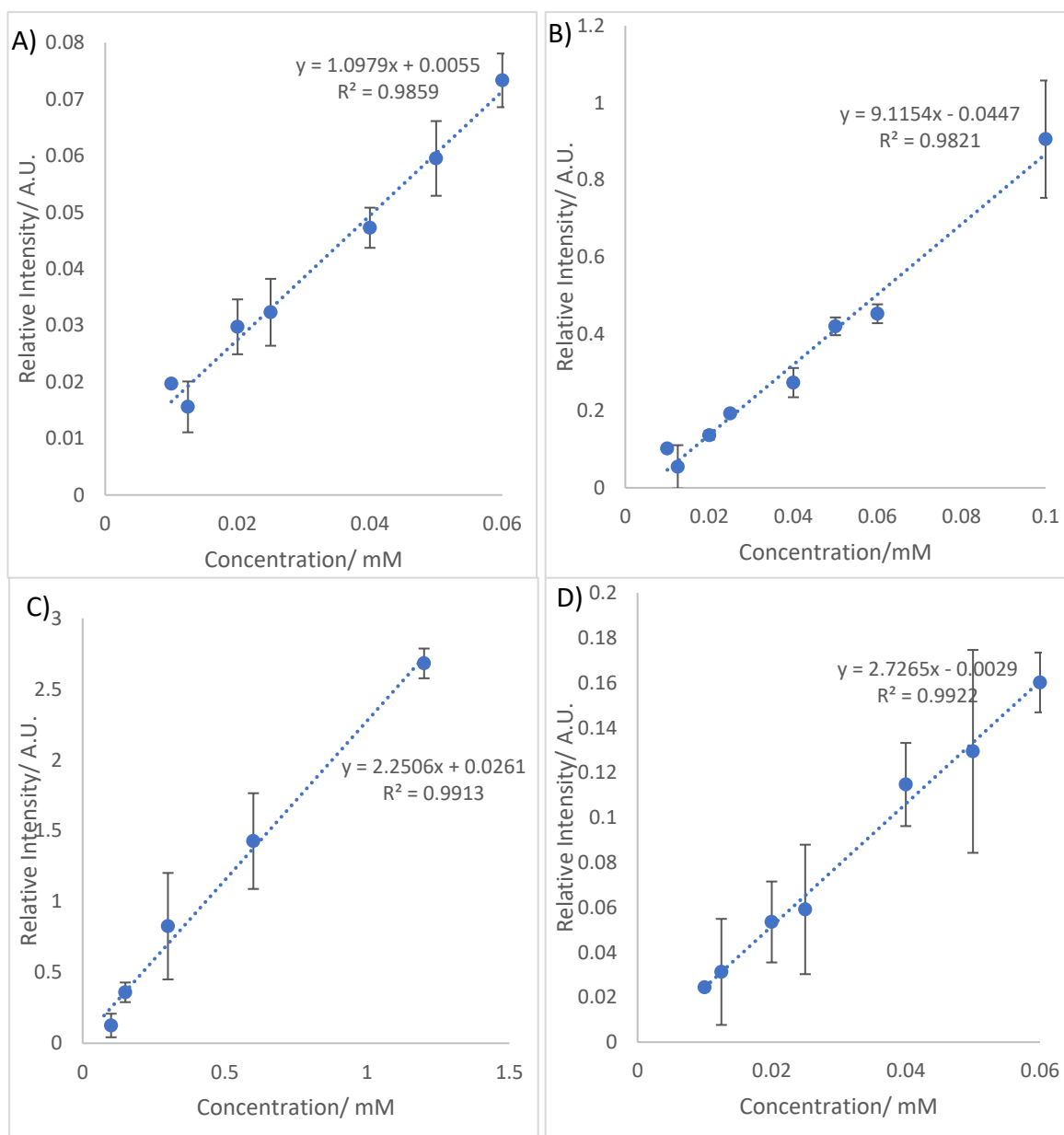


Figure 2-11 Example standard curves generated for amino acid standards a) threonine, b) aspartic acid, c) histidine and d) cysteine with  $R^2$  values and equations of a straight line used for latter calculations shown on the graph that were used for evaluation and further calculations. Standard deviations are also shown for  $n=3$ .

All of details referring to the calculative concentration capabilities for the amino acids that were utilised are shown in Table 2-7. All amino acids measured in sample sets, bar Threonine and Tryptophan fell within the dynamic ranges of these graphs. In these instances, the variance from the dynamic range was minimal. And should not impact the concentration determination in a strong enough way to impact upon significance evaluations between repeated data sets. Additionally, it should be noted



that some of the CV's generated are quite large, which can be indicative of technical variation rather than one of biological. Where this has arisen, more scrutiny is taken when considering the metabolic markers to ensure the significant variation seen is beyond the variation in CV. For instance, the marker for glutamate had a CV of 22.6%, where significant variation was seen between healthy and cancerous markers, the strong change seen in MIA PaCa-2 vastly exceeds the change in CV and therefore can confidently be attributed to a biological fluctuation. It should also be noted that the majority of the larger CV's reported occur at the higher end of the dynamic range (corresponding to large standard deviations at higher concentrations). This can also be considered when scrutinising the markers' ability applying more caution to higher end markers.

Table 2-7-Table showing the average percentage CV, R<sup>2</sup>, Linear equation and dynamic range of each

Amino Acid	Average CV/ %	R <sup>2</sup>	Linear equation	Dynamic Range/ mM
Leucine	28.35	0.9436	Y=0.7446X-0.0019	0.01-0.1
Isoleucine	18.09	0.9454	Y=10.72X-0.2345	0.02-0.3
Threonine	18.16	0.959	Y=1.5125X-0.0054	0.01-0.1
Aspartate	18.85	0.9821	Y=9.1154X-0.0447	0.01-0.1
Asparagine	24.63	0.9246	Y=0.2336X-0.041	0.02-0.1
Valine	25.79	0.9306	Y=3.3674X-0.0129	0.02-0.3
Glycine	13.07	0.9808	Y=3.5669X-0.051	0.01-0.16
Histidine	31.76	0.9913	Y=2.2506X+0.0261	0.1-1.2
Tyrosine	27.83	0.9452	Y=2.6607X+0.0366	0.0125-0.3
Methionine	28.91	0.9331	Y=1.0481X+0.007	0.01-0.1
Cysteine	29.58	0.9897	Y=1.3126X-0.0144	0.0125-0.06
Glutamate	22.26	0.9611	Y=5.0955X-0.159	0.025-0.15
Glutamine	23.54	0.9798	Y=7.8173X-0.0169	0.01-0.15
Arginine	32.20	0.9645	Y=0.4843X+0.0008	0.01-0.075
Lysine	26.29	0.9753	Y=3.4722X-0.05	0.02-0.3
Tryptophan	26.00	0.9866	Y=2.5251X+0.0019	0.01-0.06

amino acid utilised in this analysis

Univariate analysis was conducted on all of the amino acid features to determine any significantly varying markers between the groups of interest. No FDR correction was applied at this point as there were only 12 markers of interest entered into the analysis. Nine of the markers were found to show significant variation between the groups, Table 2-8.

Table 2-8-ANOVA results for significantly different amino acid markers with corresponding f. value, p. value, False Discovery rate (FDR) and Post-hoc analysis conducted using Fisher's LSD.

Amino acid	f.value	p.value	False discovery rates (FDR)	Post-hoc tests Fisher's LSD
Aspartic Acid	17.706	3.35E-08	4.02E-07	MiaPaCa-2 - A549; MiaPaCa-2 - Beas-2B; MiaPaCa-2 - HDPE; MiaPaCa-2 - Panc-1
Threonine	14.241	3.97E-07	2.38E-06	Beas-2B - A549; MiaPaCa-2 - A549; Beas-2B - HDPE; MiaPaCa-2 - Beas-2B; Beas-2B - Panc-1; MiaPaCa-2 - HDPE; MiaPaCa-2 - Panc-1
Tyrosine	8.313	6.88E-05	0.000208	MiaPaCa-2 - A549; Beas-2B - HDPE; MiaPaCa-2 - Beas-2B; MiaPaCa-2 - HDPE; MiaPaCa-2 - Panc-1
Glycine	8.3068	6.93E-05	0.000208	MiaPaCa-2 - A549; MiaPaCa-2 - Beas-2B; MiaPaCa-2 - HDPE; MiaPaCa-2 - Panc-1
Glutamate	7.2368	0.000207	0.000498	MiaPaCa-2 - A549; Beas-2B - HDPE; MiaPaCa-2 - Beas-2B; Beas-2B - Panc-1; MiaPaCa-2 - HDPE; MiaPaCa-2 - Panc-1
Methionine	5.1314	0.002164	0.004327	Beas-2B - A549; MiaPaCa-2 - A549; Beas-2B - HDPE; Beas-2B - Panc-1; MiaPaCa-2 - HDPE; MiaPaCa-2 - Panc-1
Cysteine	4.7793	0.003288	0.005636	A549 - Panc-1; MiaPaCa-2 - Beas-2B; MiaPaCa-2 - HDPE; MiaPaCa-2 - Panc-1
Tryptophan	4.4872	0.00468	0.00702	Beas-2B - HDPE; Beas-2B - Panc-1; MiaPaCa-2 - HDPE; MiaPaCa-2 - Panc-1
Arginine	4.1538	0.007049	0.009399	HDPE - A549; HDPE - Beas-2B; HDPE - MiaPaCa-2; HDPE - Panc-1

Heatmaps were used to analyse how the presence of the targeted amino acids varied in different cell lines and how these different fluctuations related the different groups. The heatmaps depicted poor group differentiation of the lung cells as these were more sparsely set about the map, whereas the pancreatic cell lines showed closer grouping with 78% of the MIA PaCa-2 coming off in the first left hand major arm. The HDPE and Panc-1 cell lines were then found in two separate clusters on the next arm split mostly over two sub arms, . When considering the group averages, the Beas-2b were found to be the most similar to the MiaPaCa-2 cell line with the Panc-

1 and A549 being the next most similar groups with HDPE being the most different. These relationships show how variable the different cancer groups can be with relation to amino acid demands.

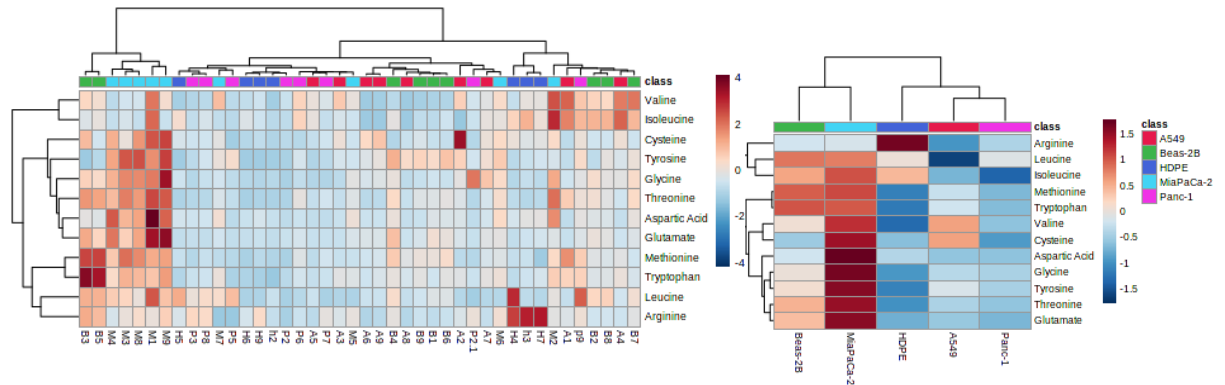


Figure 2-12 PCA evaluation of amino acid relationships between different cell lines

When comparing significant differences between the lung cell lines Leucine, Threonine, Methionine, Aspartic acid, and Glutamate showed significant variance with all amino acids decreasing in concentration in the cancerous cell lines when compared to the equivalent healthy cell line. HDPE showed significant variance with Panc-1 lines for Threonine and Tyrosine whereby the concentration increased in each instance. HDPE and MIA PaCa-2 showed significant variance for the majority of the amino acids analysed including, Valine, Glycine, Threonine, Methionine, Aspartic Acid, Cysteine, Glutamate, Tyrosine and Tryptophan; being significantly more concentrated in each instance. When looking at both cancerous and non-cancerous groups combined only Tyrosine showed significant up-regulation in the cancerous group when compared to the non- cancerous group. Though this is due to the large impact the Mia PaCa-2 had as the other groups showed a decrease. In all of the comparisons made valine was always up regulated in the cancerous form of the cell type and conversely, arginine was always found to be less prominent in the cancerous cell line when compared to healthy equivalent, Figure 2-13.

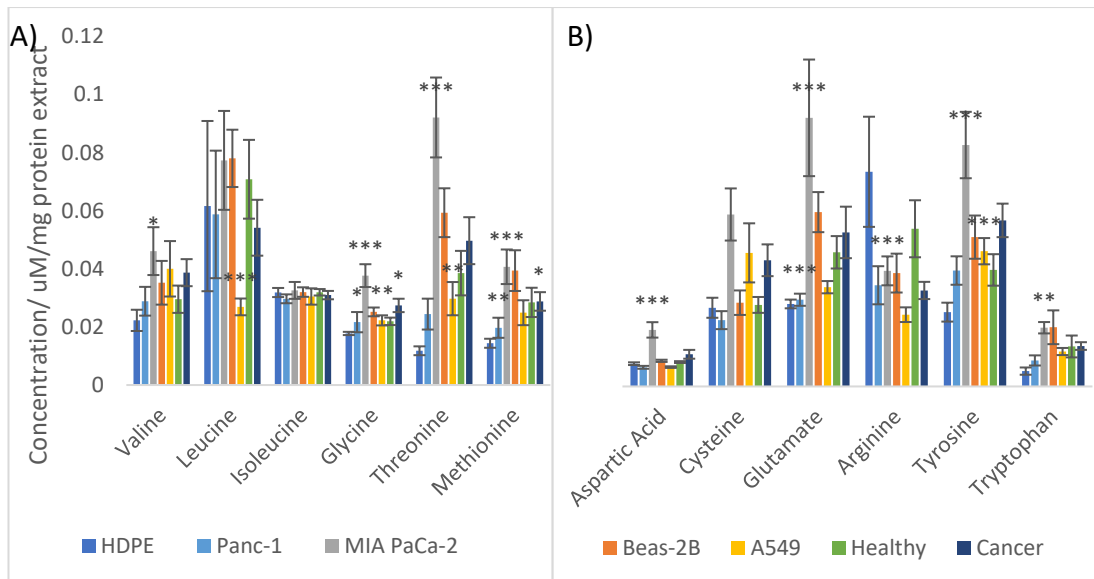


Figure 2-13 Amino acids in lung cancer and pancreatic cancer cell lines arranged according to whether they are essential A) or non-essential B) in nature. Significant changes were made between each mutant group and the healthy groups values with bars noted with \* $>0.05$  \*\* $>0.025$  \*\*\*  $>0.005$

Looking at the percentage change of the various cancerous cell lines, Figure 2-14, when compared to their healthy equivalent a more simplistic picture is shown with the majority of amino acids being up-regulated in the cancerous cell lines, almost all in MIA PaCa-2 and most are down-regulated in the lung cancer cell lines.

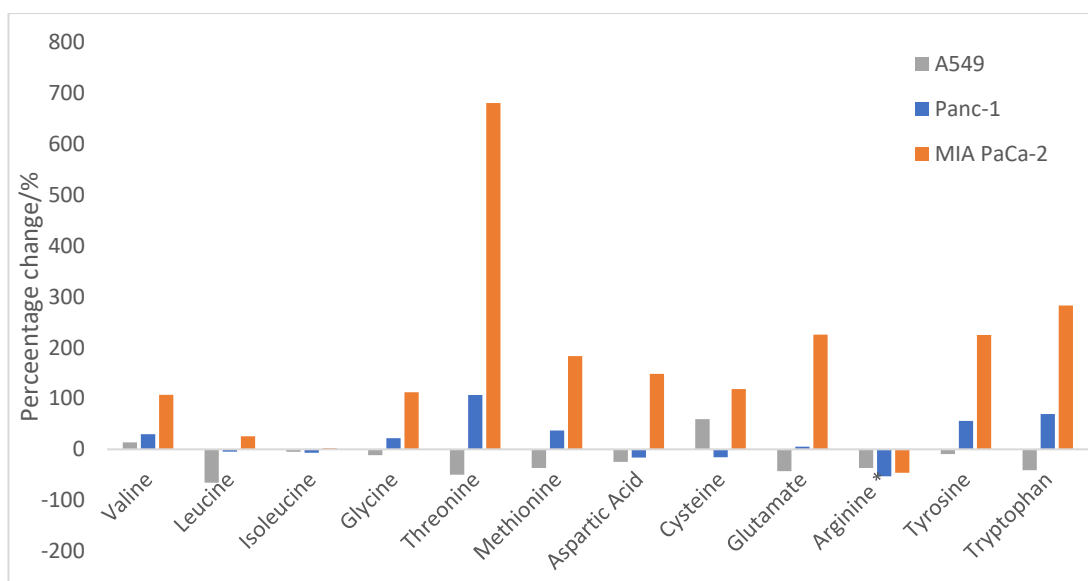


Figure 2-14 Percentage change of quantified amino acids in cancerous cell lines when compared to their healthy equivalent cell line.

Commonly, as noted in the untargeted analysis, the MIA PaCa cell line generated the greatest changes in amino acid concentrations. Interestingly, the lung cancer cell line mainly demonstrated an opposite change with a decrease in amino acid levels. Of potential note is cysteine, which in this analysis exhibited elevation in the lung cancer A549 cell line, but only one of the two pancreatic cancer cell lines.

### 2.43.3. MALDI Analysis

#### 2.4.3.1. Negative Lipidomic analysis

The first analysis undertaken was a lipidomic analysis in negative ionisation mode using MALDI ToF mass spectrometry for which 9AA was selected as a matrix for the analysis of free lipids and phospholipids present in the cell culture and was successfully used to quantify multiple phospholipids. Figure 2-19, below shows an example spectra of a pancreatic cancer and healthy cell line extract, which indicate a reasonably similar spectra of lipid species.

Very similar profiles were obtained for both lung and pancreatic cell lines at first appearance with similar conservations relating to dominant peaks- in both cases

760.6 is the most dominant, with 782.6, 786.6 and 732.6 representing the next most abundant species present. Some peaks are more unique/ prominent in each of the cell types of interest. 517.6- pancreas and 482.4, 504.4, 770.6 were all much more prominent in lung cancer cell lines. A number of significant differences were identified that were capable of distinguishing each of the cells from each other as well as offering suggestions for underlying metabolic alterations that present in cancers.

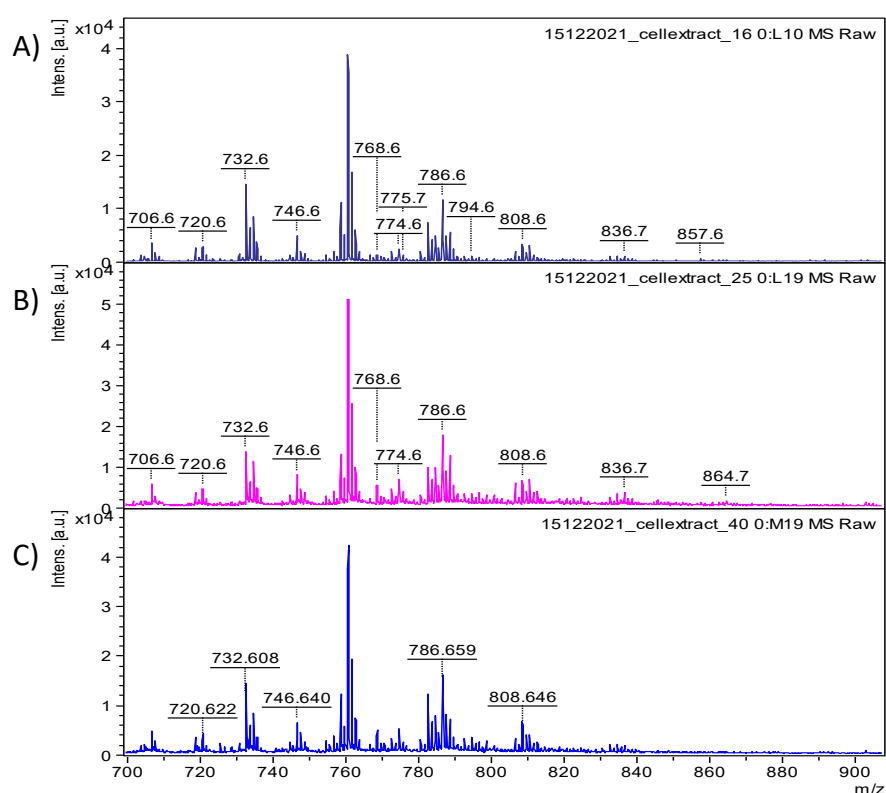


Figure 2-19- Example spectra of Pancreatic cells analysed in negative mode using MALDI analysis with 9aa matrix. A) HDPE (Healthy) dark blue , B) Panc-1 (cancerous) purple, C) Mia PaCa-2 royal blue lower.

Multivariate analysis was conducted using metaboanalyst whereby unsupervised PCA analysis was conducted to look for any associations between the different groups based upon the top 20 markers observed. PCA analysis on these shared

significantly different lipid markers generated a very good degree of separation as can be seen by the five different distinct group clusters shown in the PCA.

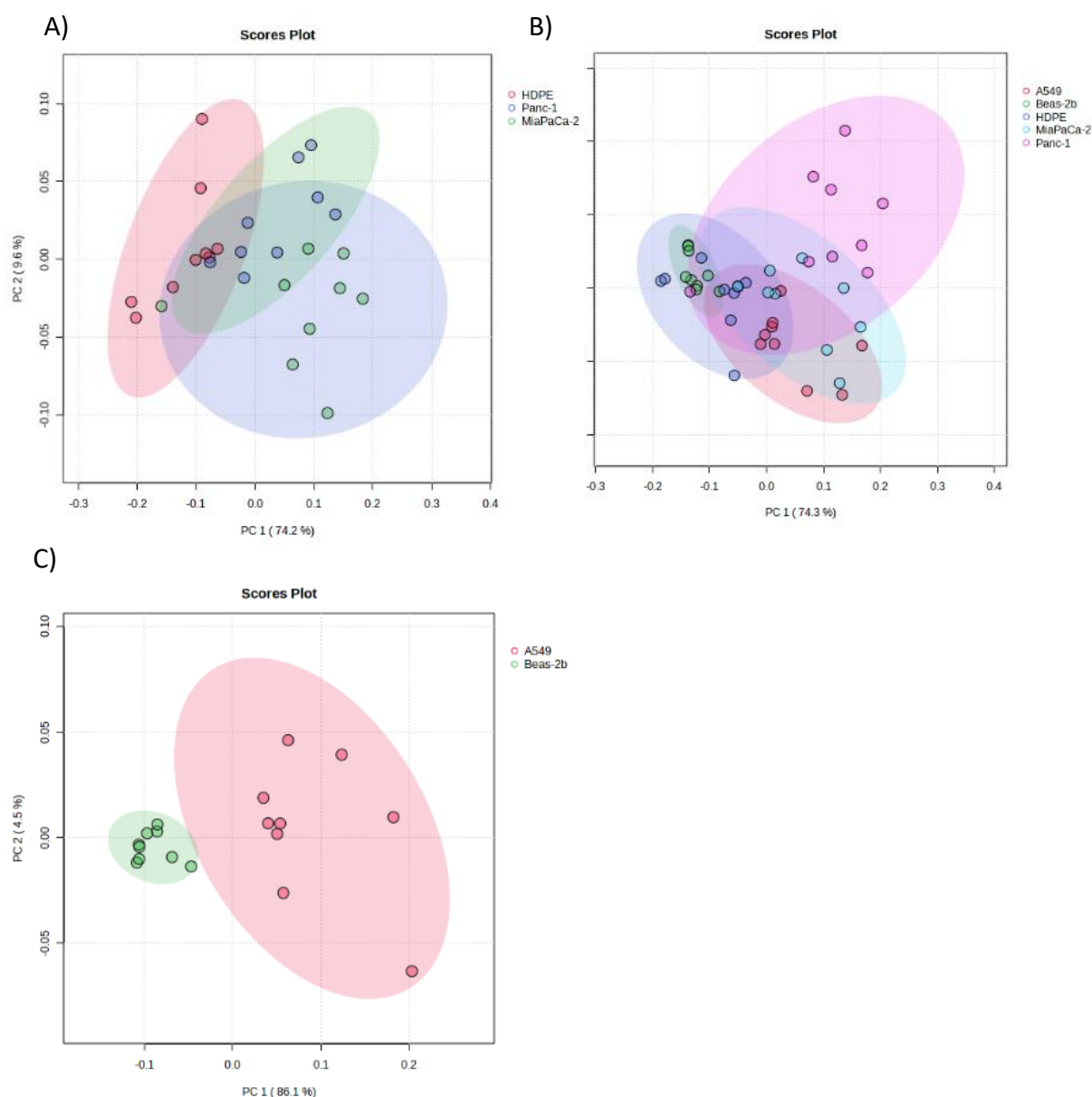


Figure 2-20- PCA analysis of the 5 different cell line groups analysed in negative MALDI analysis when assessed using the 20 common significantly different metabolite markers. A) shows the pancreatic cell lines, B) The lung cell lines and C) all cell lines in the analysis.

The healthy lines have a tendency to the left of the plot, with Beas-2b having a tighter cluster than HDPE. The cancerous cell lines have a tendency to be more sparse of occupy the right hand side of the chart.

Further evaluation of the separation through use of heatmaps shows another very good degree of separation with a clear discrimination in the first big branch of the



dendrogram with the left hand arm holding all of the healthy samples plus two cancerous ones. The second layer of separation, working from left to right, separates the Beas-2b healthy cells from the HDPE healthy pancreatic cells with the exception of two HDPE samples and one Beas-2b. The other arm divides into two further arms, the first holding all of the A549 cancerous lung cells and three MIA PaCa-2 cancerous pancreatic cells. The next arm holds all of the pancreatic cancer cells bar 3 and 2 lung cancer cells and splits into a further 2 branches, with one branch holding half of the MIA PaCa-2 samples and the other 70% of the Panc-1.

The average lipid signal intensity for the 5 cell types, again show that elevated lipid levels are found in the cancer cell types compared to the healthy comparison cells, with more subtle differences such as the levels of LPC 18:0 and PG37:3 / PA 42:1 being seen as being altered in levels between the 2 healthy cell lines. This latter level of also allows for the differentiation between the 2 pancreatic cancer cell lines (in which the levels are elevated) compared to the lung cancer cell line analysed (in which it has a lower average level compared to the healthy lung cell line). PE 38:4 also is found solely to be significantly elevated in the pancreatic cancer cell lines hence suggesting that interpretation of lipidomic profiles at a detailed level may allow for both organ of origin distinction as well as cancer location.

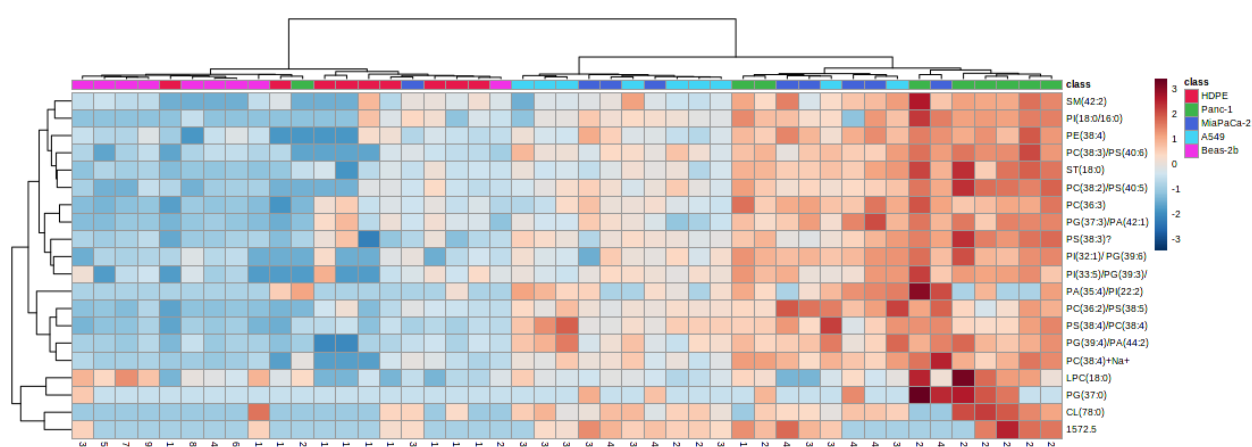


Figure 2-21- Heatmaps illustrating the variation and relationships between all cell lines groups under study for altered lipidomic presence

Determining the diagnostic properties of the markers found to distinguish the groups from each other was then conducted

### Pancreatic cell line comparisons

Comparisons were made separately between the healthy pancreatic cell line and the Panc-1 cancerous cell line. There were one hundred and nineteen significant differences found to vary between healthy pancreatic cell lines (HDPE) and pancreatic cancer cell line Panc-1. After Bayesian corrections were applied and ROC analysis conducted forty-one differences showed better than “good” potential as individual biomarker candidates (Figure 2-22)

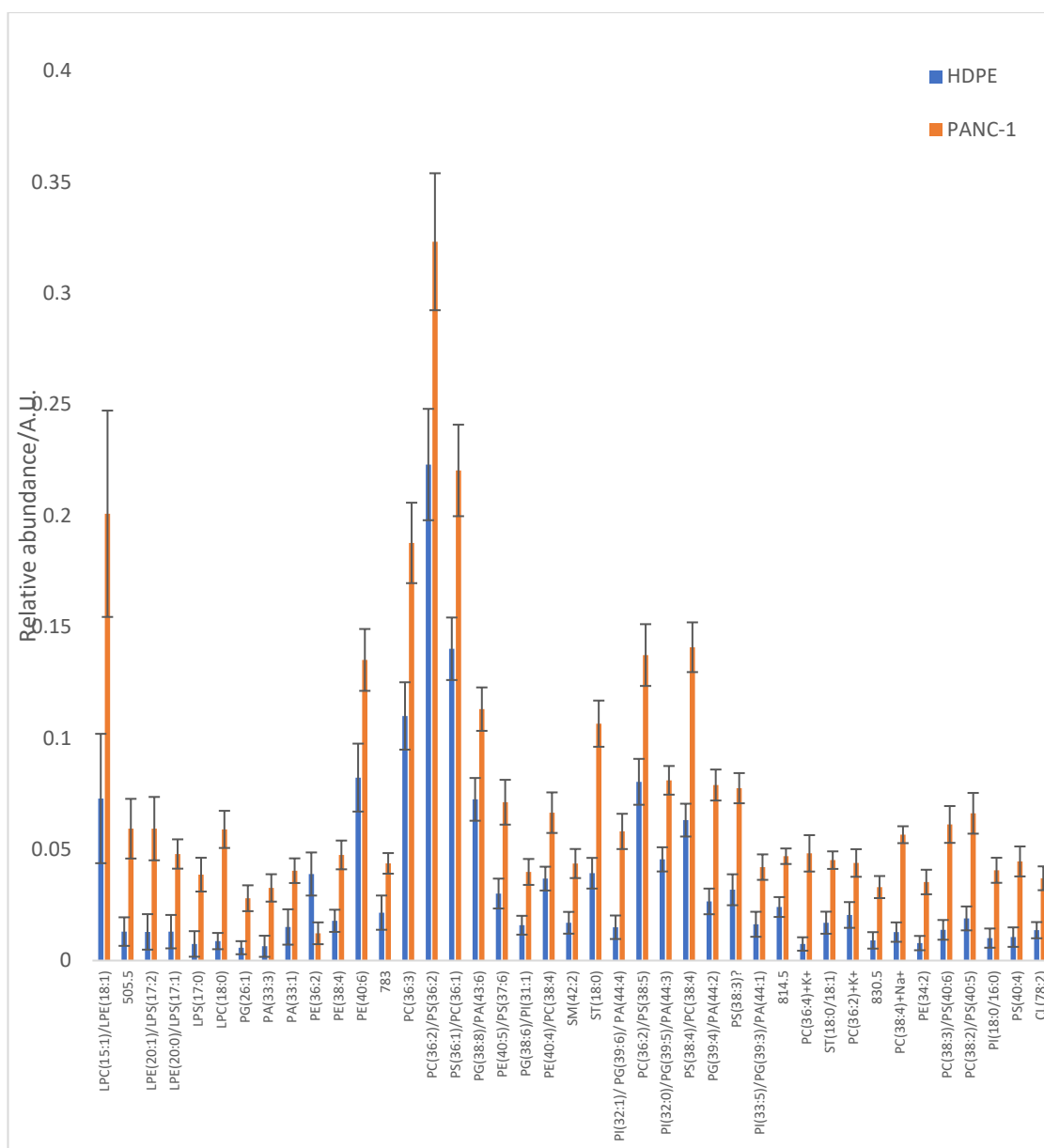


Figure 2-22- Significantly different markers that distinguished Panc-1 and HDPE cell lines analysed in MALDI using negative mode analysis.

There were ninety-six significant markers found to distinguish MIA PaCa-2 cells from HDPE cells with sixteen achieving better than good when discriminated as unique biomarker candidates (Figure 2-23)

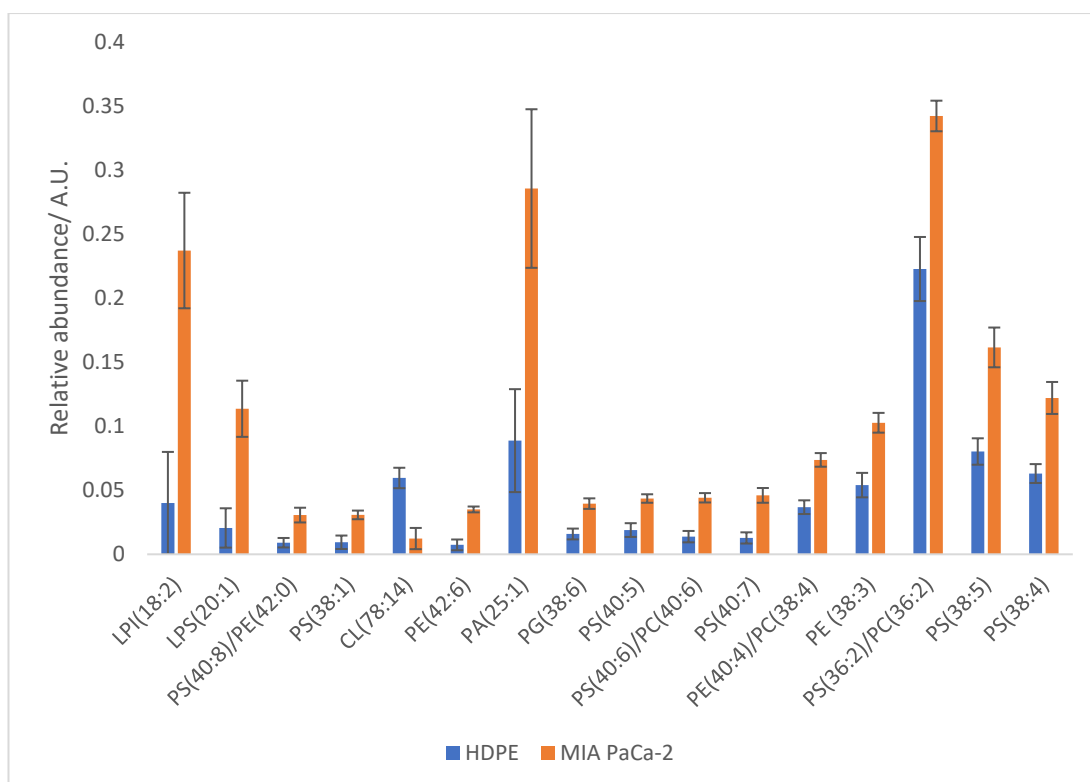


Figure 2-23-Significantly different markers that were found to be better than good diagnostically that distinguish MIA PaCa-2 from HDPE in MALDI operated in negative mode

### Lung cell line comparisons

Two hundred and fifteen different peaks were found to significantly vary between healthy and cancerous lung cancer cell lines under MALDI analysis in negative mode. After Bayesian correction and ROC curve selection to evaluate the markers effectiveness sixty-seven markers were found to have better than “good” diagnostic attributes, with twelve “excellent”, Figure 2-24. Tentative identifications for markers of interest were generated based on previously published literature and lipidmaps, these identifications are not absolute as no fragmentation detail/ standards were ran (Berry, K. A., et al.,2011; Pirman, D. A., et al., 2013 & Jackson, S. N., 2014).

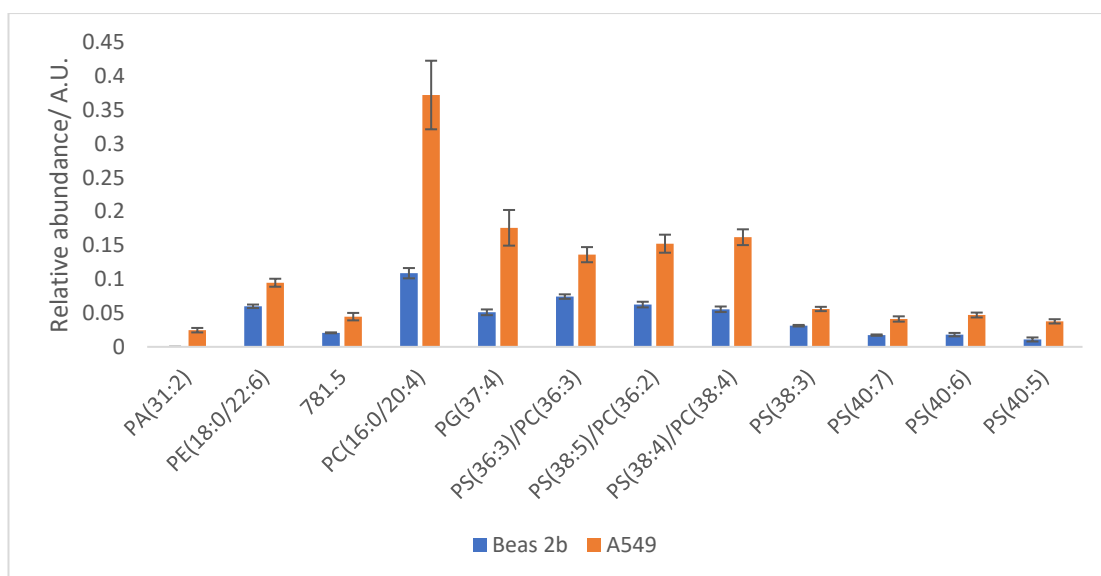


Figure 2-24- Significantly different markers that distinguish Beas-2b and A549 cell lines from each other with a better than excellent biomarker rating.

### Comparing all cell types

Twenty differences were found to be consistently significantly different across all comparisons made, these twenty common lipids, Figure 2-25, were tentatively classified as mostly being PC, PA, PI, PE and were entered into a further trend analysis.

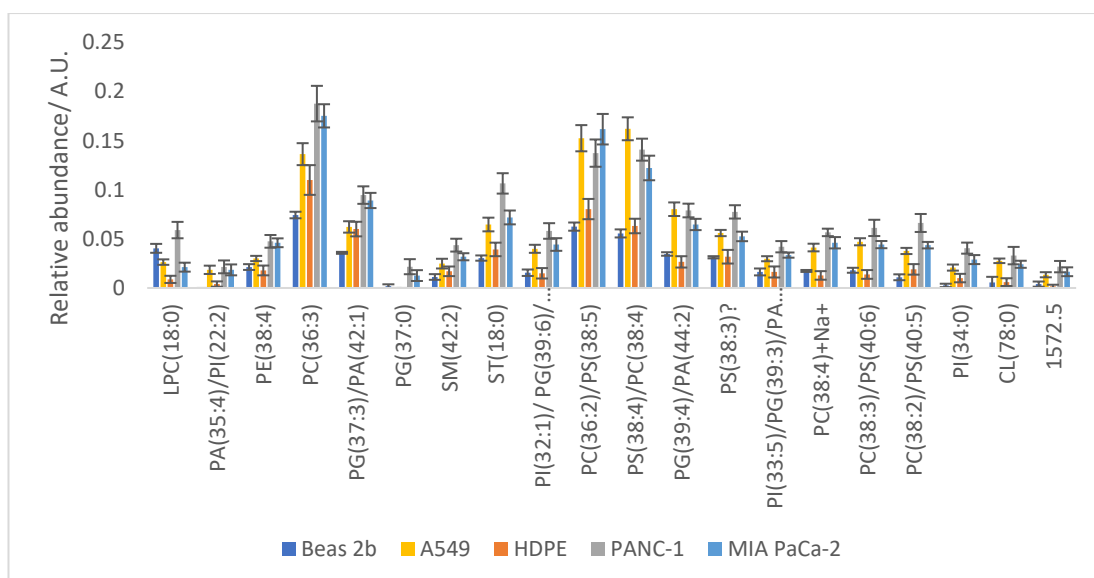


Figure 2-25- Common significantly different metabolites found to distinguish all of the groups of interest from each other.

All markers were found to be significantly up-regulated in cancerous lines compared to non-cancerous with the exception of LPC(18:0) which were down-regulated in both types of cancerous lines compared to healthy equivalents. PG(37:0) which was down-regulated in lung but up-regulated in pancreatic

#### *2.4.3.2. Positive Lipidomic and peptidomic analysis*

Extracted cell profiles were additionally analysed in positive mode using CHCA as a matrix to compare the profiles of healthy and cancerous cell lines, this time focusing on fatty acids, glycerophospholipids and peptides as it was anticipated that CHCA would allow sufficient coverage of all three of these biomolecule types. The CHCA matrix has a strong affinity for metal cation, in particular  $K^+$ , formation which was taken into account when assigning tentative identifications where it was possible to do so. Identifications were considered using previously published literature for detected metabolites under similar conditions as well as utilising lipidmaps and HMDB (Angerer, et al., 2022; Guo, S., et al., 2014; Hong, J. H., et al., 2016; Jones, E. E., et al., 2014; Kaya, I., et al., 2018; Fernández, R., et al., 2014; Jiang, L., et al., 2015; Desbenoit, N., 2018; Dória, M. L., et al., 2016; Chagovets, V. V., et al., 2020; Sommella, E et al., 2022). These identifications are not certain as no fragmentation profiles were generated as this was not the main goal of the analysis. Eight hundred and two different markers were entered into the analysis with significance comparisons being made based upon for relative intensity changes.

PCA analysis was performed using metaboanalyst to visually show the metabolic profile changes and group relationships between the cancerous and non-cancerous groups under analysis. The common significantly different metabolites that were found to distinguish each of the cancerous cell lines from their equivalent healthy counterpart were entered into this analysis. Both the clustering of replicate samples and spread of different groups were considered when analysing the potential of these markers in this way. Generally tighter clustering was seen in the healthy groups than the cancerous groups, which is anticipated due to cancer causing variations to healthy process which can fluctuate at different stages in development. A better degree of intra-sample set clustering and inter sample set separation was achieved after normalisation using square root transformation and so this was used for the PCA analysis.

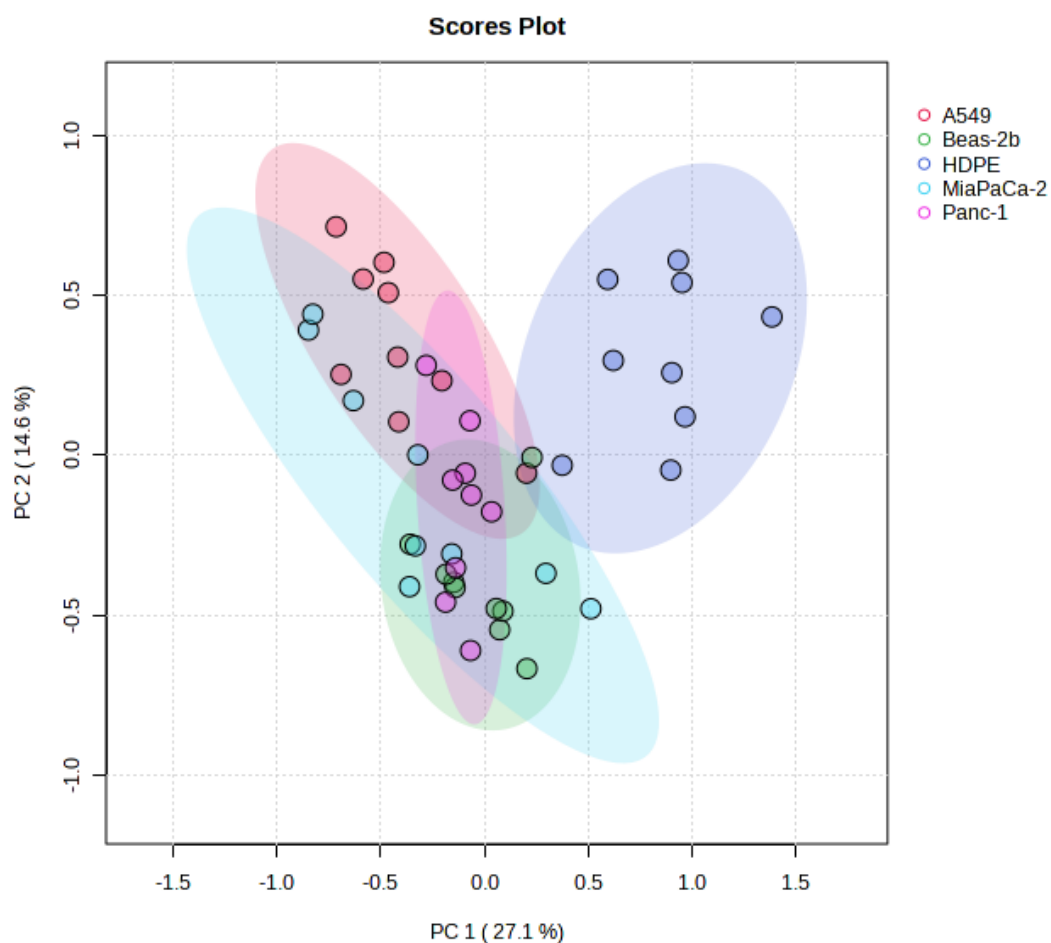


Figure 2-26- PCA plot showing the separation of the 5 different groups under analysis when considering the shared significantly different metabolite marker profiles

Very distinct separations were seen between the healthy and cancerous cell lines in each different group class under comparison, . Complete group discrimination was seen between the pancreatic cancer and healthy cells with a completely distinguished group seen for HDPE.



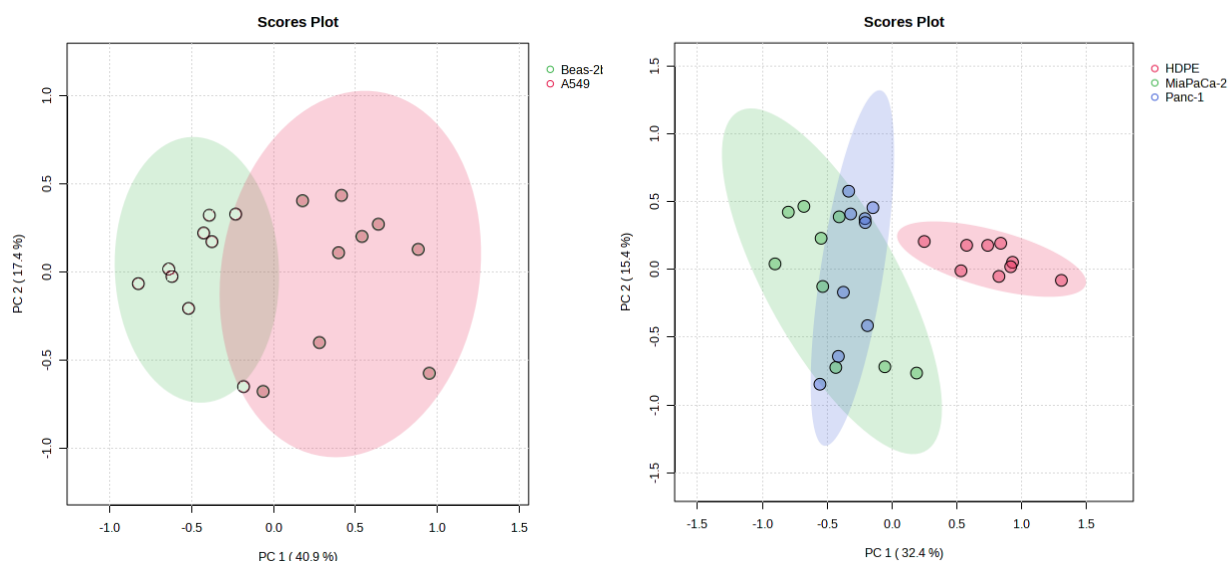


Figure 2-27- PCA plots generated to show the separation of a) Lung cancer and healthy lung cells b) healthy pancreatic cell lines and Panc-1/MiaPaCa-2 with confidence intervals shown with ovals.

Heatmaps were generated using the shared significantly different metabolite markers to further assess the relationship across the different groups under evaluation. The heatmaps were able to generate good group differentiation with all bar one of the HDPE sample set coming off on the first major branch separating from all of the other markers of interest. The other arm of the major branch broke off into two further arms, one consisting of a mix of the majority A549 cancerous lung and MIA PaCa-2 cancerous pancreatic cells and the other arm splitting into two further arms one made up of the majority of Panc-1 cancerous cell lines and the final branch all of the Beas-2B healthy lung cell line with 2 of the MIA PaCa-2 cancerous pancreatic cell lines.

When averaging the groups the profile relationship changes slightly suggesting that HDPE is still the most distinguishable group with MIA PaCa-2 and a Panc-1 being the most similar, followed by A549 and then Beas-2b.

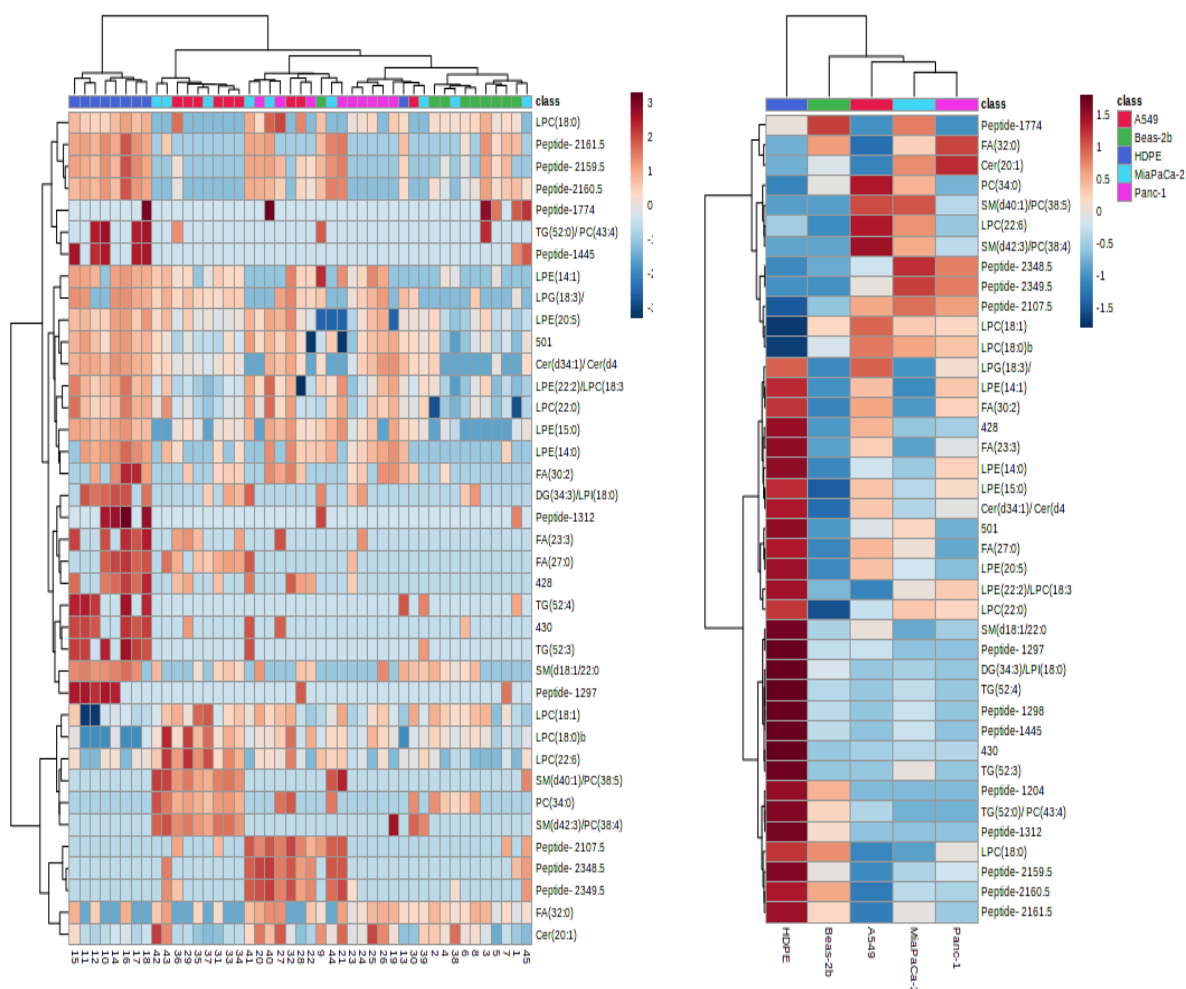


Figure 2-28- Heatmaps generated using shared significantly different metabolites to analyse profile changes and group relationships between the 5 different groups under study.

When combining all of the groups together a good level of separation is still achieved for defining whether the group is cancerous or healthy. The first 12 metabolites shown on are the best at distinguishing the healthy from the cancerous cell lines with all being lower in the healthy cell lines than the cancerous ones, in particular Cer(20:1), PC(34:0), SM(d40:1), LPC(22:6), SM(d42:3), Peptide 23485, Peptide 23495, Peptide 21075, LPC(18:1) and LPC(18:0). HDPE showed a much higher and distinguishing level of all of the lower 27 markers whereas an up-regulation of markers LPC(18:1), LPC(18:0) and Peptide 21075 were distinguishingly higher in all cancerous cell lines and down-regulated in the healthy lines.

## Pancreatic cell line comparisons

### HDPE vs Panc-1

There were seventy-two significantly different metabolite markers found to distinguish HDPE from Panc-1 cell lines in positive mode MALDI analysis, similar to the prior comparison these markers included constituents of all of the different groups under analysis including fatty acids, glycerophospholipids and peptides.

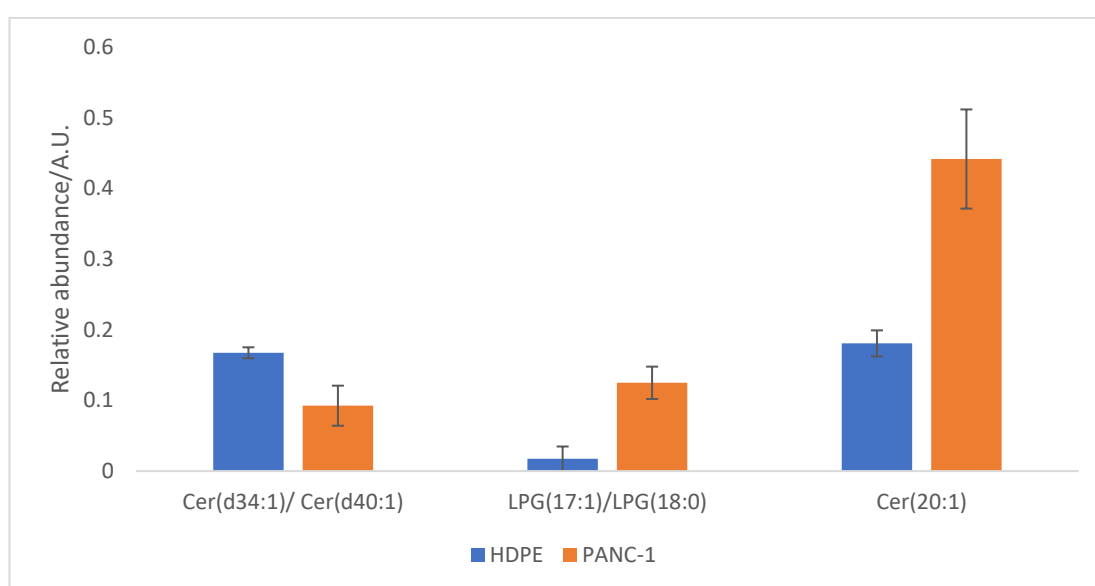


Figure 2-29- Significantly different markers found to distinguish Panc-1 cell lines from HDPE healthy cell lines whilst maintaining a better than "good" biomarker quality under ROC analysis.

This was narrowed to 3 markers that obtained better than "good" diagnostic stats as unique markers, as shown in Figure 2-29, which are also annotated on the chromatogram, Figure 2-30, to illustrate an example of the visual difference of the peak changes observed.

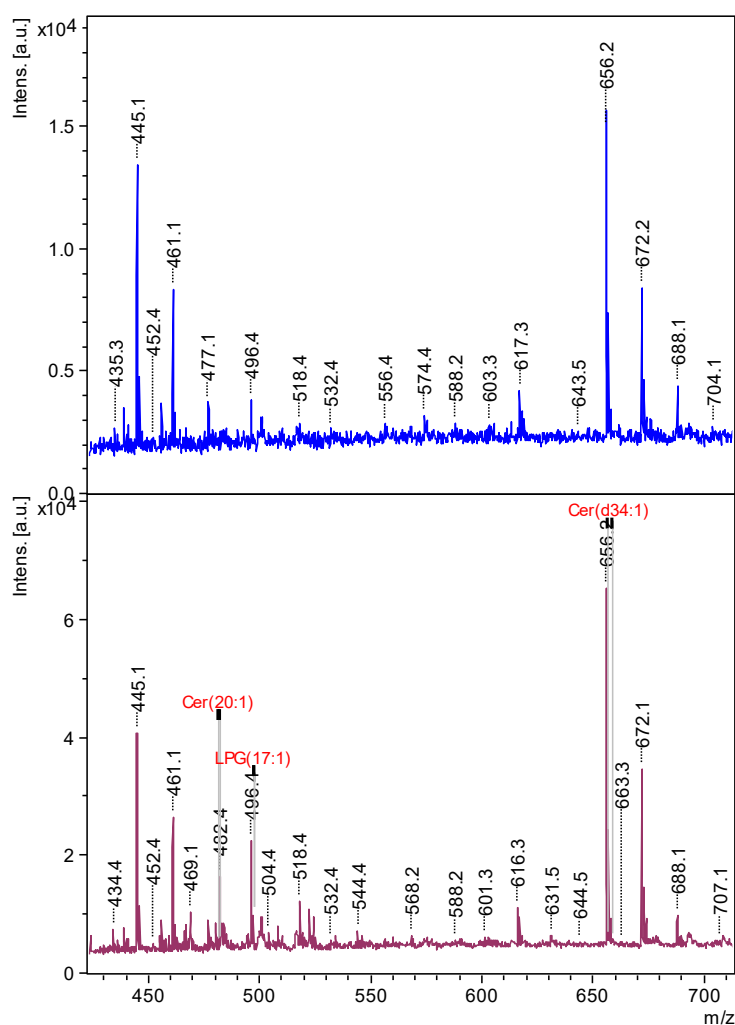


Figure 2-30-Example spectra for HDPE and Mia PaCa-2 cell lines respectively with the three “good” diagnostic markers of interest labelled.

There were eighty-five significantly different variables found to distinguish MIA PaCa-2 cell lines from the healthy HDPE cell lines using MALDI in positive mode across all different groups under evaluation. Only one of the metabolites of interest scored better than “good” in terms of diagnostic potential relating to ROC scoring, Cer(d34:1) as shown in Figure 2-31.

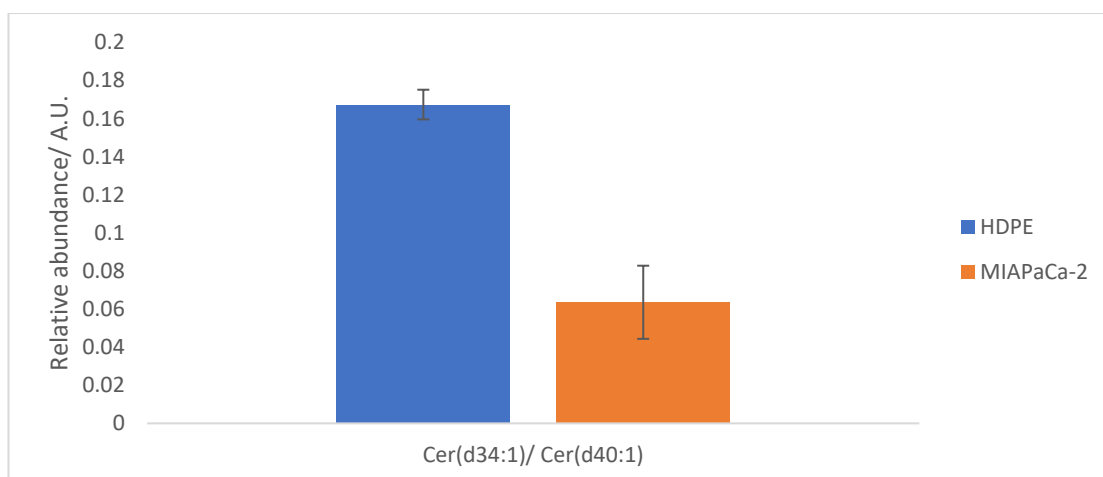


Figure 2-31- Significantly different metabolite marker found to distinguish MiaPaCa-2 cancerous cell lines from HDPE healthy cell lines whilst achieving a better than "good" ROC scoring.

### Lung cell line comparisons

Seventy-seven markers were found to be significantly different between A549 cancerous lung and Beas-2b non-cancerous lung cells. These consisted of fatty acids, lysophospholipids, glycerophospholipids and peptides, Figure 2-33, shows example spectra that cover the key regions found to be of interest, with a number of the significantly different markers that exist between 400 and 800  $m/z$ . After discriminating the biomarker potential of the seventy-seven markers using ROC analysis, 17 markers were selected for performing better than "good", all of which were a fatty acid constituent as shown in, Figure 2-32.

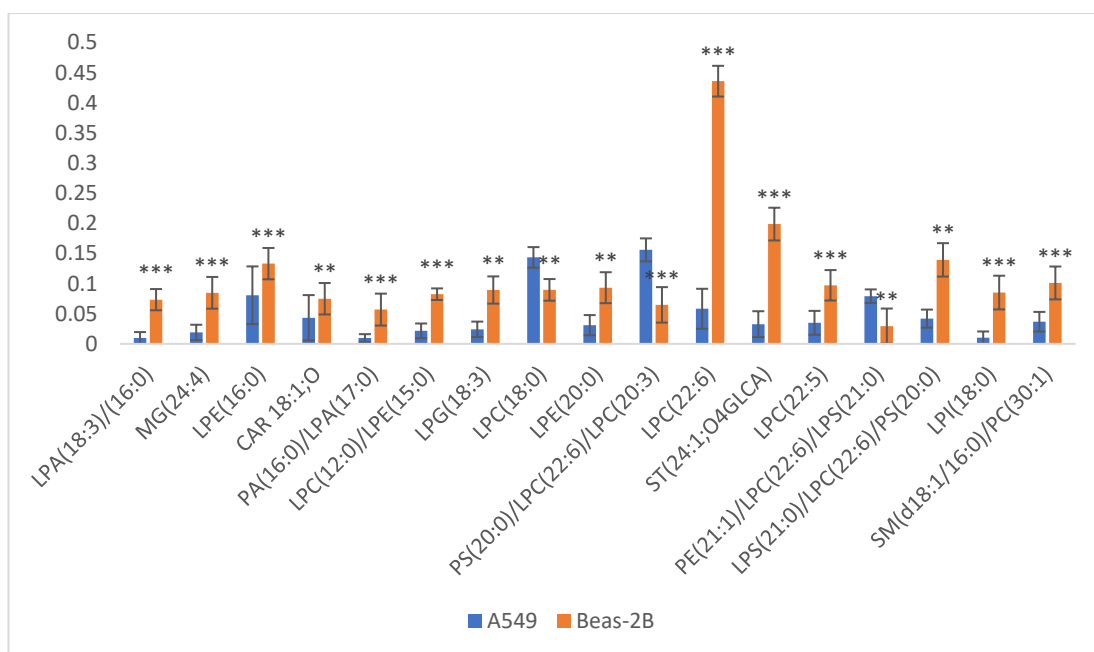


Figure 2-32- Significantly different profile markers obtained using MALDI analysis in positive mode that distinguish cancerous and non-cancerous lung cells from each other whilst achieving better than “good” ROC scoring. Standard error bars are shown where n=9. here multiple identifications are possible these shown. Asterisks for significance are shown, with \* < 0.05 \*\* < 0.025, \*\*\* < 0.01.

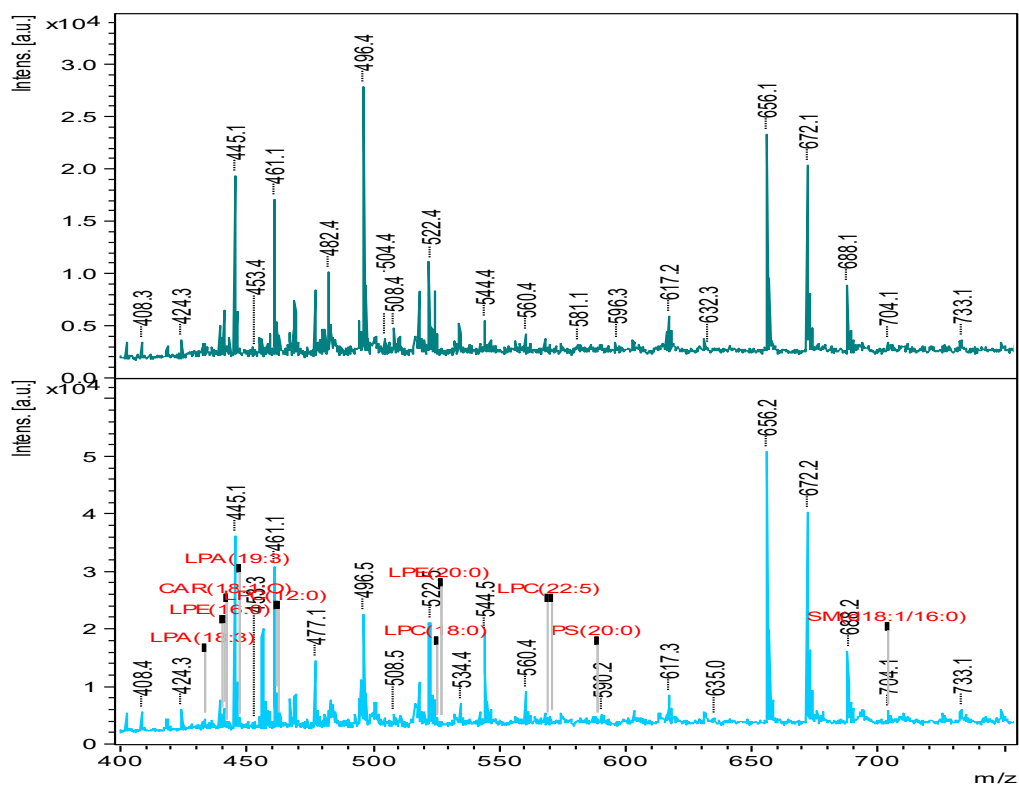


Figure 2-33- Mean intensity comparison of lipid differences found to significantly discriminate lung cancer and healthy cell lines from each other whilst maintaining better than “good” ROC scoring as individual markers.

### Comparing all cell types

Forty significantly different markers were found to be shared across more than one of the cancerous comparisons made and selected for further comparative evaluation in the study. Free fatty acids held a large group of these common markers. All of the fatty acids were found to be lower in intensity in cancerous cell lines than the healthy equivalent with the exception of FA(32:0) which was alternatively found to be up-regulated in cancerous when compared to healthy.

A less prominent group of fatty acids shown to significantly discriminate different cancerous groups from each other consisted of ceramides which were found to be significantly up-regulated in cancerous cell lines when compared to their healthy equivalent, Figure 2-34.

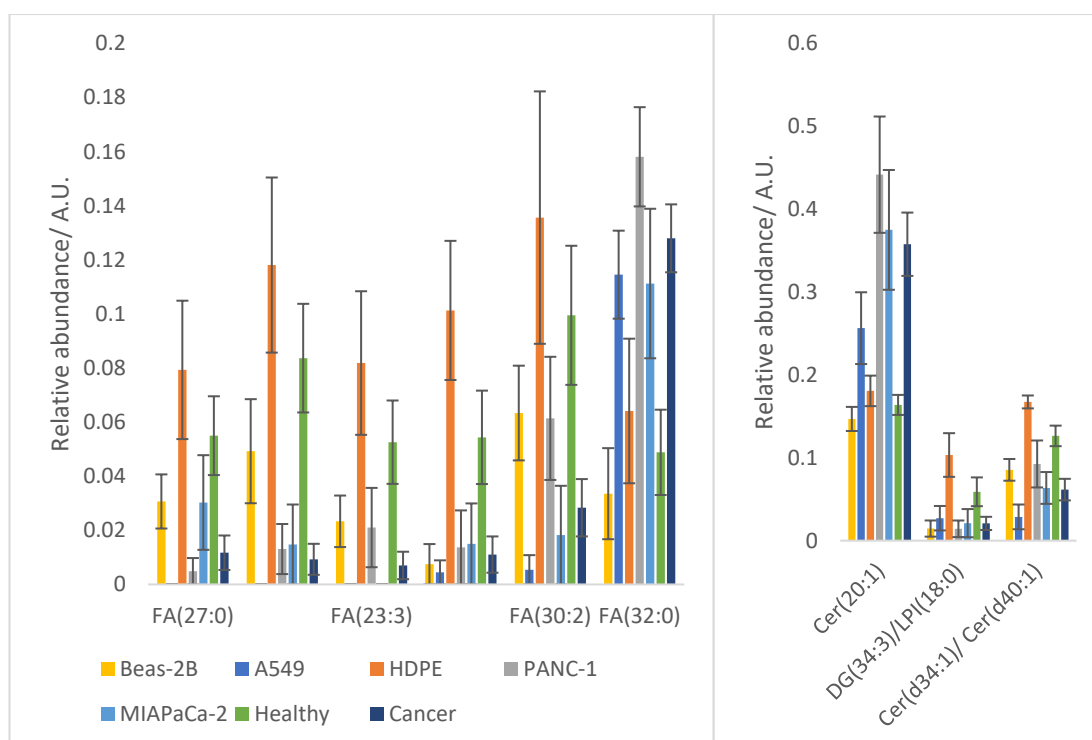


Figure 2-34 A) - shared significantly different fatty acids found to distinguish cancerous and non-cancerous cell lines from each other. B) Significantly different ceramide and DG lipids found to be

significantly different between healthy and cancerous lines across more than one different cell line comparison.

The majority of lysophospholipids were found to be significantly down-regulated in cancerous cell lines when compared to their healthy equivalents with the exception of a selection that were found to be down-regulated in lung cancer but up-regulated in both of the pancreatic cancer cell lines, these include LPC(18:1) (this change was not significant in lung), LPC(18:0) b sig in all, LPC(22:6) and PC(34:0). Additionally, LPE(22:2) was found to be up-regulated in lung cells but upregulated in pancreatic cancer, however the change in lung cancer was not found to be significant.



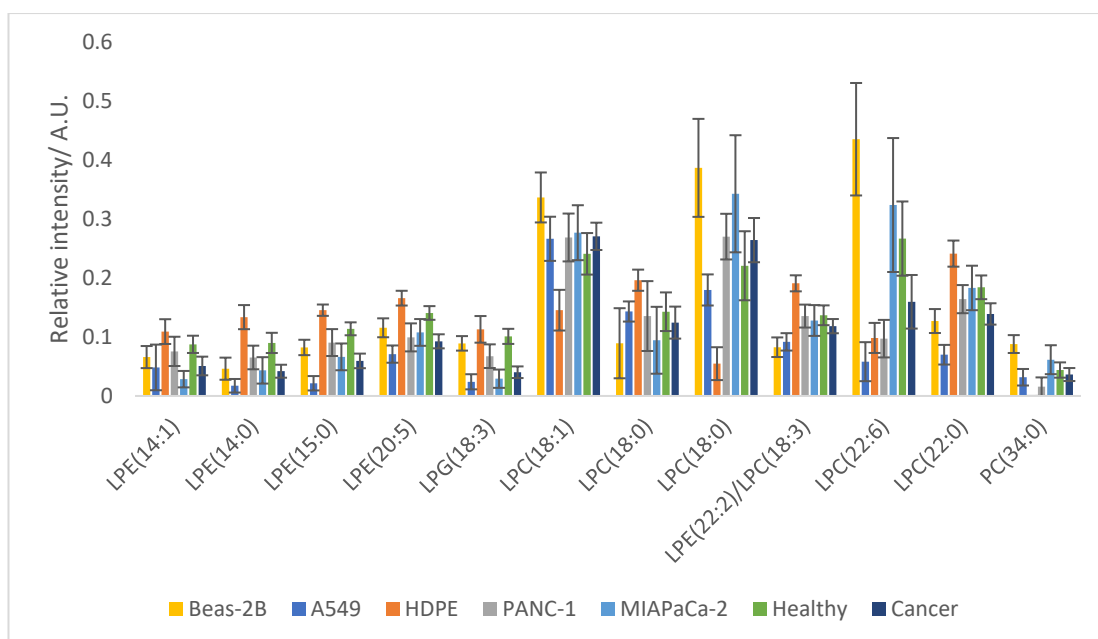


Figure 2-35- Significantly different lysophospholipids found to discriminate multiple different cancerous cell lines from their healthy equivalent

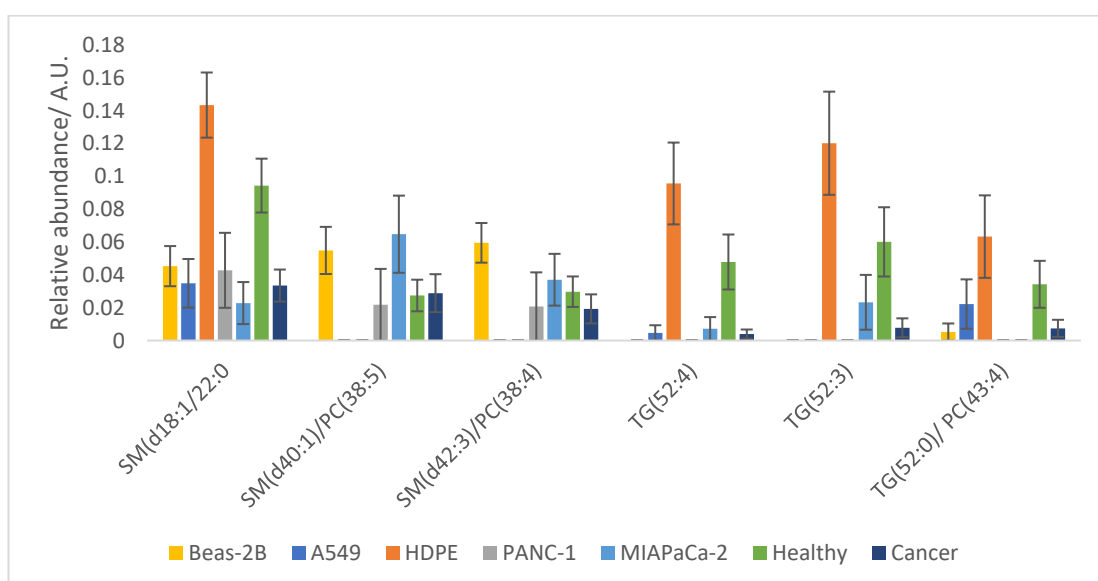


Figure 2-36- Significantly different Sphingolipids/ Triacylglycerol markers found to differentiate multiple different group comparisons when looking at Beas-2B, A549, HDPE, Panc-1 and MIAPaCa-2 cell lines relative abundances.

The MALDI analysis in positive ionisation mode also allowed for the study of suspected peptides presented by the extracted sample. A number of putative peptides were found to be significantly different across multiple group comparisons.

The patterns seen in these significant variations changed entirely with cell source (lung/pancreatic) under analysis but was found to remain consistent within the groups (Panc-1 and MIA PaCa-2).

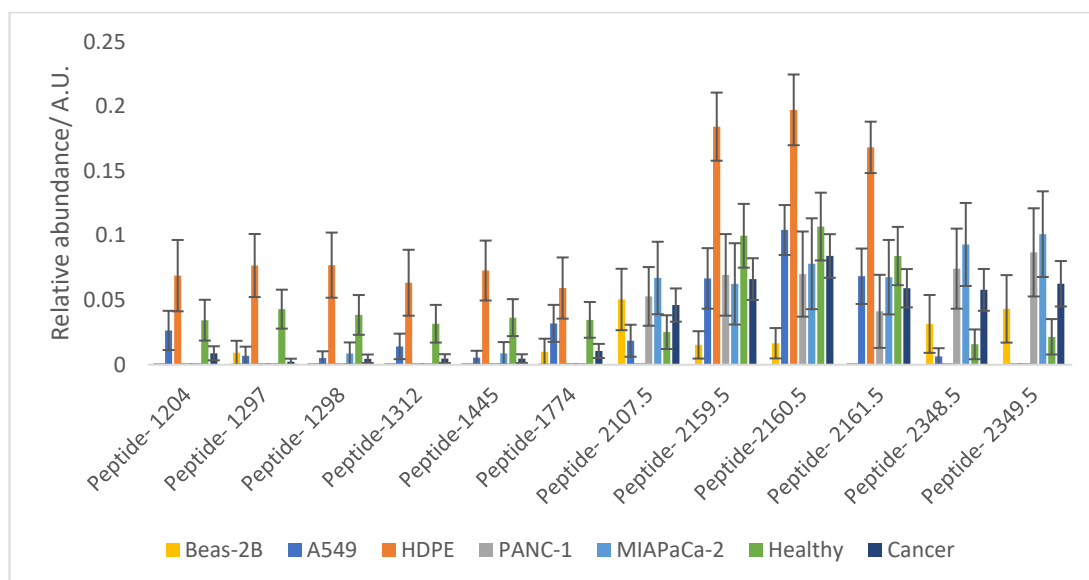


Figure 2-37- Significantly different peptide markers found to distinguish multiple different cancerous groups from healthy equivalent cell lines

No Glycerophospholipids were found to be significantly different across multiple group comparisons.

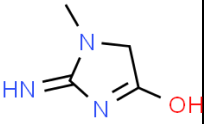
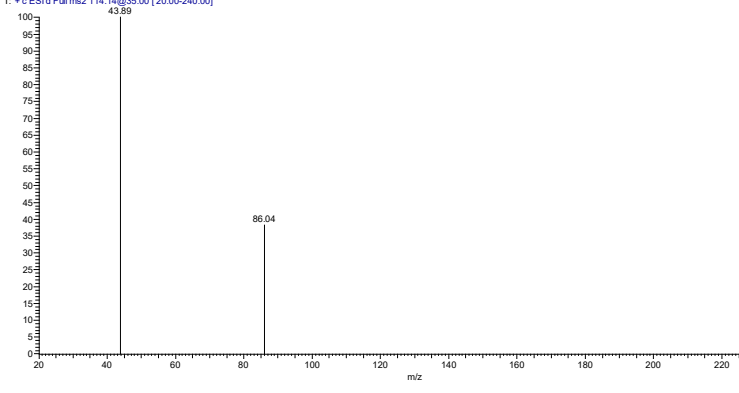
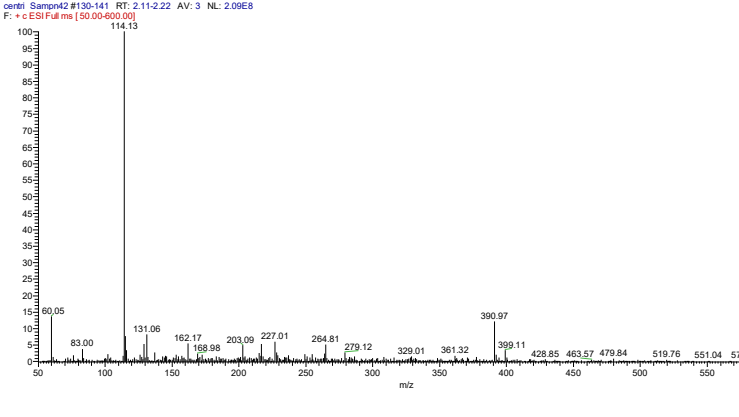
#### 2.4.4. Urinary modified nucleoside analysis

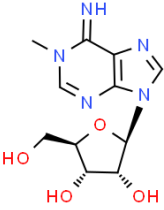
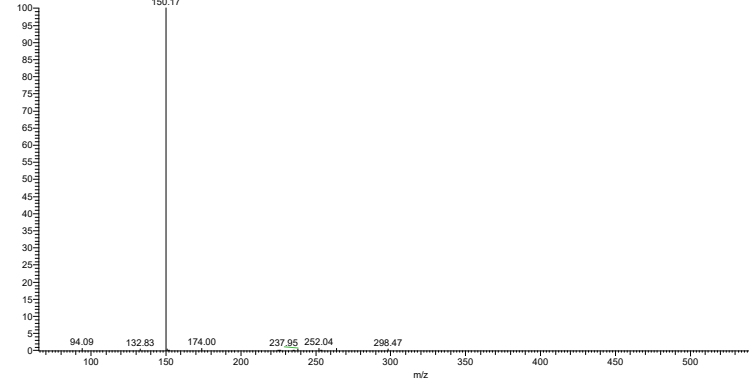
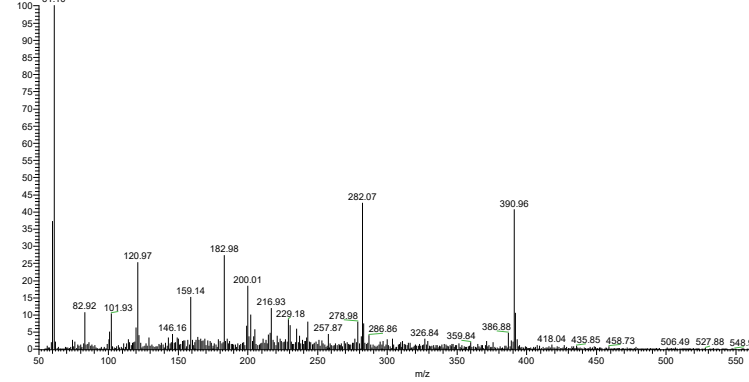
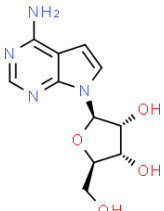
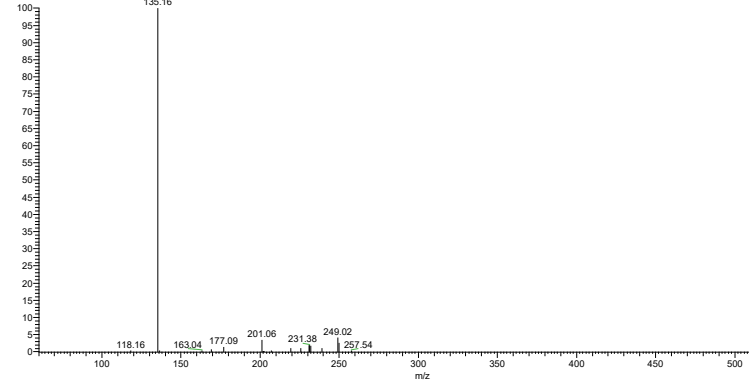
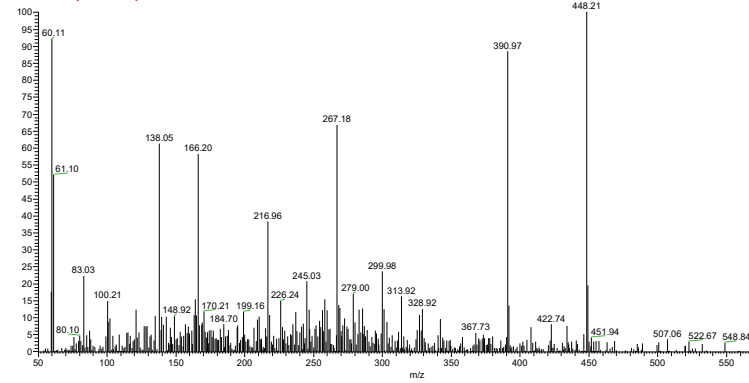
As a separate analysis looking at potential biomarkers of pancreatic cancer, urine samples were collected from patients with pancreatic cancer, other pancreatic disorders and healthy volunteers and extracted as described in order to partially purify urinary metabolites. Whereas the cancer cell lines provide a source of diagnostic marker, the purpose of the urinary analysis is to seek a non-invasive screening tool to identify candidates likely to have pancreatic cancer. The samples were analysed by HPLC-data dependent tandem mass spectrometry in order to generate full scan spectra and tandem mass spectra of abundant ions in an automated fashion. The data produced was then analysed using Xcalibur, reporting

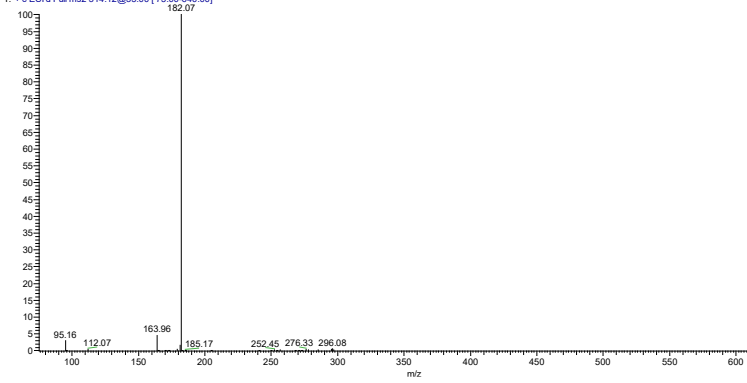
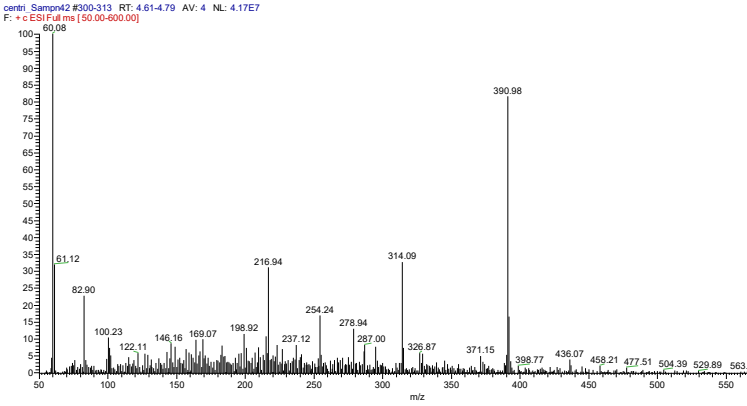
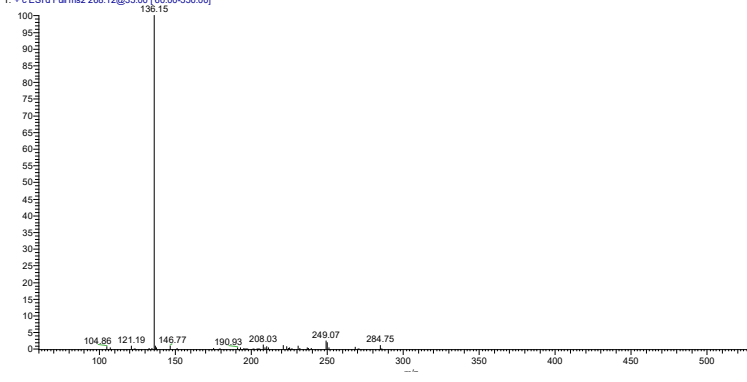
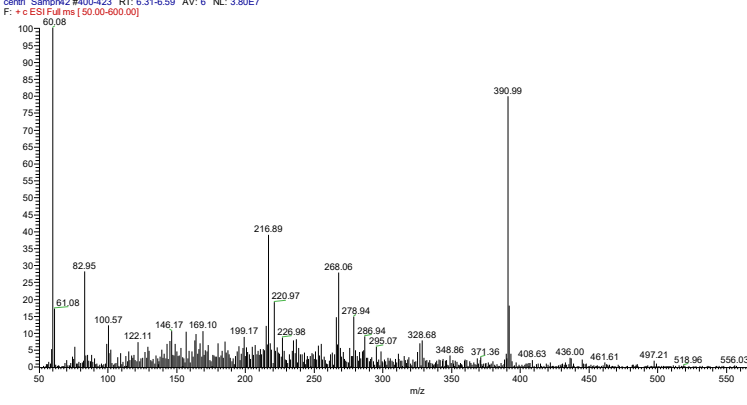
any  $m/z$  peaks that showed a constant neutral loss of 132 during tandem mass spectrometric fragmentation. This loss can be attributed to the loss of a ribose sugar moiety indicative of a nucleoside and for such ions the corresponding peak area, retention time and the fragmentation detail were recorded. The intact mass and tandem mass spectra were compared with existing literature and in doing so a total of 13 peaks of interest were noted from each sample including the internal standard tubercidin and aside from this, the urinary metabolite creatinine was also identified and its levels used for urinary volume normalisation. The metabolites and compounds identified in this manner are summarised in Table 2-9. For the metabolite 1-methylguanosine there were two separate peaks found to elute for this nucleoside, one at 12.8 and the other at 13.2 minutes. These have been assigned tentative N1 and N2 identifications following the findings of Dudley E., et al. (2004) whereby the N1 variant was found to elute first utilising similar methodology.

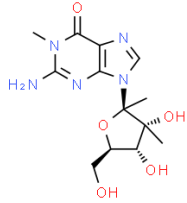
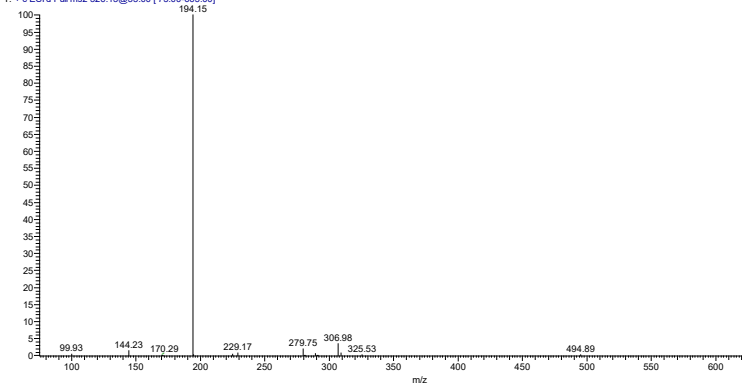
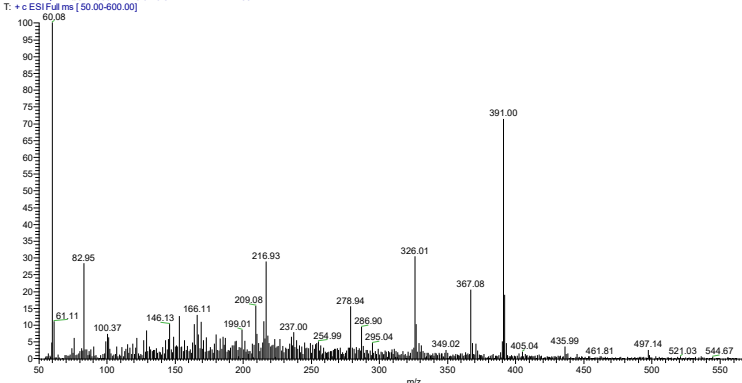
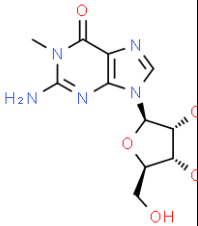
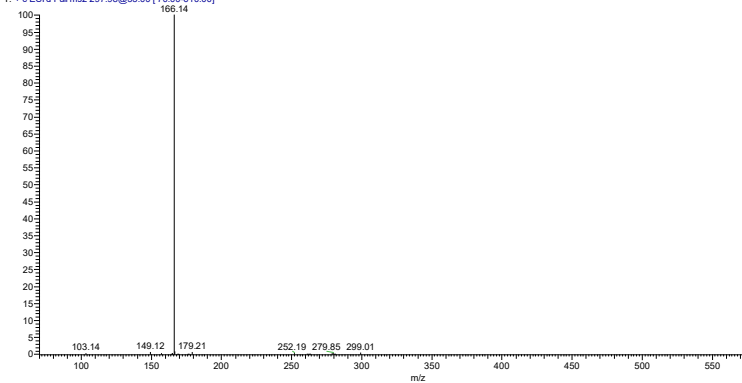
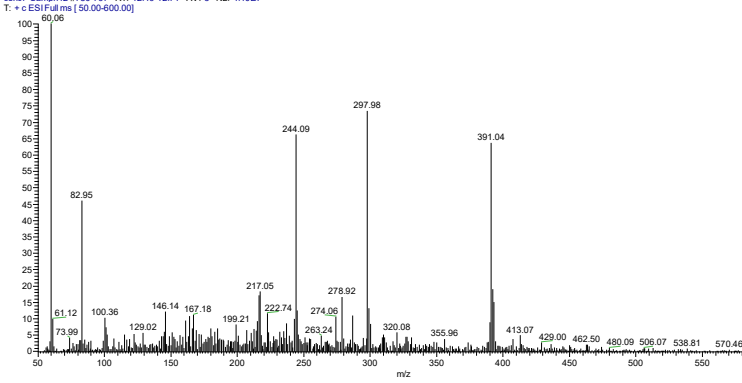
The data was normalised, firstly for experimental changes according to the relative intensity of the internal standard tubercidin before then being normalised for kidney functioning utilising creatinine levels. Following this, the levels were compared between the pancreatic cancer cohort and healthy individuals.

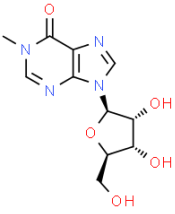
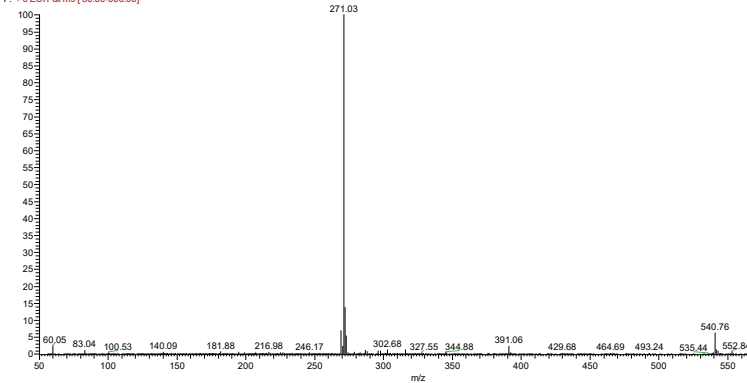
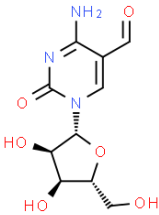
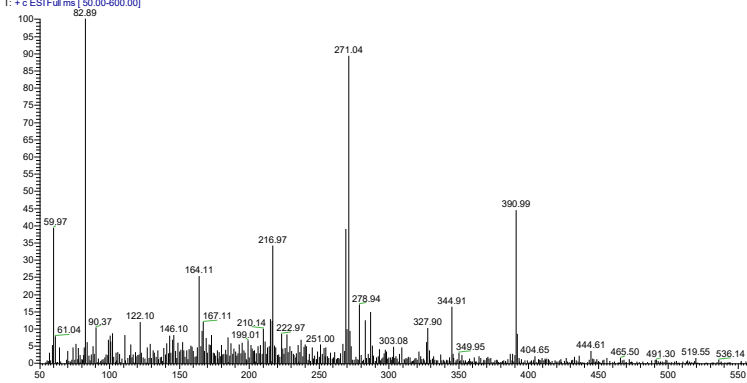
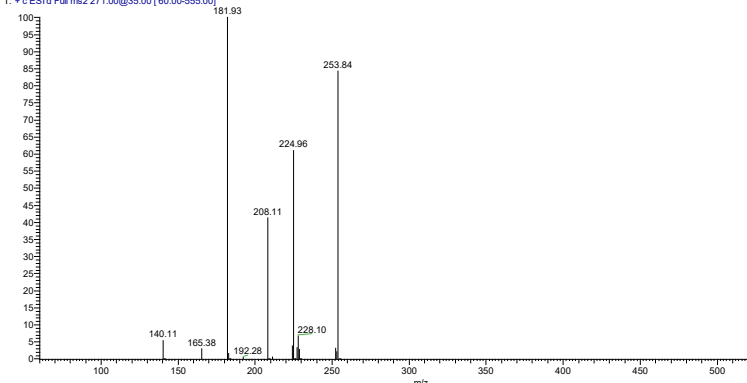
Table 2-9- Modified nucleosides of interest identified in urine samples using HPLC analysis; with their corresponding retention times, compound  $m/z$  anticipated in positive ionisation mode, fragment ion of interest, in the case of the modified nucleosides this is indicative of a loss of 132, nucleoside identity, formula, Mr and structure with fragmentation indicated in red and blue detail.

Retention time/min	$m/z$	Fragment	Identity Formula Mr	Scan MS <sup>2</sup> Full ms
2	114	86	Creatinine $C_4H_7N_3O$ Mr: 113.12g.mol <sup>-1</sup> 	<p>centri Sampn42 #130 RT: 2.08 AV: 1 NL: 3.17E5 T: + c ESI d Full ms [20.00-240.00]</p>  <p>centri Sampn42 #130-141 RT: 2.11-2.22 AV: 3 NL: 2.09E8 F: + c ESI Full ms [50.00-600.00]</p> 

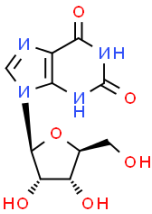
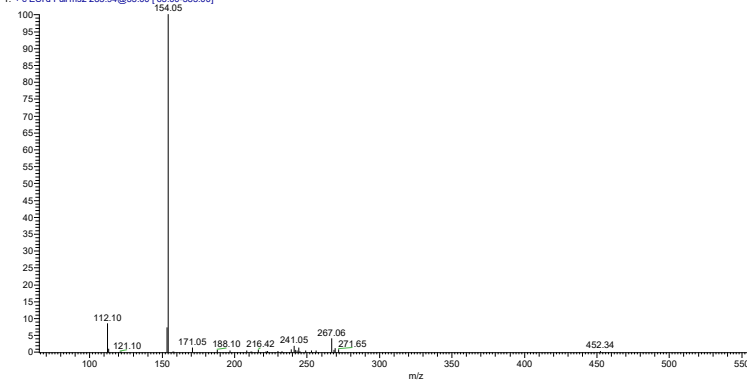
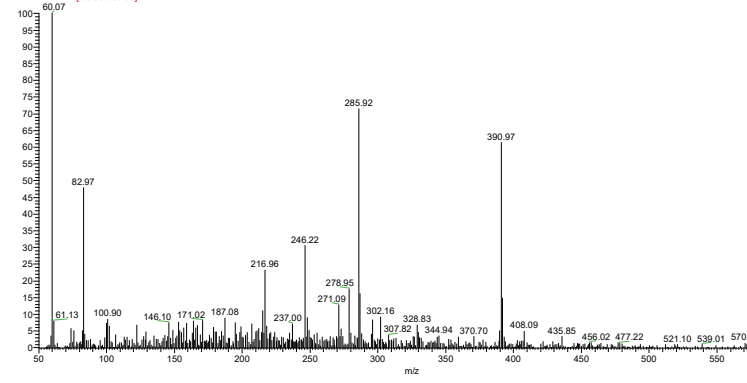
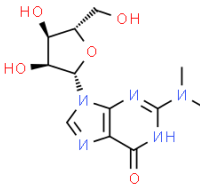
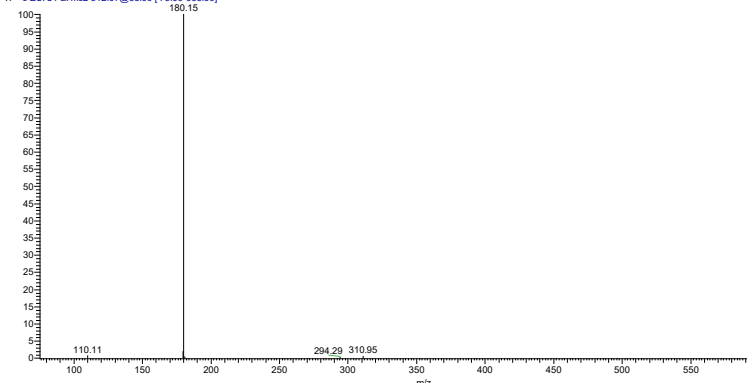
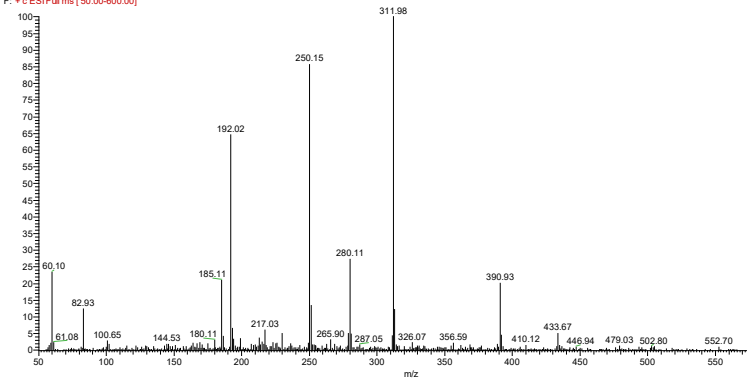
3.8	282	150	<p>1-</p> <p>methyladenosine</p> <p><math>C_{11}H_{15}N_5O_4</math></p> <p>Mr: 281.27 g.mol<sup>-1</sup></p> <p>1</p> 	<p>centri_Sampn#2 #168 RT: 2.57 AV: 1 NL: 3.38E7 T: + c ESI d Full ms2 282.05@35.00 [65.00-575.00]</p>  <p>centri_Sampn#2 #165-222 RT: 2.53-3.30 AV: 15 NL: 7.89E7 F: + c ESI Full ms [50.00-600.00]</p> 
4.5	267	135	<p>Tubercidin</p> <p><math>C_{11}H_{14}N_4O_4</math></p> <p>Mr: 266.253 g.mol<sup>-1</sup></p> 	<p>centri_Sampn#2 #259 RT: 3.92 AV: 1 NL: 1.49E7 T: + c ESI d Full ms2 267.22@35.00 [60.00-545.00]</p>  <p>centri_Sampn#2 #254-262 RT: 3.90-3.95 AV: 2 NL: 3.71E7 F: + c ESI Full ms [50.00-600.00]</p> 

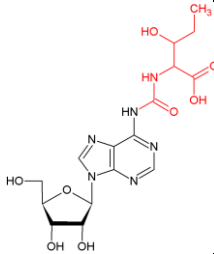
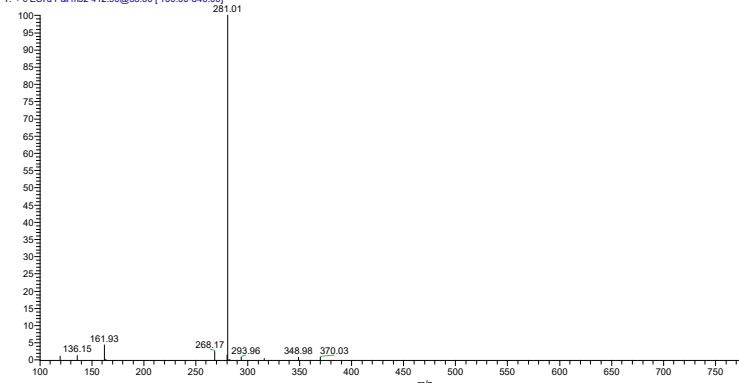
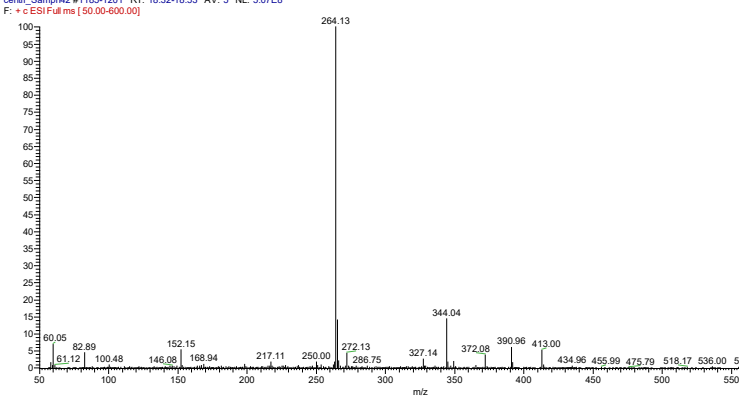
6	314	182	Ethylacetatecytidine	<p>centr1_Samprn42 #303 RT: 4.64 AV: 1 NL: 7.64E6 T: + c ESI'd Full ms2 314.12@35.00 [75.00-640.00]</p>  <p>centr1_Samprn42 #300-313 RT: 4.61-4.79 AV: 4 NL: 4.17E7 F: + c ESI'd Full ms [50.00-600.00]</p> 
7.5	268	136	Adenosine $C_{10}H_{13}N_5O_4$ Mr: 241 g.mol <sup>-1</sup>	<p>centr1_Samprn42 #404 RT: 6.35 AV: 1 NL: 3.44E6 T: + c ESI'd Full ms2 268.12@35.00 [60.00-550.00]</p>  <p>centr1_Samprn42 #400-423 RT: 6.31-6.59 AV: 6 NL: 3.80E7 F: + c ESI'd Full ms [50.00-600.00]</p> 

11	326	194	<p>N2,N2,2'O-trimethylguanosi ne C<sub>13</sub>H<sub>19</sub>N<sub>5</sub>O<sub>5</sub></p> <p>Mr: 325.320 g.mol<sup>-1</sup></p> 	<p>centri_Sampn#2 #527 RT: 8.34 AV: 1 NL: 2.00E6 T: + c ESI d Full ms2 326.13@35.00 [ 75.00-665.00]</p>  <p>centri_Sampn#2 #522-569 RT: 8.31-9.01 AV: 12 NL: 4.06E7 T: + c ESI d Full ms [ 50.00-600.00]</p> 
12.8	298	166	<p>N1-methylguanosine C<sub>11</sub>H<sub>15</sub>N<sub>5</sub>O<sub>5</sub></p> <p>Mr: 297.267 g.mol<sup>-1</sup></p> 	<p>centri_Sampn#2 #786 RT: 12.56 AV: 1 NL: 1.94E7 T: + c ESI d Full ms2 297.98@35.00 [ 70.00-610.00]</p>  <p>centri_Sampn#2 #780-797 RT: 12.49-12.71 AV: 5 NL: 4.40E7 T: + c ESI d Full ms [ 50.00-600.00]</p> 
13.2	298	166	N2-methylguanosine	

13.2	283	151	<p>1-methylinosine</p> <p>282.25g/mol</p> <p>Mr: C<sub>11</sub>H<sub>14</sub>N<sub>4</sub>O<sub>5</sub></p> 	<p>centri_Sampn42 #808-822 RT: 12.87-13.07 AV: 4 NL: 1.34E9 F: + c ESI Full ms [ 50.00-600.00]</p> 
11.9 9	271	139	<p>5-formylcytidine</p> <p>C<sub>10</sub>H<sub>13</sub>N<sub>3</sub>O<sub>6</sub></p> <p>Mr: 271.227 g.mol<sup>-1</sup></p> 	<p>centri_Samp79 #739-759 RT: 11.74-11.97 AV: 5 NL: 4.69E7 T: + c ESI Full ms [ 50.00-600.00]</p>  <p>centri_Samp79 #750 RT: 11.86 AV: 1 NL: 1.75E7 T: + c ESI d Full ms2 271.00@35.00 [ 60.00-555.00]</p> 



13.9	286	154	<p><b>Xanthosine</b></p> <p><math>C_{10}H_{12}N_4O_6</math></p> <p>Mr: 284.225 g.mol<sup>-1</sup></p> 	<p>centri_Sampn42 #854 RT: 13.59 AV: 1 NL: 4.80E6 T: + c ESI d Full ms2 285.94@35.00 [65.00-585.00]</p>  <p>centri_Sampn42 #852-864 RT: 13.58-13.70 AV: 3 NL: 4.69E7 F: + c ESI d Full ms [50.00-600.00]</p> 
14.5	312	180	<p><b>Dimethylguanosi ne</b></p> <p>?(162h2o; 153hcn; 137ch3nch2, 110ncn(ch3)2; 128 unidentified)</p> <p><math>C_{12}H_{17}N_5O_5</math></p> <p>Mr: 311.294 g.mol<sup>-1</sup></p> 	<p>centri_Sampn42 #895 RT: 14.21 AV: 1 NL: 3.29E7 T: + c ESI d Full ms2 312.07@35.00 [75.00-635.00]</p>  <p>centri_Sampn42 #891-908 RT: 14.19-14.35 AV: 4 NL: 1.49E8 F: + c ESI d Full ms [50.00-600.00]</p> 

18.5	413	281	<p>N6-threonylcarbamoyladenine</p> <p><math>C_{15}O_8N_6H_{20}</math></p> <p>Mr: 413.355 g.mol<sup>-1</sup></p> 	<p>centri_Sampn42 #1191 RT: 18.39 AV: 1 NL: 4.68E6 T: + c ESI d Full ms [ 100.00-840.00 ]</p>  <p>centri_Sampn42 #1185-1201 RT: 18.32-18.53 AV: 5 NL: 5.07E8 F: + c ESI Full ms [ 50.00-600.00 ]</p> 
------	-----	-----	---	--

Initial analysis was conducted to seek out notable profile differences between male and female individuals. This was conducted to determine if gender was a factor to consider for modified nucleoside analysis. When looking at all data (and not factoring in gender bias amongst diagnosis) no significant alterations were seen. When then moving on to consider inter-group variability with gender a single significant variation was noted, when looking at the first 5-formyl cytidine peak's level between male and female pancreatitis groups. No other subgroups showed any significant alterations based on gender and as such it was concluded that this does not require further consideration as a factor in the analysis.

Generally, the vast majority of the nucleosides of interest are found to be upregulated in PDAC conditions when being compared to healthy and, less so, other candidate cohorts, Figure 2-38.

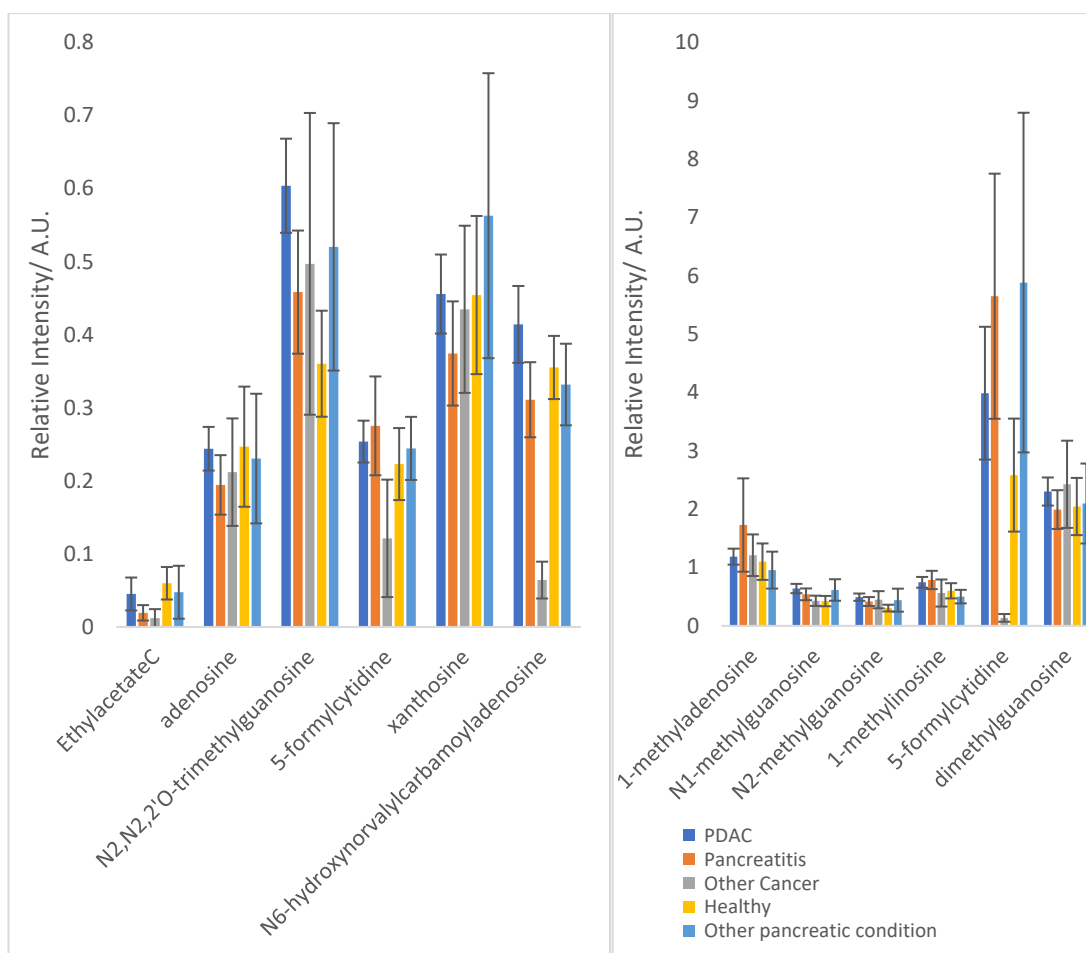


Figure 2-38- The average intensities of all of the modified nucleosides of interest comparing all of the different groups of interest; Pancreatic ductal adenocarcinoma, Pancreatitis, other cancer, healthy and other pancreatic conditions

The first stage of analysis involved drawing comparisons between the PDAC patients versus all other groups combined. Of these, N2,N2,2'O-trimethyl guanosine, and both of the 1-methyl guanosine peaks were found to be significantly upregulated in PDAC conditions when compared to all of the other groups under analysis combined, achieving p values of 0.001, 0.047 and 0.034 respectively. This is shown in Figure 2-39.

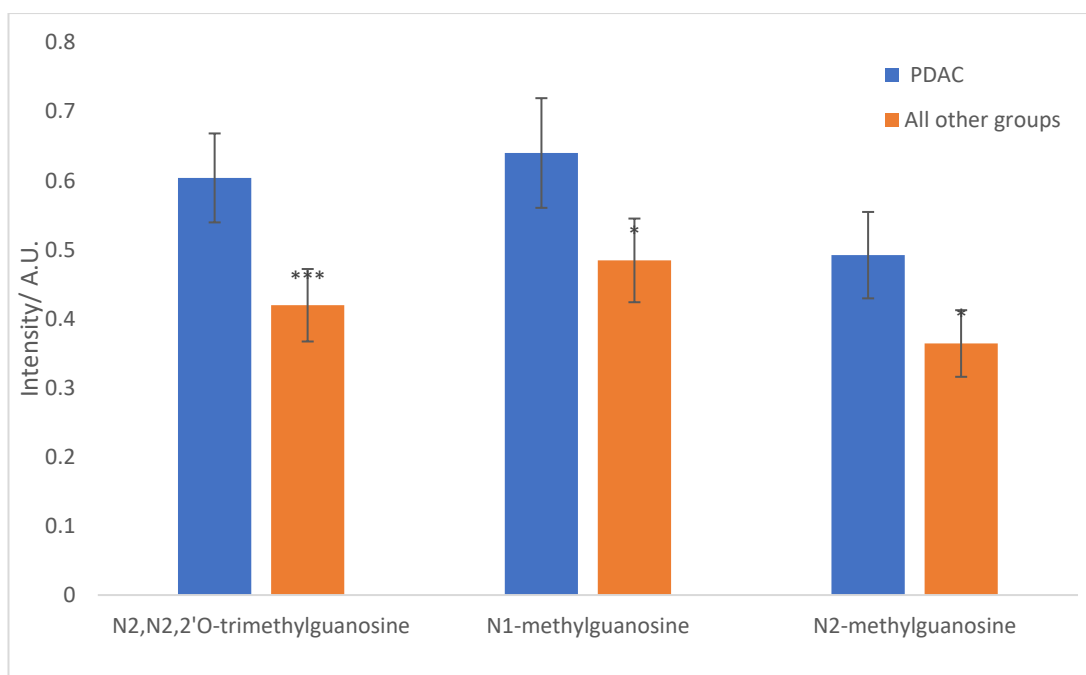


Figure 2-39- Significantly upregulated modified nucleosides in Pancreatic ductal adenocarcinoma vs all other candidate groups combined. Significant differences have been signified with asterisks with \* denoting a p. value of less than 0.05, \*\* less than 0.025 and \*\*\* less than 0.005.

The next stage of comparison was to check which modified nucleosides were found to be significantly distinguishable between healthy and all other study candidates combined, and in doing so considering a large number of hepatic conditions. These include 1-methyl adenosine, N2,N2,2'O-trimethyl guanosine, both of the 1-methyl guanosine peaks, 1-methyl inosine and di-methylguanosine, Figure 2-40. The other groups were always found to hold significantly higher intensities than healthy candidates and achieved p values of 0.015, 0.000, 0.003, 0.008, 0.034 and 0.017 respectively.

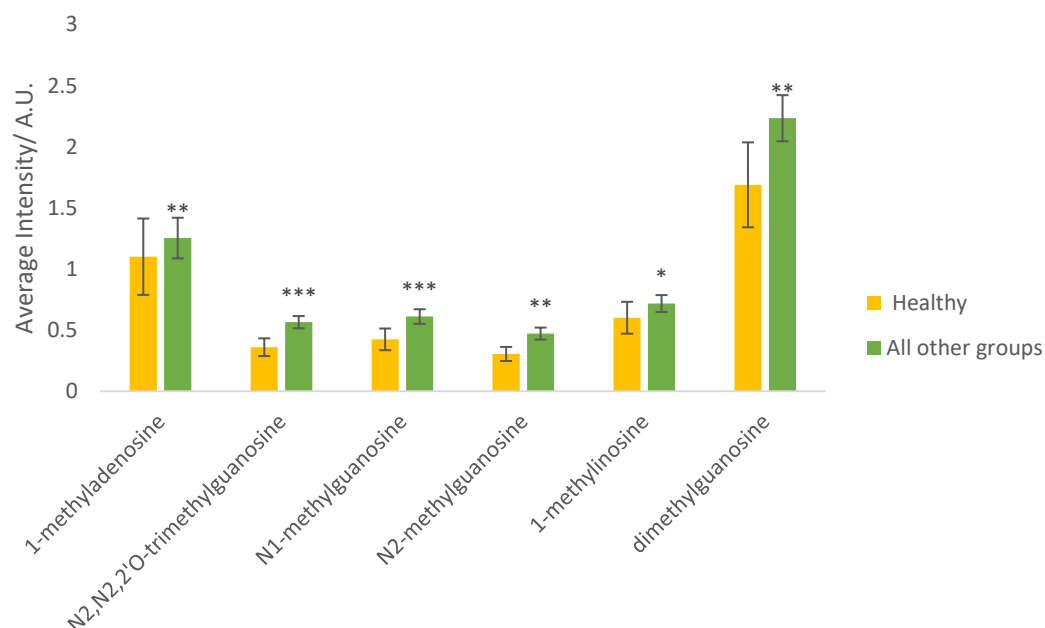


Figure 2-40- Significantly different modified nucleoside intensities between healthy and all other group candidates combined. Significant differences have been signified with asterisks with one denoting a p. value of less than 0.05, two less than 0.025 and three less than 0.005.

The next comparison sought to study variance between healthy and PDAC candidates specifically. Six nucleoside markers were found to be significantly upregulated in PDAC sufferers urine when compared to the healthy cohort. There were; 1-methyl adenosine, N2,N2,2'O-trimethylguanosine, both 1-methylguanosine peaks, 1-methylinosine and di-methylguanosine, which achieved p.values of 0.016, 0.000, 0.004, 0.008, 0.029 and 0.020 respectively, as shown in

Figure 2-41.

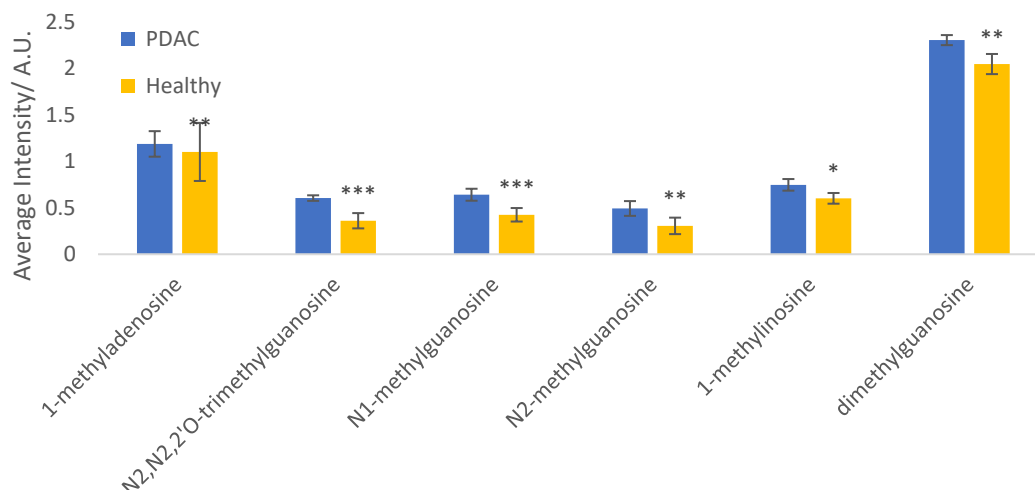


Figure 2-41- the significantly different modified nucleosides between pancreatic ductal adenocarcinoma and healthy cohorts. Significant differences have been signified with asterisks with one denoting a p. value of less than 0.05, two less than 0.025 and three less than 0.005.

No significantly different urinary nucleoside markers were found to distinguish PDAC from the pancreatitis group. Despite the values not showing significant alterations between pancreatitis and PDAC conditions, all of the nucleosides, with the exception of 1-methyladenosine and 1-methylinosine, are upregulated in PDAC when compared to pancreatitis. Furthermore, all of the nucleosides are found with lowest intensities in healthy candidates. Additionally, none of the nucleosides of interest are found to bare significant differences between pancreatitis and healthy cohorts, Figure 2-42, though this may reflect on the more limited cohort size available in pancreatitis groups. However, this does also show that the significantly different markers are exclusive for PDAC and are not significantly varied in other pancreatic related ailments including pancreatitis.

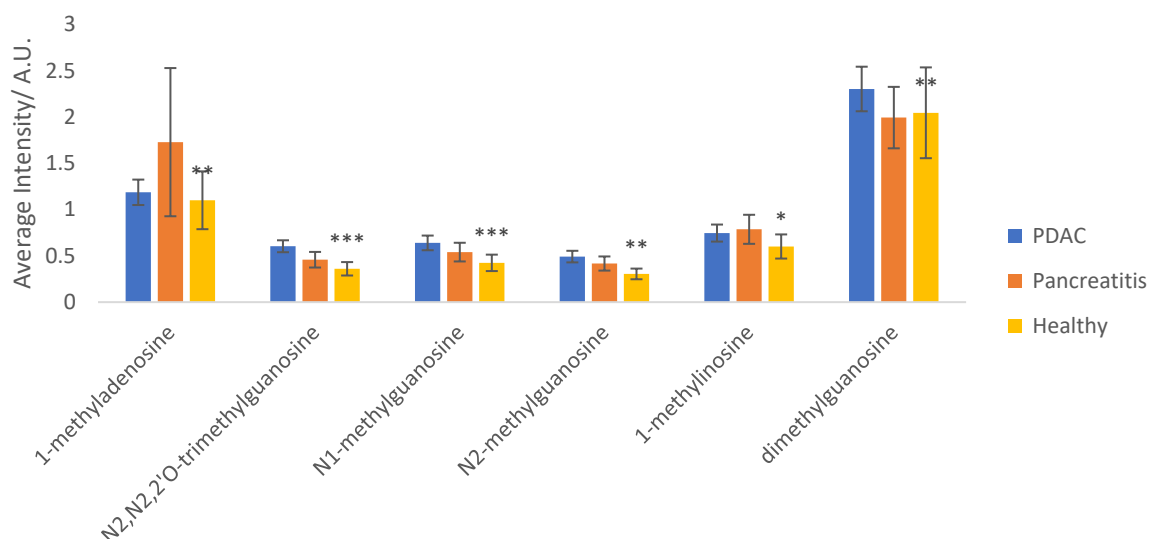


Figure 2-42- Alterations in nucleosides found to be significantly different between healthy and pancreatic ductal adenocarcinoma cohorts, with additional comparison of pancreatitis candidate levels. Significant differences have been signified with asterisks with one denoting a p. value of less than 0.05, two less than 0.025 and three less than 0.005.

When looking at the diagnostic quality of the biomarkers of interest, none of the markers identified provided strong enough ROC scoring to be considered suitable individual biomarker candidates. Even when used in a panel of the most significant markers the ROC scoring is still not better than “good” in all elements.

An advantage of the HPLC MS methodology utilised in this study is that the profile of urinary modified nucleosides is collected without defining the  $m/z$  to be analysed at the outset of the analysis due to the application of data dependent analysis acquisition. When looking at specific nucleoside classes, methylguanosines as a subclass of modified nucleosides were seen to generally be elevated in the urine of pancreatic cancer patients compared to the other groups included in the study (Figure 2-43). As a result of this, the utilisation of both the total ion current (combined normalised peak areas) of all the identified nucleosides and just the methylated guanosine nucleosides was studied for their diagnostic ability to distinguish pancreatic cancer patients from the healthy control cohort. The outcome of the analysis is summarised in Table 2-10, and whilst compared to the thresholds laid out in the Ray et al study, these exhibit some promise. Specifically, the sensitivity of the analysis of the total combined nucleoside profile 88% and both the total nucleoside

profile and the methylated guanosine nucleosides alone generate an accuracy of 66%, suggesting some promise in diagnostic potential.

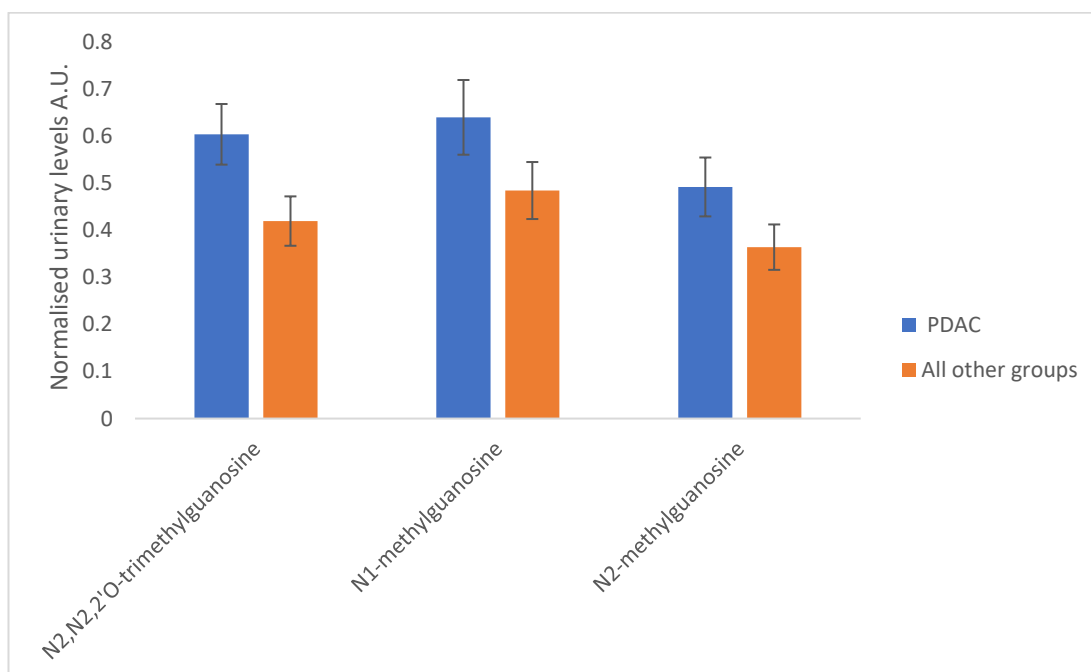


Figure 2-43 The average variation of methylguanosine levels between pancreatic ductal adenocarcinoma and all other groups under study

Table 2-10- The diagnostic properties of the combined nucleosides and combined guanosine modifications under study during the urine analysis used to differentiate PDAC from all other groups under analysis.

Comparison	AUC	Sensitivity	Specificity	Accuracy	+LHR	-LHR
Combined nucleosides	0.645	0.883	0.594	66%	1.48	0.288
Combined guanosine modifications	0.672	0.70	0.584	66%	1.68	0.51



#### 2.4.5 SILAC metabolic pathway analysis

Upon noting a number of significant differences between the metabolic profiles of different pancreatic cancer cell lines a SILAC study was undertaken to better understand the specific pathways that have been altered during metabolic alterations. Uniformly labelled glucose (U- $^{13}\text{C}$ ) and glutamine (U- $^{13}\text{C}$ ) tracing experiments were conducted in order to look at the proportion of labelled to unlabelled marker isotopes present in specific compound peaks that relate to the TCA cycle. This will allow the identification of  $\text{C}^{13}$  labelled carbons carried over to downstream of glucose/ glutamine and hence enable the identification of pathways that are more heavily utilised by the different cells studied in this study.

Glucose is a major metabolite and cellular energy source, generating ATP via glycolysis and further through feeding into either aerobic (the TCA cycle) or anaerobic (lactate) production, Figure 2-44. The up-regulation of these pathways is a known mechanism utilised by cancer cells to help keep up with the increased metabolic requirements. The increased feed through of glycolysis can lead to pyruvate accumulation, being the common source for both pathways. The Warburg effect involves the anaerobic energy production accompanied by lactate production, which is heavily utilised by cancer cells even in the presence of oxygen. Another potential source of carbon utilisation in the TCA cycle, especially for highly proliferating cancerous cells, is through glutamine entry at the point of  $\alpha$ -ketoglutarate (Vanhove, K., et al., 2019). This pathway has also been analysed to establish how cancerous and non-cancerous cells differ in their flux. A final further analysis was then conducted to establish the TCA related consequences of glucose starvation relating to the glutamine utilisation pathway with a focus on how this differently impacted the different cancerous cell lines under evaluation.

Glucose labelling was the first to be analysed in order to establish the most prevalent energy resource utilised by cancerous and non-cancerous lung and pancreatic cell

lines and establish regions of dys-regulation that are shared and unique to the different cell types under evaluation and common with diseased status.

This study found the turn-around of pyruvate, in accordance with the percentage of labelled to unlabelled metabolite presence, indicating the utilisation of glucose through glycolysis was for Beas-2b 74%, A549 76%, HDPE 44%, Panc-1 83% and MIA PaCa-2 20% as shown in Figure 2-45. These are calculated through taking the extracted chromatogram from the selected peaks associated ion's (M+0) and labelled isotopologue's peak area and calculating the ratio of these two factors. The variation was only seen to significantly vary in the pancreatic cell lines when comparing cancerous and non-cancerous presence.

Pyruvate is the metabolic product of glycolysis and elicits four main potential uses at cellular level to support growth and proliferation. These can all be described through the pyruvate junction and include the formation of either lactate, alanine or one of two pathways into the TCA cycle either through formation of acetyl Co-A, which will feed into citric acid production or through the formation of oxaloacetate, this can be directly utilised in the TCA cycle. Analysing the labelled pyruvate products flux through these four different pathways can be determined through evaluating the carbon labelling observed during Glucose C<sup>13</sup> labelling sets of experiments using the schematic set out in Figure 2-45 whereby the production of lactate or alanine can be monitored through the measurement of the (M+3) isotopologue's of each metabolite in comparison to the unlabelled form. The alternative pathways both involved the entry into the TCA cycle and as such can be monitored through citric acid isotopologue evaluation using the (M+2) and (M+3) variants. These different forms establish the utilisation of the acetyl-co-A and oxaloacetate conversion respectively as these give indication of the product formed during the first turn of the TCA cycle. Statistical analysis was conducted using one-way ANOVA followed by Tukey's post-hoc analysis.

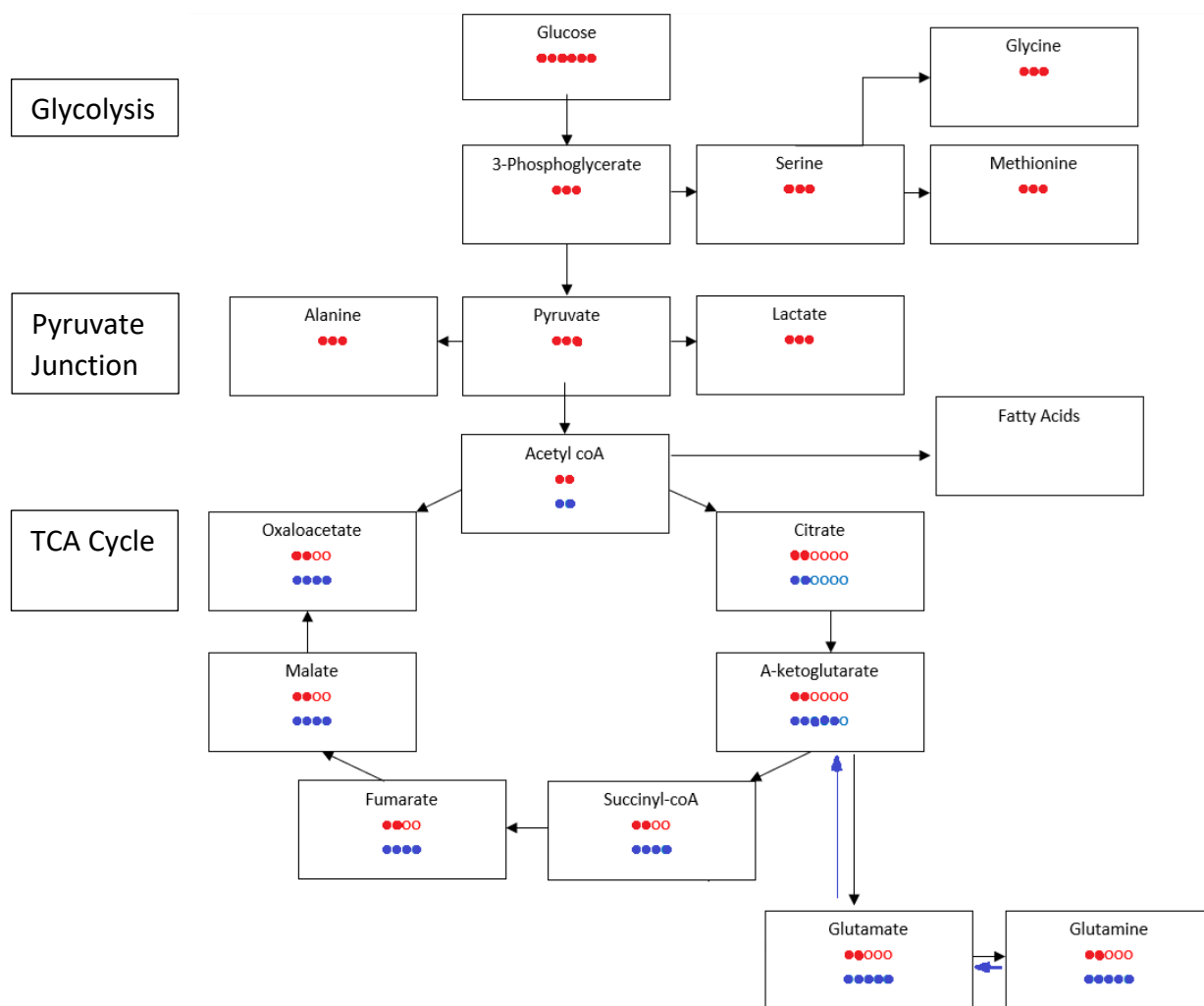


Figure 2-44- The metabolic pathway under evaluation in the study showing the carbon presence from glucose labelling (red) and glutamine (blue). Carbons are shown by the circles with coloured circles indicating a labelled carbon and a hollow circle no labelling upon the carbon.

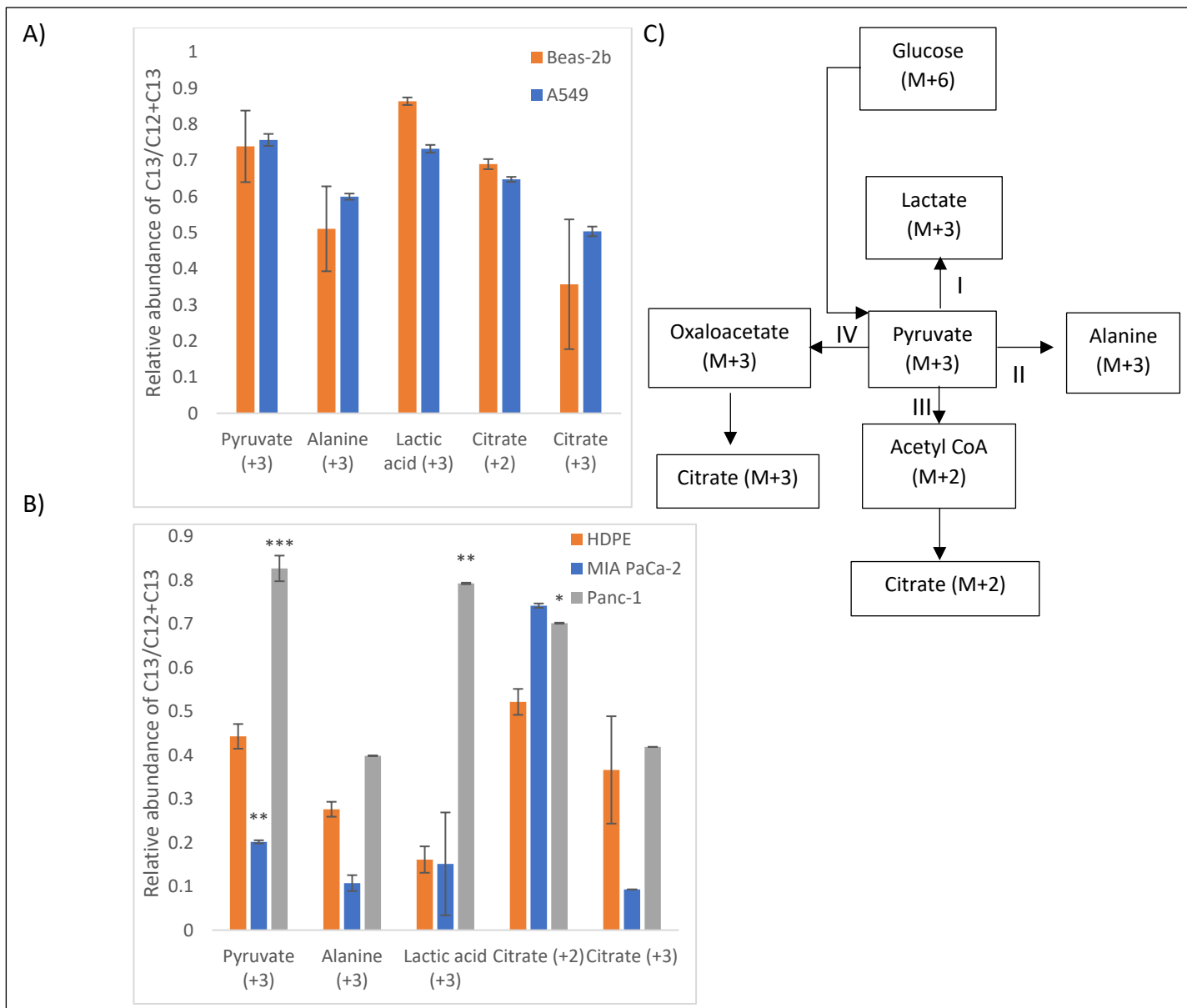


Figure 2-45- Identifying the fate of pyruvate produced during glucose metabolism A) Alterations in key metabolites for lung cell lines B) Alterations in key metabolites from pancreatic cell extracts C) Schematic showing the possible pathways pyruvate can enter and how these can be identified; I) can be measured using the generation of M+3 lactate II) Glutamine mediated amination into alanine, this can be measured utilising the presence of M+3 alanine III) Decarboxylation of pyruvate into acetyl CoA, this can be measured looking at the presence of M+2 citrate. IV) The carboxylation of pyruvate into oxaloacetate This can be measured looking at M+3 labelled citrate. Significant variance between the cancerous cell lines and healthy equivalent is illustrated with asterisk.

Whilst this cannot be used to give absolute definitions regarding how much each pathway is utilised due to these metabolites having further downstream uses in

addition to other potential influxes other than glucose, it can give a reflection on how much the glucose contributed to the total presence.

Both healthy and cancerous lung cell lines share similar turn-over profiles of the metabolites described in Figure 2-45 relating to the entry of glucose in proportion to pre-existing cellular sources, when comparing the ratio of labelled to unlabelled metabolite presence. Both have a high turn-over of lactic acid from the initial labelled glucose source, with over 70% carrying the (M+3) marker from direct conversion pyruvate. This turnover is higher in healthy Beas-2b line (86%) than in cancerous A549 (73%). Direct entry into the TCA cycle, with observation of Citrate (M+2) being seen, is the next highest priority (69 & 65% respectively healthy vs cancerous) followed by Alanine (51% & 60%) and then Citrate M+3 (34% & 50%). Surprisingly, the turn-over is highest in the cancerous cell line only in Citrate M+3 and Alanine. Comparatively pancreatic cells show a greater alteration in pathways than the lung cells did when malignancy is present. For HDPE and MIA PaCa-2 lines direct entry into the TCA cycle (Citrate M+ 2) is the main route shown (52% & 74%), whereas for Panc-1 this is the lactic acid pathway first (79%) followed by the TCA (70%) relating into the proportion of labelled marker presence. For healthy lines, the lactic acid pathway has the lowest turnaround with 16%, comparable with MIA PaCa-2 15%. The proportion of labelled Alanine generated through labelled glucose utilisation was lower in pancreatic cell lines (HDPE 28%, MIAPaCa-2 11%, Panc-1 40%) when compared to lungs (51% Beas-2b and 59% A549) and was generally found to be higher in cancerous cell lines when compared to non-cancerous, with the exception of MIA PaCa-2. There is also a pyruvate salvage pathway leading from malate back to pyruvate to be considered. This can be captured through the production of pyruvate (M+2) isotopologue, rather than the M+3 already discussed. This form comes if the malate carries the labelled carbons and is shuttled to the cytosol and decarboxylated into pyruvate with the by-product formation of NADPH. The levels between lung and pancreatic cell lines were comparable when tracing the proportion of labelled carbons sourced from glucose. The pathway was found to be less utilised by cancerous cell lines than their healthy equivalent, significantly so in MIA-PaCa-2 cell lines.

Next when considering the flux through the TCA cycle there are a number of markers and their isotopologues that can be considered to establish the fate of the carbons after entry. Citrate is a good example of a metabolite to monitor to observe these trends. One of the previously discussed entry routes, involving the conversions of M+3 labelled pyruvate into M+2 acetyl-CoA. This will initially generate Citrate isotopologue M+2. The citrate can continue to cycle around the citric acid cycle, which will result in the gradual increase of the isotopologue label as more labelled acetyl CoA is incorporated. The M+2 labelled citrate will combine with another M+2 labelled acetyl-CoA to form M+4 citrate and finally, after another cycle, the citrate will form the M+6 isotopologue whereby it is fully saturated in labelled carbons. Alternatively, the other source of entry previously discussed would generate the (M+3) isotopologue of Citrate. This would go on to generate the M+5 and then M+6 fully saturated isotopologues of Citrate with turns of the cycle. There is also the potential to form an (M+5) isotopologue of citrate through the product of an (M+2) acetyl-CoA combining with this (M+3) oxaloacetate, Figure 2-45.

The proportional presence of labelled citrate was comparable between lung and pancreatic cell lines and was found to be at a high level in all cells under study. Lung cancer showed a slight decreased presence of the carried carbons from glucose in all bar Citrate M+3 which was increased when compared to its healthy counterpart. In comparison Pancreatic cancerous cell lines had a higher level of Citrate 2, Panc-1 showed an elevated Citrate M+3 whereas Mia-PaCa-2 was decreased. Mia PaCa-2 was increased in all other citrate variations whereas Panc-1 was decreased.

If Citrate is considered to be the start point of the TCA cycle the downstream metabolites that were found at suitable levels to be detectable in this analysis were  $\alpha$ -ketoglutarate (M+2 +4 & M+5) succinate (M+2 and M+4) and Malate (M+2 and M+4). In lung cancer cell lines almost all other discernible TCA constituents were found to be increased in cancerous than non-cancerous cells with the exception of one isotopologue of  $\alpha$ -ketoglutarate. The same was true of MIA-PaCa-2 cell lines except this time with the exception of one isotopologue of succinate, whereas Panc-1 showed a decrease in all constituents, Figure 2-46.

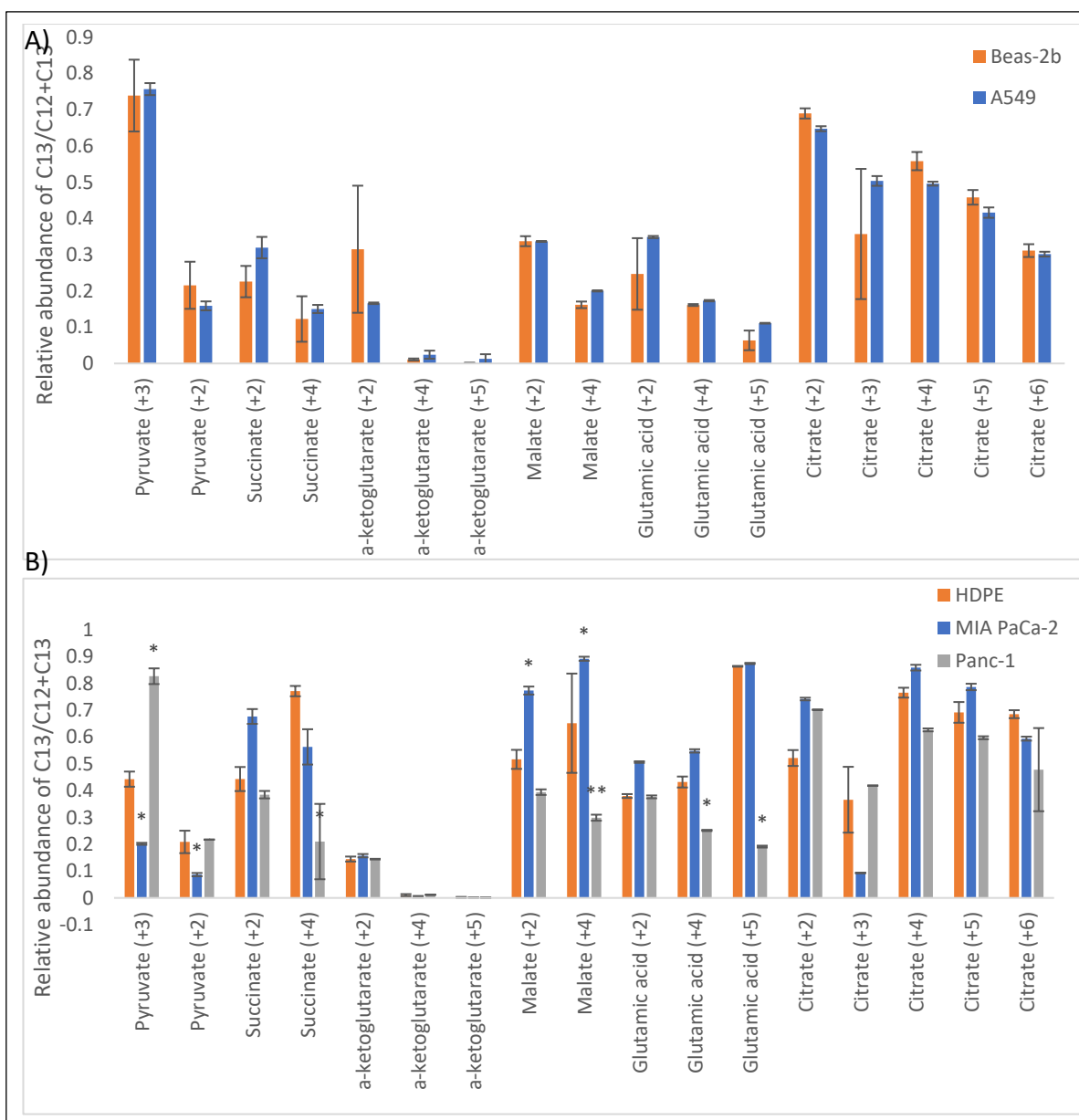


Figure 2-46- The TCA related compounds average relative abundance of labelled marker compared to unlabelled through incorporation of glucose ( $U^{13}$ ) for A) Lung cell lines and B) Pancreatic cell lines with standard error plotted and significant values  $<0.05$  indicated with asterisk.

One output of the TCA cycle, which is also an input we are interested in is as the metabolite Glutamic acid (also referred to here as glutamate). The production of glutamate was analysed using the glucose marker as shown in Figure 2-47.

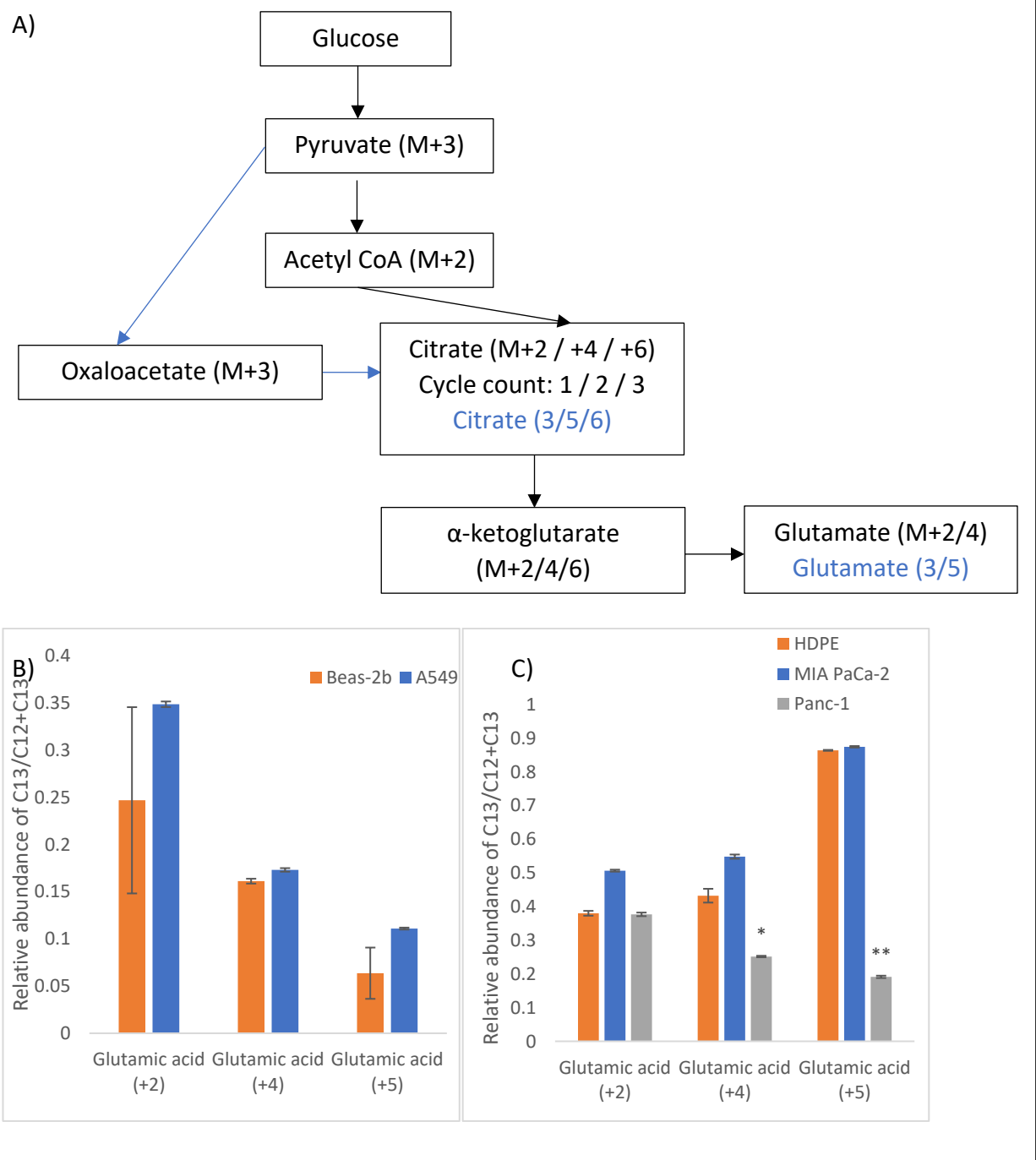


Figure 2-47- A) Illustrates the pathway for Glutamine production that can produce detected isotopologues from the carbons incorporated from Glucose<sup>13</sup> B) Illustrates the proportion of each



labelled form of glucose in comparison to unlabelled in Lung cell lines. B) Illustrates the proportion of each labelled form of glucose in comparison to unlabelled in Pancreatic cell lines.

The marker presence from glucose incorporation was found to be increased in both A549 and MIA PaCa-2 cancerous cell lines when compared to their healthy equivalents. It was however lower in Panc-1 cell lines. Though in every case this appears to be a small source of glutamic acid, indicating a lesser level of glutamine synthesis.

Next using labelled ( $U^{13}$ ) Glutamine the flux of Glutamine relating to TCA influx was analysed with fully supplemented conditions (presence of glucose) and then separately under glucose deprived conditions. Glutamine has uses as a source of reduced nitrogen as well as an alternative supply of carbon for the TCA cycle for which it is first transported into cells before conversion into glutamate utilising glutamine dehydrogenase (GLUD 1/2) in the mitochondria, the glutamate product will also carry the (M+5) label. The glutamine is also a precursor for nucleotide and lipid synthesis (Cluntun, A. A., et al., 2017) with this pathway being of most importance for more highly proliferating cancer cells. The glutamate will then be fed into the TCA cycle via transamination generating  $\alpha$ -ketoglutarate (M+5) and an amino acid by-product (which will be discussed later). From here the  $\alpha$ -ketoglutarate can enter the TCA cycle and eventually generate the (M+4) isotopologue of oxaloacetate and then the (M+4) citrate. This citrate can then either be used in fatty acid synthesis or enter another round of TCA cycle potentially forming (M+6) citric acid. An alternative fate of  $\alpha$ -ketoglutarate is through its reduction into citrate, which can be observed through the production of an (M+5) citrate isotopologue as depicted in Figure 2-48.

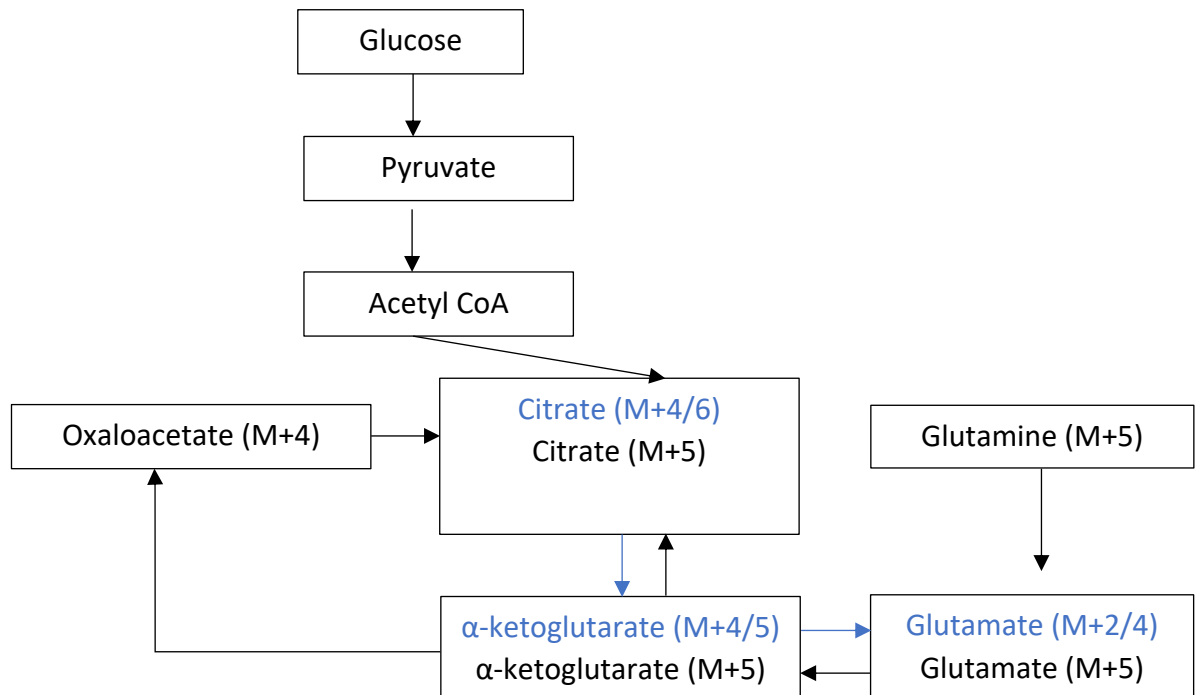


Figure 2-48- Potential isotopologue formation possible through the incorporation of uniformly labelled glutamine  $C^{13}$ , upon entry into the TCA cycle via transamination. The black labelled metabolites show the initial route of entry into the cycle, the blue show the possible down-stream isotopologues.

As such the earliest marker to represent the flux of glutamine entering into the TCA cycle can be evaluated through monitoring the initial production of (M+5) glutamate. Additionally, at this point the amount exiting the cycle as glutamate from this source was also considered through looking at the M+2 and M+4 isotopologues, Figure 2-49.

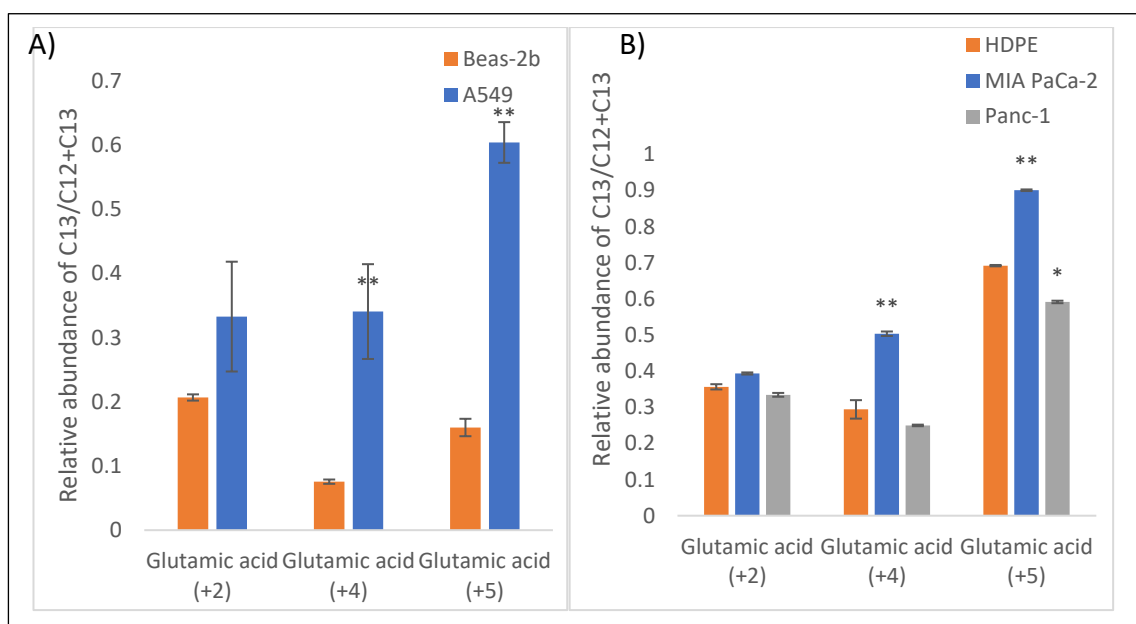


Figure 2-49- comparative labelled abundances of glutamic acid isotopologues compared to unlabelled marker presence in A) Lung cell lines and B) Pancreatic cell lines

The influx of glutamic acid, as shown by the (M+5 isotopologue), was found to be significantly elevated in A549 cells when compared to the healthy equivalent cell line Beas-2b. This was also found to also be elevated in MIA-PaCa-2 cell lines but not in Panc-1. Similarly the efflux of glutamic acid (M+2 and 4 isotopologue) was found to follow this same trend.

When removing the glucose source from the cell cultures the dependence on this glutamate source is seen to increase in Beas-2b and decrease in A549 when compared to what was shown when glucose was present. This presence in A549 is still significantly above the level of Beas-2b. When considering the impact of glucose removal in pancreatic cell lines, Panc-1 saw an increase in its proportion of incorporated glutamine increasing the percentage turnaround of Glutamic acid measured, whereas both HDPE and MIA-PaCa-2 were both lower than this level when glucose was removed from the system as a resource, Figure 2-50.

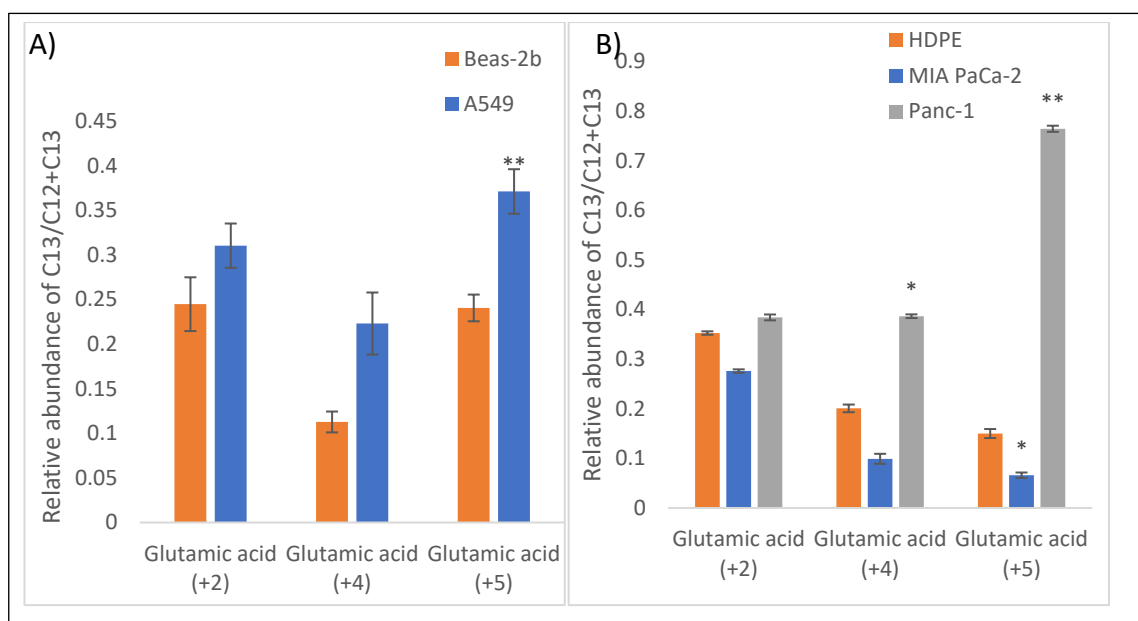


Figure 2-50- Average proportion of labelled isotopologue to unlabelled for glutamic acid in A) Lung cells and B) pancreatic cells with the absence of glucose from the media.

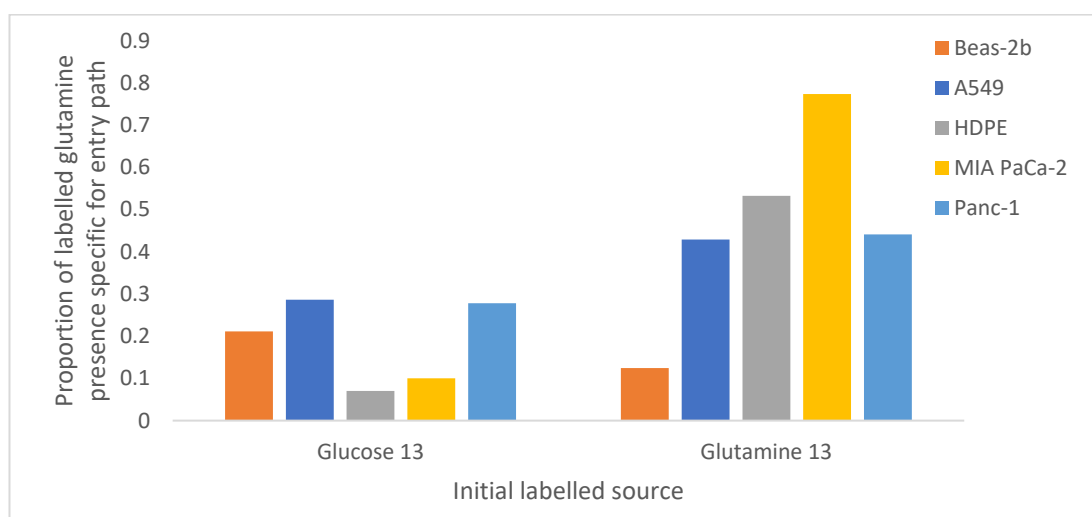


Figure 2-51-Illustrates the proportion of M+2 glucose labelled and M+5 glutamine labelled glutamic acid present during glucose and glutamine tracing experiments respectively compared to the sum of all other labelled and unlabelled markers present. Illustrating the main initial influx of each marker for the metabolite of interest suggesting whether the glutamine has come from glucose or glutamine initially.

Another way of considering the tracing experiments impact on the presence of glutamine is to look at the proportion of the first possible isotopologue of glutamic acid able to be present under each condition. For glucose labelling this is demonstrated by the presence of glutamic acid M+2 as this represents the product of turn one of the TCA cycle sourced from glucose. For glutamine tracing this will be

M+5. This is then represented as a proportion of the sum of all other possible isotopologue markers present, which in both cases is (M+ 0, 2 4 and 5). This shows a greater proportion of the metabolite has been sourced from glutamine in all cases bar Beas-2b. The utilisation of glutamine as a proportion is higher in pancreatic cell lines than it is in lung. The utilisation of glucose as a source of glutamine is higher in all cancer cell lines compared to healthy cell lines. The influx of glutamine as a proportion of a whole is higher in cancer cell lines than healthy cell lines with the exception of Panc-1, which is relatively unchanged.

Next looking at the flux through the TCA cycle leading from the (M+5) glutamic acid entry. We can follow its already discussed pathway around the TCA cycle through generation of isotopologue citrate (M+4) or its pyruvate cycle use with the generation of isotopologue (M+3). There is also the possibility to generate citrate (M+5) through the reductive metabolism of  $\alpha$ -ketoglutarate which will then largely be fed into fatty acid biosynthesis. The utilisation of the glutamine source was lesser in lung cells when compared to pancreatic and is overall far more utilised in cancerous cells than their healthy equivalents. Panc-1 is the only cell line that showed a decrease in some of these pathways, including the formation of citrate (M+2) and (M+5).

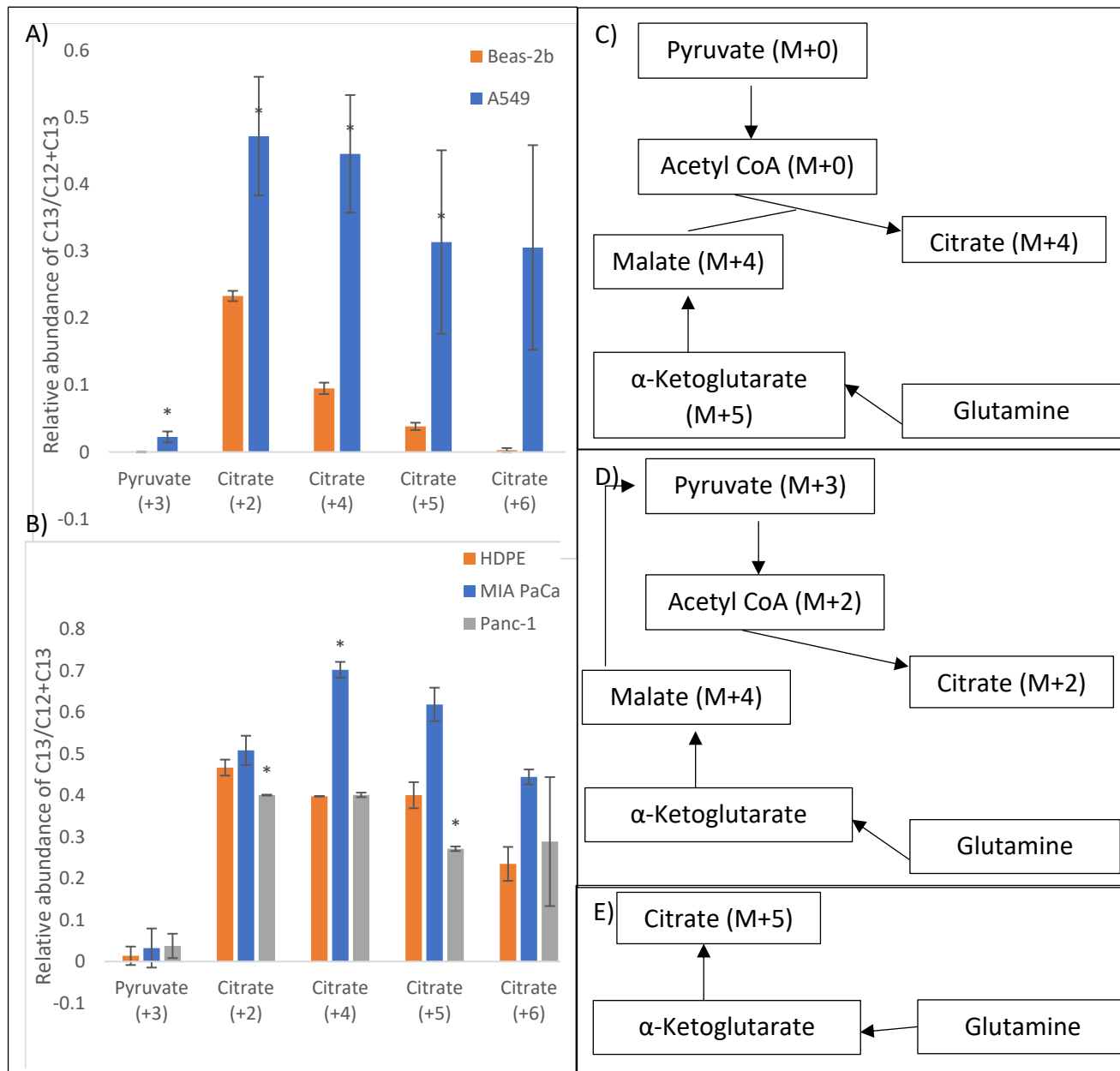


Figure 2-52- The proportional change in the presence of isotopologues of measurable TCA intermediates in A) lung cells and B) Pancreatic cells with metabolic pathways described in C) detailing the “traditional” entry of glutamine into the TCA cycle generating citrate isotopologue (M+4) B) The pyruvate salvage pathway leading from glutamine that could explain the presence of pyruvate (M+3) and citrate (M+2). E) The alternative pathway of glutamine to enable to formation of ((M+5) isotopologue of citrate.

Glutamine tracing allows a different source for pyruvate cycling that may be preferentially utilised by cancerous cells. This is evaluated firstly with glucose supplementation and then in a state of glucose starvation. Under “normal” growth conditions . Pyruvate salvage was very small in this way, less than 5% of the pyruvate was regenerated from labelled glutamine, generating pyruvate M+3, whereas the vast majority of glutamine was fed into the TCA cycle directly . In lung cells cancerous cell line had a much higher turnaround represented by the proportion of labelled citrate (M+2) being 47% compared to healthy 23% presence. In pancreatic cells Healthy cells had a similar 46% turnaround compared to MIA PaCa-2 modestly higher 51% and pancreatic cancers slightly lower 40%. The reductive metabolism was found to be higher in A549 and MIA-PaCa-2 but lower in Panc-1 than healthy equivalent, with this pathway largely feeding into lipid metabolism after generation of acetyl-CoA and then eventually palmitic acid. The lower proportion of this peak indicates the preference for the forward Krebs cycle.

When these pathways are evaluated in glucose deprived conditions, Figure 2-53, the pathways become far less utilised in lung cells whereas they become generally more utilised in pancreatic cells shown by higher overall proportions of labelled markers being seen.

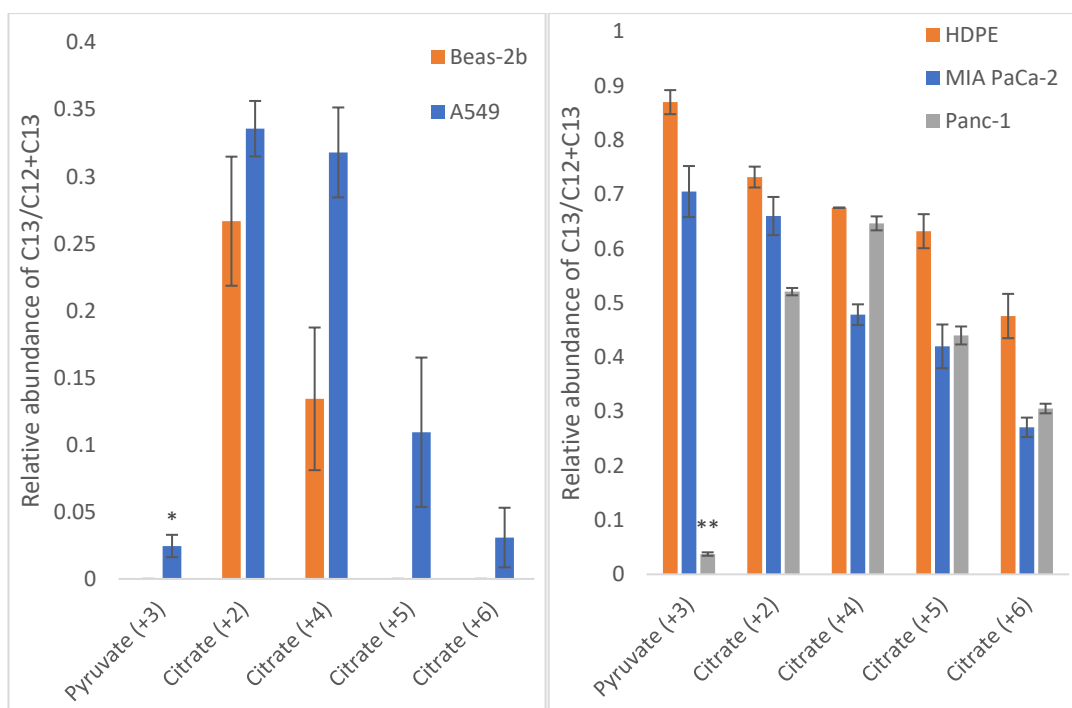


Figure 2-53 The average proportionate changes of labelled isotopologues of metabolites proportionate to unlabelled for A) Lung cell lines and B) Pancreatic cell lines.

The lung cells see a decrease in the turnaround of both entry pathways (Citrate 2 and 4) with the pyruvate salvage re-entry into the TCA cycle being the more dominant (Citrate M+2) with a 27% turnaround in Beas-2b and 34% in A549. The direct entry into TCA (Citrate M+4) shows a glutamine carbon proportion of 32% in A549 and 13% in HDPE. Reductive metabolism with generation of citrate (M+5) is exclusive A549 lung cells, with Beas-2b being absent or below the detection limit of evaluation.

Pancreatic cells show a very large increase in the pyruvate salvage pathway in both the healthy and MIA PaCa-2 cell lines with 87% and 70% labelled presence, respectively. Panc-1 has a significantly lower presence with only 4%. Of the carbons originating from a glutamine source. The TCA cycle does still show a very high amount of turnaround with 73% HDPE, 66% MIA PaCa-2 and 52% in Panc-1 with the indirect citrate M+2 isotopologue and the direct Citrate M+4 showing a labelled percentage presence of 67% ,48% and 65% for HDPE, MIA-PaCa-2 and Panc-1 cell lines respectively. The reductive metabolism pathway is comparable across all cell lines though percentage of labelled presence is slightly higher in HDPE cell lines with contributions of 63%, 42% and 44%.



Interestingly, the other candidate markers of the TCA cycle that were observed using glutamine tracing profiles, including succinate,  $\alpha$ -ketoglutarate and malate, were seen to be quite comparable between lung cancer and healthy cell lines. Mia-PaCa-2 cell lines did show a significantly increased presence in marker labelling ratios for all TCA intermediates observed whereas Panc-1 cells showed a significant decrease in succinate and malate markers.

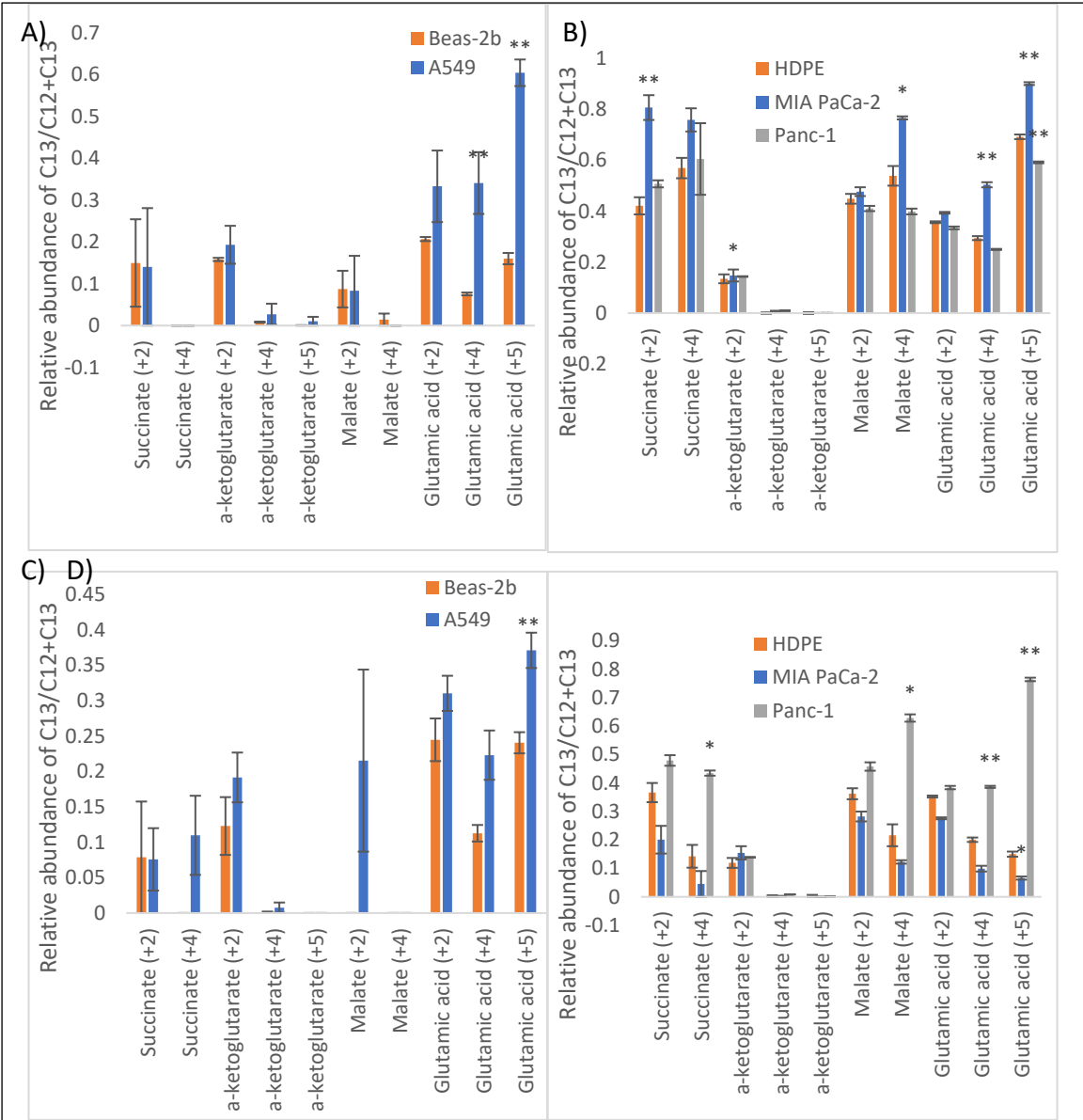


Figure 2-54- Alterations in the proportion of labelled isotopologues and unlabelled for different metabolites detected that relate to the TCA cycle in A) Lung cells supplemented with glucose and labelled glutamine 2) Pancreatic cells supplemented with glucose and labelled glutamine 3) Lung cells supplemented with labelled glutamine but no glucose 4) Pancreatic cells supplemented with labelled glutamine but no glucose

When considering the glucose deprived cell lines, Lung cells showed an overall decrease in the presence of labelled TCA candidate markers than in the cells grown with glucose available, with the exception of Malate +2 which was seen to be higher in A549. A number of markers were found to be absent/ below the detection limit in Beas-2b cell lines with the majority of the observed markers being more prominent in the cancerous line than in the non-cancerous. Pancreatic cancer cells showed a more comparative flux through the TCA cycle to that shown with glucose supplementation being available. Of interest the MIA-PaCa-2 markers were all decreased compared to both glucose supplemented MIA-PaCa-2 lines and healthy lines, whereas Panc-1 showed a number of up-regulated markers with a much more prominent succinate M+4 than previous and healthy and both malate markers were seen to be significantly more prevalent than HDPE cell lines, whereas in the presence of glucose these were seen to be slightly lower.

#### Other metabolites of interest

A possible reason for the alterations in metabolite flux being seen is due to numerous adjacent metabolic pathways that feed-out from the TCA cycle and associated metabolites. This includes a number of fatty acid, nucleic acid and amino acids. A number of the amino acids were traceable using this methodology and so can also be considered during this analysis as a way of evaluating how the alterations on metabolic demand during pathogenesis impact how metabolites might be re-routed or increased to keep up with other demands besides increased energy requirements.

Glucose tracing monitored the production of serine and glycine that carried the carbon labelling leading off from 3-phosphoglycerate during glycolysis. The proportion of serine and glycine that carried the labelled marker was significantly elevated in lung cancer cell lines when compared to healthy, whereas in pancreatic cell lines showed a significant decrease in serine and a slight decrease in glycine,

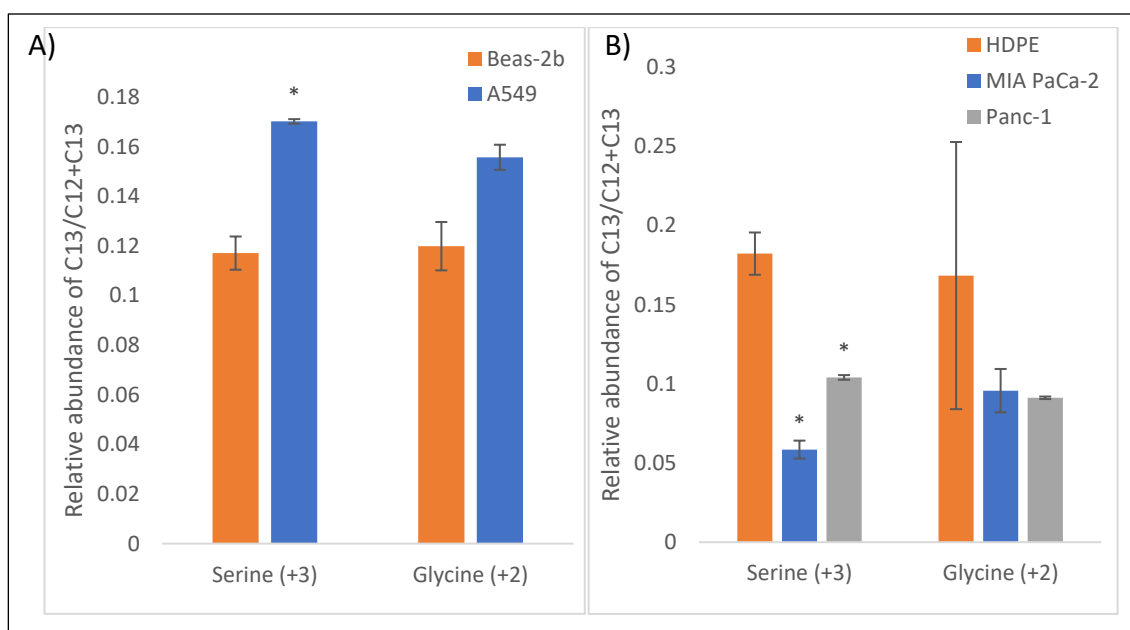


Figure 2-55- The variation in labelled to unlabelled ratio of isotopologues of serine and glycine generated from labelled glucose metabolism in A) lung and B) Pancreatic cell lines

The amino acids listed in, Figure 2-56 all showed an increase in labelling carry down from glucose in A549 and MIA-PaCa-1 cell lines when compared to their healthy counter parts. Panc-1 showed comparable carry over of labelled: unlabelled carbons to HDPE healthy lines with the exception of aspartic acid that showed a decrease.

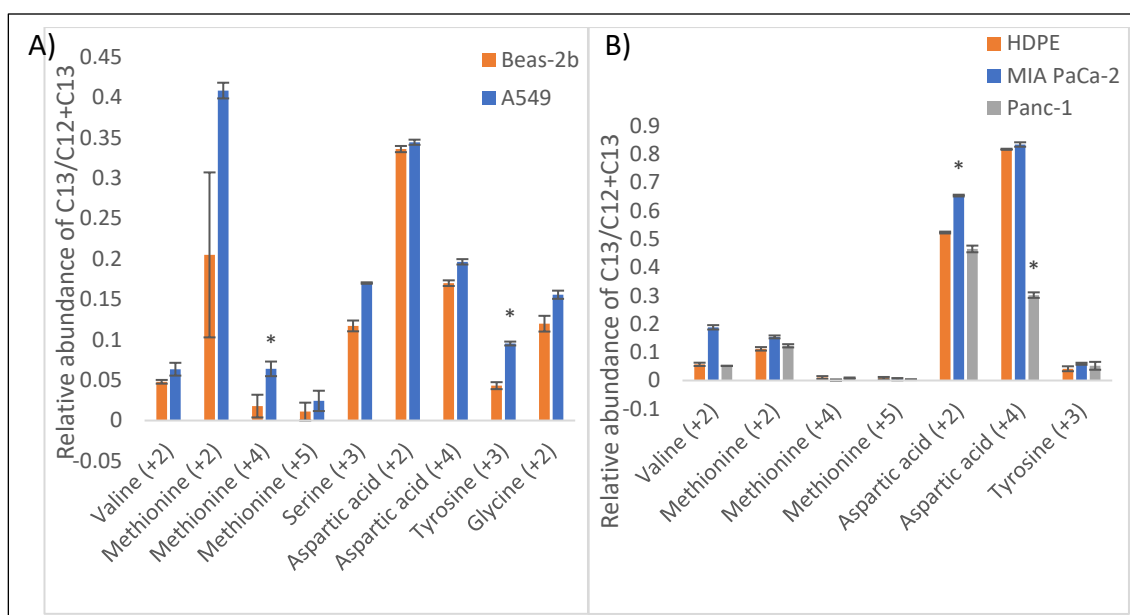


Figure 2-56- Average proportion of labelled to unlabelled isotopologue of amino acids after introduction of glucose labelled media in A) Lung cells and B) Pancreatic cells with standard error shown and significant variance to healthy equivalent cells shown with asterisk.

## 2.5. Discussion

This chapter has demonstrated the use of a range of mass spectrometry based analysis to cover a broad range of metabolic markers both in cellular composition and waste products, it is hoped that a better understanding of aberrant pathways will be encapsulated and the discovery of new biomarkers that can prove most effective for quick and easy screening/ confirmed diagnosis of pancreatic cancer sufferers.

This study utilised cancer cell-line profiles to evaluate a range of markers using GCMS and MALDI based techniques as well as using urinary markers via HPLC.

The first two aims of this chapter were to identify novel biomarkers that were capable of distinguishing pancreatic cancer cell lines from healthy pancreatic cell lines. These were then to be compared to markers found to distinguish lung cancer and healthy lung cells from each other and draw discussions on which markers were conserved/ exclusive for the different cell types. This section of the analysis used multiple

different strategy in an untargeted manner in order to capture as broad a range of metabolic markers as possible.

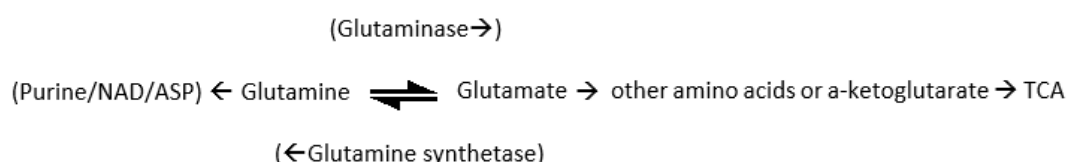
## GCMS

The GCMS analysis was used to analyse six hundred and forty-three metabolite markers that were extracted and analysed for statistical variation in the study comparing the global untargeted and amino acid targeted profiles between non-cancerous and cancerous cell-lines that included two forms of pancreatic cancers and one form of lung cancer cell line. Whilst the targeted analysis provided a comparative quantitation between the comparative groups, the absolute amount present in each sample could be subject to matrix effects as standard curves were prepared in water solutions rather than an equivalent artificial matrix that could introduce extraction inefficiencies.

Comparisons were then made to note any exclusive and shared differences between the different forms of cancers and their healthy equivalent lines. A number of interesting metabolites, mainly amino acid based, were found to fluctuate between the cell lines;

Leucine was found to be up-regulated in pancreatic cancer cell lines but significantly down-regulated in lung cancer cell lines during the study. Leucine is an essential amino acid and is one of the most utilised in the body with roles in protein synthesis signalling molecule of rapamycin and has been found in other literature to be one of the most utilised amino acids in cancer cell lines (Kang J. S. 2020).

## Glutamate & Glutamine



Glutamine is a conditionally-essential amino acid and one of the most abundant of all amino acids found in human plasma. It is additionally one often monitored in cancer studies due to its diverse roles in cellular metabolism and specifically to

tumorigenesis neoplasia, disease progression and metastasis (Yang, L., et al., 2016). During this study it was found to have a greater presence in all cancer cell-lines under analysis, most obviously MIA PaCa-2. Glutamine is typically taken into the cell by glutamine transporters in the plasma membrane, it can also be synthesised from glutamate using glutamine synthetase but this is a minor source.

In contrast to this, glutamate is considered non-essential and is typically found to be quite low in abundance in healthy cells and is not taken up directly in larger amounts in cancerous cells, instead it is an intrinsic product most commonly formed from glutamine via glutaminase (Choi, B. H., & Coloff, J. L., 2019). The oncogene cMYC is responsible for the activation of glutamine transporters and additionally increases the expression of glutaminase (Bott, A. J., et al., 2016). cMYC also positively regulates LAT1 which will be discussed in more detail later in this section. cMYC is known to be up-regulated in numerous cancerous cell lines causing the expression of LAT1, glutamine transporters and glutaminase to all be up-regulated as a result mirroring what is being seen in this study. Additional studies have shown that the level of LAT1 expression in lung cancer cell lines is less than that of pancreatic. This may hold weight in explaining the difference noted in this area between lung cancer and pancreatic cancer cell lines as if there is less cMYC expression in lung cells there will then be less glutamine taken up into the cells and therefore less substrate available for conversion into glutamate. In addition due to the cMYC alteration there will also be less enzyme, glutaminase, available for this process, further impacting the amount of glutamate being produced (Zhang, J., et al. 2020, Kaira K., et al., 2008 & 2012).

Glutamate is an important metabolite due to its requirement in the biosynthesis of cysteine, glycine, asparagine and arginine as well as the formation of  $\alpha$ -ketoglutarate which is then fed into the TCA cycle as an energy source. In this study the glutamate was found to be up-regulated in the pancreatic cancer cells lines and yet down-regulated in lung cancer cells when compared to the equivalent healthy model, both of these findings are supported in other studies (Kamphorst, J. J., et al., 2015).

Possible additional explanations for the lower levels of glutamate seen in lung cancer cells can be attributed to the rate of glutamate utilisation, especially the high demand

in the TCA cycle, does not keep up with the biosynthesis in lung cancer cell lines. An alternative reason links with the higher levels of cysteine noted in both lung and pancreatic cancer in this study, most prominently MIA PaCa-2. Cysteine can be transported into the cell in exchange for a glutamate via the xCT transporter when demands are particularly high. This can in turn cause a shortage of glutamate, which could be the case in lung cancer cells (Choi B.H., & Coloff J.L., 2019). Glutamine can also be intrinsically synthesised using glutamine synthetase which converts glutamate with ammonia to form glutamine, driven by ATP (Yang L., et al., 2016). In cancer cells glutamine synthetase expression has been found to be upregulated increasing the levels of glutamine and asparagine, nucleotides and enhances the uptake of essential amino acids, also largely found to be up-regulated in pancreatic cancer cells.

### Glucose

Glucose was found to be significantly more abundant in A549 and MIA PaCa-2 cell lines than their healthy equivalent cell lines (Panc-1 showed no significant variation). Increased glucose metabolism is a key feature of many different cancers, including lung and pancreatic. The **Warburg effect** is largely accountable for this as a key metabolic switch aiding upregulation of glycolytic pathways to support the new and higher energy demands of highly proliferating cancer cells and incorporate more nutrients for additional growth and biomass where alternatively in healthy cells this energy was largely provided by the more efficient mitochondrial oxidation. The main end goal of this effect in cancer cells is the increase of glucose uptake and the increased production of lactate and other downstream metabolites. Under healthy conditions, aerobic metabolism of glucose to pyruvate via the TCA cycle- minimal production of lactate (36 molecules of ATP/glucose). Cancer cells, require a boost via anaerobic pathway, produces large amounts of lactate via a “fermentation” pathway (2 molecules of ATP/ glucose). Though less efficient allows a faster route of biomass accumulation, better facilitating high proliferation. This is a key adaptation to support the biosynthetic requirements for uncontrolled proliferation through the de-novo generation of lipids, nucleotides and proteins required (Liberti, M. V. & Locasale J.W., 2016; Vander Heiden, M. G., et al., 2009). This boost in glucose requirement calls for

more enhanced presence of glucose transporters in cancerous cell lines. Various studies have found such transporters, such as Glut 1, to be significantly elevated in cancerous cell lines including lung and pancreatic (Davis-Yadley, A. H., et al., 2016).

#### Serine/ Glycine

Serine was found to have a significantly elevated presence in both pancreatic and lung cancer cell lines in comparison to their healthy equivalent. Glycine was found to be significantly up-regulated in MIA PaCa-2, up-regulated in PANC-1, same in A549. The Warburg effect is also accountable for these changes noted in serine levels. Through diversion of glucose carbons to other pathways including serine biosynthesis. Serine can also be taken up by the cell via transporters or synthesized via a branching pathway from glycolysis from 3-phosphoglycerate via phosphoglycerate dehydrogenase enzyme and eventually is converted into glycine via the same pathway via serine hydroxymethyltransferase 1 (SHMT1). Serine is largely utilised in protein, purine, thymidine, choline and methionine synthesis (Ma, X., et al., 2019). Glycine metabolism and Glycine decarboxylase enzyme have been shown to drive tumorigenesis. Alterations in GLDC expression can impact changes to glycolysis and glycine metabolism impacting pyrimidine metabolism and cell proliferation. GLDC catalyses glycine catabolism to CO<sub>2</sub>, NH<sub>3</sub> and 5,10- methylene-tetrahydrofolate. GLDC regulates glycine metabolism. GLDC promotes serine synthesis/uptake. Lung and pancreas overexpress GLDC. GLDC promotes growth Zhang, W. C., et al., 2012.

Overall, a number of these amino acid markers can be utilised in giving a clear patient diagnosis for pancreatic cancer presence, with some also holding promise for future development in to a prognostic marker. The downside of GCMS analysis as a tool for diagnosis is the requirement for a more invasive sample source and a more intense sample preparation and analysis than some of the other strategies used. This could however, be balanced against the presence of such strong diagnostic biomarker and so would be more suited as a solid tool for diagnosis than one of screening.



**Negative mode**

Abnormal lipid metabolism has known association with numerous different cancers and has provided useful markers in other studies. In this study negative mode analysis has been used to analyse such lipidomic profiles, with a focus on glycerophospholipids and free fatty acids. Glycerophospholipids are a major component of cells, with most notable presence in cellular membranes and as such fluctuations can have impacts on various features including phagocytosis, fluidity, and signalling. Multiple PI, PE, PS and PC lipids were found to be significantly altered in the cancerous cell samples, the main result of these changes is drastically altered membrane compositions and therefore properties. Almost all markers of interest were found to be significantly up-regulated in the cancerous cell lines when compared to its equivalent healthy candidate cell lines. The up-regulation of these illustrates an increase in lipid synthesis, particularly with the increase in PE and PI noted. These specific sub-classes are of interest due to their known involvement in various cellular processes that better enable the cancers progression.

The majority of the significantly different markers were given tentative identification of class PC, with all of these markers showing up-regulation in the cancerous cell lines. Phosphatidylcholine (**PC**) is a glycerophospholipid that is utilised heavily in the cell membranes, being the most abundant glycerophospholipid, across the lipid bilayer. PC also acts as a precursor to other lipids SM and PE. Its increase in cancer cells is well documented and typically associated with a need to support high rates of proliferation (Fu, Y. et al., 2020; Koundouros, N., & Poulogiannis, G. 2020; Lima, A. R., et al., 2022). The Up-regulation of PC is also associated with various other pathways that result in greater energy accumulation supporting higher rates of proliferation. Additionally, with its utilisation in membrane structures, with highly proliferating cell types/ cancers there is an expectation for a higher level of PE/PC presence overall. Furthermore, cancer cell membranes undergo modifications for altered functioning, which can result through alterations in phospholipid metabolism such as those observed in this study.

Very high levels of fatty acid synthase have been found in cancerous cell lines, including lung and pancreatic (Fazli, H. R., et al., 2021; Swinnen, J. V., et al., 2003). This enzyme is utilised in the synthesis of palmitate from acetyl-CoA and malonyl-CoA utilising glucose and glutamine. This enables cancer cell to produce more phospholipids in order to build membranes in supporting enhanced cell growth and proliferation and modulation of membrane fluidity and functioning. The up regulation of lipids is not only by enhanced synthesis but also an increased level of fatty acid uptake. Provide an energy source during metabolic stress. Generation of secondary messengers that modulate tumour progression (Koundouros, N., & Poulogiannis, G. 2020).

**PS** phosphatidylserine is typically associated with the inner leaflet of the cellular membrane, but it can be externalised in some types of cancers (Desai, T. J., et al., 2016). Up-regulation of PS can cause a decrease in the immune response through its exposure, resulting in protection from NK cells and other cytotoxic immune cells. The degradation of PS results in PE formation. Phosphatidylinositol (**PI**) is located in the plasma membrane and endomembrane systems and has an important function in exocytosis. PI enables the formation of GPI-anchors, enabling motile protein membrane interactions and act as a regulator of membrane binding proteins as well as contributing to membrane fluidity. Additionally it can alter G- protein mediated signalling pathways (Lima, A. R., et al., 2022). Furthermore, PI is additionally a source of lipid second messengers in activating the Akt signalling pathway (Marien, E., et al., 2015). **PE** phosphatidylethanolamine second most common membrane phospholipid. PE partakes in various roles within the membrane including structurally, membrane curvature secondary messengers and folding of membrane proteins. Positive regulator of autophagy prevents the accumulation and uncontrolled degradation are underlying cause of carcinogenesis.

Overall, the vast differences seen in lipidomic profiles in cancerous cell lines show extremely strong promise for use as diagnostic markers for pancreatic cancers as well as holding potential for other cancerous types.

### **Positive mode**

MALDI provided a successful way of conducting an untargeted analysis on a large number of different metabolite profile markers extracted from cell line culture in a fast and effective way. Discriminatory markers in various cell lines that were capable of distinguishing cancerous and non-cancerous groups from one and other with a high level of accuracy were highlighted and discriminated using ROC analysis to ensure their capability as individual biomarkers.

Lung cells showed a panel of lysophospholipids that were found to be significantly less abundant in cancerous cell lines when compared to healthy and all individually provided better than good ROC scoring. Phospholipids are a major component of cell membranes and have important roles in energy and cell signalling pathways, illustrating a likely variation of these metabolic pathways. Sphingolipids are constituents of cell membranes, lipoproteins and rafts which are synthesised within the cell and function as messengers mediating cell apoptosis and modulate functioning in the signal transduction network (Wang, S., et al., 2016; Sommella, E., et al., 2022; Lima, A. R., et al., 2022). Modulators of signal transduction. **SM** was found to be decreased in lung and increased in pancreatic cancer cells compared with healthy. SMase breakdown SM into ceramide which functions as a second messenger to induce apoptosis. Decrease in SM through activation of SMase by the metabolic product of arachidonic acid (Sommella, E., et al., 2022). Elevated SM levels through SMase alterations have been found to be elevated in a number of different conditions including diabetes, lysosome storage disorders and cancers (Hong, J. H., et al., 2016; Jones, E. E., et al., 2014). Alternatively, ceramide could obtain a PC head to generate SM, as such the downregulation of SM seen may be through a low level of ceramide which in turn will affect the apoptosis (Wang, S., et al., 2016)

Pancreatic cancer's most successful metabolic marker was Cer(d34:1) which was found to be significantly less abundant in both of the cancer lines under evaluation and equally maintained a better than good ROC scoring in both instances. Panc-1 cell lines had two other individually strong biomarkers LPG(17:1) and Cer(20:1) both of which were found to be more abundant in cancerous lines when compared to healthy. The activity of SMases in pancreatic cancer worth consideration in future studies (Jones, E. E., et al., 2014). **Ceramide** has been shown to be involved in cancer

signalling pathways as well as cell proliferation involvement including apoptosis. ACDase cleave ceramides into fatty acids and sphingosines, ACDase's are known to be overexpressed in a number of cancers including pancreatic and lung (Jones, E. E., et al., 2014) accompanied by a higher level of ceramides. Membrane and signalling apoptosis

### **Fatty acids**

Tumorigenesis has been associated with an increase in *de novo* fatty acid synthase activity, this includes SCD1 ( converts saturated into monounsaturated) and FASN (acetyl CoA/Malonyl CoA into long chain palmitate). Markers for dysregulation of these enzymes include an increase in PC(36:1) and decrease in SM(34:1), with both being associated with overexpression of SCD1. This can lead to an accumulation of fatty acids , which can then modulate the formation of phospholipids supporting further cell proliferation and signalling (Wang, S., et al., 2016; Guo, S., et al., 2014).

The polyunsaturated fatty acids also found to be dysregulated in cancerous cell lines can have broader scope impacts through the activation of other fatty acid desaturases (Guo, S., et al., 2014). **TG's** were found to be significantly lower in pancreatic cancer, which is in agreement with other cancers in research through the use of TG's in their degradation to generate energy to support membrane synthesis and rapid proliferation (Sommella, E., et al., 2022)

Whilst there was a good degree of similarity of profiles relating to the markers of interest within cancer cell line types, there were a number of variailities noted between the two pancreatic cancer cell lines profiles which emphasises a need to better explore the diversity of pancreatic cancer through looking at more cell lines/ clinical samples to better understand the unique feature of the different cell lines and what the difference in markers may mean both diagnostically and prognostically and how these markers may influence the cancers phenotypical traits.

Furthermore with the unique traits found between the different cell lines can these be used to distinguish specific pancreatic types/ stages/ phenotypes in further studies.

Overall, a significant alteration in amino acid presence observed in this study signals a change in the intake/ biosynthesis and metabolism of these metabolites in order to support the high cellular proliferation rates.

LAT1 imports essential amino acids and is known to be up-regulated in a number of highly proliferating cancerous cells including lung and pancreatic promoting the cells proliferation and survival needs (Kaira, K., et al., 2012 & Zhang, J., et al., 2020). In pancreatic cancer its presence is also suggestive of the aggressive nature of the disease. MIA PaCa-2 was found to have a significantly more elevated level of these essential amino acids and is known to have a shorter doubling time and therefore a greater rate of proliferation. This aggressiveness of the cell lines may explain the more pronounced alterations noted in the profile.

In some instances in the lung cancer cell line the presence of amino acids were seen to decrease, though literature suggests that the cancerous cells should still also have a more increased level of LAT1 amino acid transporters present, cell type is still lower than pancreatic cancer, possibly due to its better blood supply supporting nutrient transition and as such the level of amino acid in-take may not equal the amount being used and hence seen as a net decrease in the cellular profile (Kaira K., et al., 2012 & Zhang, J., et al., 2020).

The next aim of this chapter was to use targeted analysis to seek modified nucleoside alterations in urine samples from pancreatic cancer patients. Modified nucleosides are nucleic acid breakdown products that can act as a RNA turnaround/ degradation measures through their excretion rates due to their inability to be re-used as their un-modified counter-parts and typically their excretion in healthy candidates is minimal. This study has checked how usable these markers are in pancreatic cancer diagnosis using urine analysis comparisons between healthy, PDAC, pancreatitis and other pancreatic conditions. This study successfully identified thirteen different modified nucleoside makers which were then able to be utilised in a semi quantitative

analysis based on relative abundances in comparison to the presence of creatinine. This was achieved with a methodology that required minimal pre-sample preparation reducing cost and time required for analytical processing. There were no significant differences reported in these modified nucleoside abundances when comparing males and females urine profiles. Overall, all modified nucleosides were seen to be up-regulated in PDAC when compared to healthy cell lines, 6 of which significantly so. A significant increase in the presence of modified nucleosides were also seen to distinguish pancreatic cancer from pancreatitis, other cancers and other pancreatic condition groups. This reflects other previously studied cancers including lymphomas (Rasmuson, T., & Björk, G. R., 1995), Liver cancer (Yang, J., et al., 2004). This increase is indicative of an increased turnover rate of tRNA in cancerous bodies than in healthy. The level of increase seen does vary between different nucleosides. There is also a noted general increase of modified nucleoside presence and tumour stage, though again this is not always the case and varies between different nucleosides. Though these markers did not achieve good enough biomarker qualities as individual or as a panel of the detected markers. However, the trend still shows potential going forward to build upon to find a successful markers. The main issue found with the biomarker analysis was the presence of false negative values in the ROC analysis leading to insufficient diagnosis potential of all markers analysed. Even when used in a panel or artificial neuronal network classification applied this was still a consistent flaw in the diagnostic potential of the modified nucleoside markers. N2,N2,2'-O-trimethylguanosine was the most promising individual biomarker found for diagnosing PDAC producing AUC scoring of 0.742, LHR+ 1.89 and LHR- of 0.39 with a sensitivity of 90% and specificity of 53%, but again these are all under the acceptable threshold limit, followed by 1-methylguanosine and dimethylguanosine. This reflects findings in (Seidel, A., et al., 2006) on a panel of different cancers that also found these to be the most successfully employed modified nucleosides for differentiation of the cancerous and non-cancerous groups. This high level of sensitivity shows promise for a cheap and accessible screening tool to identify those most at risk. When using a panel of all variations of modified guanosine nucleosides a sensitivity of 70% with a specificity of 58% was achieved slightly improving this.

Overall whilst modified nucleosides have been shown to show potential as a screening diagnostic tool to identify as many at risk candidates as possible they lack in ability for use as independent diagnostic values in a fast, high through-out and non-invasive method and can also give a potential indication of disease progression status.

In order to establish a broader understanding of the underlying processes that occur during pancreatic cancer and identify those changes that may be of use in a prognostic capacity in future studies, stable isotope labelling was utilised. This allowed the tracing of pathways leading from glucose and glutamine metabolism that were altered in cancerous cell lines. Cancer cells are understood to have altered metabolism from that of healthy cells through the increased metabolic demands and energy requirements needed to support an increased rate of cellular proliferation accompanied by altered survival constraints such as nutritional deprivation. Both glucose and glutamine are key nutritional resources acting as precursors to the citric acid cycle (TCA) utilised by cancer cells to help alleviate these problems. This study aims to establish how lung and pancreatic cells utilise these resources and establish which mechanisms are shared and which are unique to the different cell lines. Utilising this information in conjunction with the previous cell metabolite profile analysis to identify the most prominent pathways used including under glucose depleted extremities, how this impacts the different pathways that were found to be altered from that of healthy cells. The previous report regarding metabolic changes in these cell lines, normalising to protein content showed a significant increase in the presence of glucose and glutamine in all cancerous cell lines when compared to healthy equivalents, with the exception of Panc-1 which was relatively unchanged for glucose content. To better understand what this increased presence means in terms of what pathways are being utilised tracing experiments labelling both of these metabolites have been utilised along with analysing the alterations seen under strained glucose deprived conditions.

Rather than looking at net metabolic presence this study evaluated the metabolic turn-around of individual key markers that relate to the metabolism of glucose and glutamine with focus on their entry into the TCA cycle.

### **Glucose U<sup>6</sup> tracing**

Pyruvate was the earliest marker derived from glucose metabolism analysed during this analysis, being the end product of glycolysis accompanied by the provision of cellular ATP. The unit rate of production was found to be increased in only pancreatic cancer cell lines, Panc-1, when compared to healthy counterparts. Lung, A549, cells were found to be slightly higher but relatively unchanged and surprisingly MIA PaCa-2 cell lines were found to be significantly lowered. This surprising decrease in labelled proportion seen in Mia PaCa-2 cells can be explained by the cells having an overall increased presence of unlabelled glucose initially and could be metabolising glucose at the same rate as other cells the ratio will still appear lowered. Another possibility is that there is an alternative unlabelled source of pyruvate being utilised which effectively “dampens down” the proportion of labelled marker seen in comparison. The intake of pyruvate is instantly being utilised by the cell on the other pathways or the citric acid cycle is being supplemented primarily by a different source altogether. The glucose may be being used for other things prior to the end of glycolysis i.e. nucleotide production and feeding out at 3-PG feeding into serine/ glycine and so is being directed in another route other than glycolysis and pyruvate formation. In either case it can be attributed that MIA-PaCa-2 cells have a lowered overall requirement/ proportion of glucose sourced pyruvate per unit time than other cell types in this study whereas A549 and more so Panc-1 have a much higher demand for glucose sources of pyruvate.

Relating this to our previous findings that glucose had an increased presence in A549 cells accompanied with the increased pyruvate formation makes case for an increased net intake of glucose in cancerous lung cells overall from that of healthy lung cells. Panc-1 has a high demand for pyruvate derived glucose, the previous data found unchanged glucose presence to healthy cells which overall suggests that Panc-1 cells must take the glucose in at a higher rate in order to maintain its glucose store whilst being able to generate more “product”. Mia-PaCa-2 lines instead have a higher content and resource of glucose as found in the previous study and potentially



additional high throughput methods for gaining pyruvate other than glucose source utilisation. This already illustrates a broad difference in the metabolic alterations observed between cancer cells and healthy cells as well as differences between different cancer cells to one and other.

The next stage addressed the four different fates for the pyruvate produced at the “Pyruvate junction”, these include:

- 1) NADH reduction to lactate
- 2) Glutamine mediated amination to alanine
- 3) Carboxylation to oxaloacetate
- 4) Decarboxylation to acetyl-CoA

Lung cancer cells showed an increased metabolite flux towards **Lactic acid** than pancreatic cancer cells did when glucose was the metabolite source. It was also found to be overall more prioritised in cancerous cell lines than in non-cancerous with the exception of MIA PaCa-2 which was unchanged from healthy lines, though as before this could be due to the presence of an alternative source of unlabelled pyruvate.

All (healthy and cancerous) cells when proliferating convert as much as 85% of their glucose into lactate via aerobic glycolysis, when resting this rate is dramatically reduced in the cytosol catalysed by lactate dehydrogenase. Lactate levels are known to be more increased in numerous different forms of cancer including lung and pancreatic (Cai, M., et al., 2022), this preferential shunting of pyruvate to lactate in this manner is known as the Warburg effect. Otto Warburg evidenced cancer cells increased their uptake of glucose and production of lactate even in the presence of oxygen (Warburg O, et al.. 1927). Cancer cells use this pathway for multiple reasons and is seen to be particularly increased in more aggressive cancers (de la Cruz-López, K. G. et al., 2019). The lactate acid generated needs to be removed from the cells with  $H^+$  ions in order to prevent the acidification of cells. Additional to the internal pH homeostasis this movement helps to establish and optimise the tumors microenvironment through manipulation of the pH gradient. Lactate is excreted from

the cells via monocarboxylate (MCT1,4) transporters removes protons from GA3P dehydrogenase reaction maintaining internal pH and acidifying the microenvironment- better promotes tumour growth, immune cell evasion, invasion and metastasis promotion (Pérez-Tomás, R., & Pérez-Guillén, I. 2020; Cai, M., et al., 2022). This is achieved through its ability to suppresses the proliferation and cytokine production of human cytotoxic T lymphocytes inhibiting their actions against cancer cells. In addition lactic acid can also reduced the cytotoxic activity of CTLs, induce the apoptosis of natural killer cells and effect dendritic cells activity relating to immune stimulation (Fischer, K., et al., 2007).

The actual generation of lactate via NADH dependent reduction enables the recycling of 2x NAD<sup>+</sup> which enables the continuation of higher levels of glycolysis through recycling a critical factor, Furthermore the reaction is reversible through lactate exchange and so if more pyruvate was required this lactate could pose as a store for latter pyruvate generation (Cai, M., et al., 2022; Pérez-Tomás, R., & Pérez-Guillén, I. 2020).

These reactions are catalysed by Nictotinamine adenine dinucleotide oxidoreductase LDH enzyme as follows:

LDHA: pyruvate (NADH) → lactate (NAD<sup>+</sup>) reduction

LDHB: Lactate (NAD) → pyruvate (NADH) oxidation

Alanine is another potential product stemming from the pyruvate junction to consider under this analysis. This pathway stemming from glucose when compared to the presence from other sources was again found to be higher in lung cells when compared to pancreatic cells and overall more prevalent in cancerous cells when compared to their healthy equivalent with the exception of MIA PaCa-2 which again was found to be significantly lower.

Pyruvate is converted into alanine utilising alanine aminotransferase or glutamic pyruvic transaminase (GPT2) via amination reaction, this can occur in conjunction to glutamine entering the TCA cycle when it becomes α-ketoglutarate. This pathway is

considered to be an insignificant source for intra-cellular alanine, as the majority will be imported from the extracellular fluid, this source of alanine has a tendency to be utilised as a carbon source for the TCA cycle as it is able to be converted the opposite way, back into pyruvate (Otto A. M. 2020; Sousa, C. M., et al., 2016)

This increase in alanine production may be as simple as the reduced competition for pyruvate by pyruvate dehydrogenase etc due to the increased availability of the source in the cells or alternatively a reduced requirement for conversion to acetyl Co-A due to increases in glutamine supplementation, to be discussed later, but may also be a more heavily utilised pathway due to additional requirements due to increased rates of proliferation. Research has shown the pancreatic cancer cells have become reliant on this internal source of alanine production to keep up with the increased cellular demands and alanine reserve preservation (Parker, S. J., et al., 2020). These demands arise not only through its use in anabolism but is also suggested to drive extracellular matrix formation through the co-production of  $\alpha$ -ketoglutarate, which was attributed in breast cancer cells (Vettore, L., et al., 2020). An alternative hypothesis proposed that the alanine will now be in excess and as such will be excreted into extracellular space. From here the alanine could be taken up by neighbouring cells which may then secrete pyruvate out as it will be less required making this substrate available for the cancer cells (Vettore, L., et al., 2020).

The final uses of pyruvate derived from glucose under consideration involve two slightly different entries into the TCA cycle, via acetyl-CoA or oxaloacetate generation. The TCA cycle paths the most energy efficient utilisation of glucose for the cell.

Both lung and pancreatic cells show similar glucose sourced product metabolite markers for each of these pathways with the acetyl-CoA entry being the most prevalent with a 55-80% turnaround seen, whereas the OAA pathway had a lesser 10-50% turnaround. The major pathway into the TCA cycle is the most efficient for energy production, hence its heavier utilisation. This pathway was only seen to

significantly increase in pancreatic cancer cell lines when compared to healthy equivalents.

The minor pathway involves the carboxylation of pyruvate into oxaloacetate via an ATP- dependent pyruvate carboxylase which is primarily utilised as a mechanism of replenishing TCA cycle intermediates alongside glutaminolysis. Studies have shown that the presence of pyruvate carboxylase is increased in lung cancer cells, particularly early stage, and the presence impacts tumorigenesis and proliferation. Knockout's impact TCA cycle sustainability, nucleotide and lipid biosynthesis and redox homeostasis.

This TCA cycle upregulation is also supported by increase in other intermediates being traced though this is not the case with all of Panc-1's intermediates with a number of glucose derived metabolites being decreased in presence compared to the overall pool. This could mean a greater amount of output from TCA cycle intermediates to other pathways such as fatty acid or amino acid synthesis or reflect the intake of other unlabelled sources of TCA intermediates.

Fatty acid related metabolites were not analysed during this study but some amino acid markers were. Serine and glycine were found to be increased in the proportion of glucose derived production in lung cancer cells, whereas pancreatic cancer cells found an overall decrease when compared to their healthy equivalent lines. The previous study showed that both of these metabolites were found to be increased in overall abundance in cancerous cell lines when compared to their healthy counterparts. This suggests that whilst the lung cancer cells depend more on glucose derived sources to keep up with metabolic demands for these amino acids, pancreatic cancer cell lines utilise other modes of production more than the glucose source, or have larger stores of the metabolites meaning the influx is not needed to vastly increase to keep up with the increased demand.

All other amino acids under study were found to increase in the proportion of glucose derived presence in cancerous cell lines than that of healthy. These amino acids were derived after the point of TCA entry.

Overall the glucose analysis has shown the vast differences in metabolic feed-through from glucose for all of the different cell lines under study with differences between different cell types and even different cancers of the same cell types and could therefore prove to be a useful method not only to determine if cells display malignancy but also the potential to specifically differentiate different cancers impacts and profiles from one and other and possibly with more research define levels of aggressiveness or hold prognostic marker potential. There is however such a vast alteration in profiles seen that its suitability as a diagnostic tool is more limited and better suited to one of a more distinguished profiling. Additionally, more research into other healthy and unhealthy examples of cells will be needed to best establish the key markers capable of giving clear diagnosis patterns.

Relating to our specific evaluation's lung cells seem to prioritise glucose sources for their pyruvate generation via glycolysis with less coming from alternative sources than pancreatic cancer cells do and even more so in cancerous lines. The lung cells have a shorter doubling time than pancreatic cells and so this may impact the way energy is acquired for the cells. Alternatively pancreatic cell lines show a lowered reliance on pyruvate generated from glucose for entry into the TCA cycle most likely due to alternative sources, though this requirement is greater in more highly proliferating cancer cells with lower pre-existing cellular reserves. To better understand if it is the case of alternative resources the next stage of the study analysed glutamine utilisation in the TCA cycle. A key indication of such uses is the decrease in the presence of certain labelled markers ratios presence in the pancreatic cancer cells despite it being shown that there is an overall increase in the feedback into the TCA cycle from the point of glucose. This finding strongly suggests that the labelled metabolites are being effectively dampened down by the influx of unlabelled  $\alpha$ -ketoglutarate with the most credible source being glutamine via glutaminolysis.

A final less prominent utilisation of glucose during cancer cell development is through its use in amino acid biosynthesis. Amino acids traced during this analysis included methionine, serine, aspartic acid, tyrosine and glycine. All of these amino acids showed a higher proportion of glucose incorporation over the experimental period in A549, lung cancer cells and MIA-PaCa-2 compared to healthy counterparts whereas a lower proportion was noted in Panc-1. This is likely due to Panc-1's higher utilisation of the glucose source for TCA based energy metabolism.

Serine and glycine are generated from 3-phosphoglycerate using 3-phosphoglycerate dehydrogenase to make the former before that is then converted into serine before serine hydroxymethyltransferase then generates glycine. Glycine has a key role in highly proliferating cells helping to support protein, lipid, nucleic acid, and glutathione synthesis (Guo, K., et al., 2020).

Oxaloacetate can be utilised in amino acid synthesis through its ability to form aspartic acid and methionine. Transamination is responsible for the production of aspartic acid which, as with alanine, is accompanied by the formation of  $\alpha$ -ketoglutarate from glutamine catalyzed by GPT. This process has been proven to increase proliferation of cancer cells (Garcia-Bermudez, J., et al., 2018). Prior data showed that Mia-PaCa-2 was the only cancerous cell line to have an increased net cellular presence of aspartic acid in comparison to healthy equivalent cell lines, as the others were lower. The current tracing analysis found an up-regulation of glucose formed aspartic acid in A549 and Mia PaCa-2 cell lines but lower in Panc-1. This shows that there is a very heavy utilisation of glucose in aspartic acid formation particularly in the MiaPaCa-2 cell lines to displace a larger proportion of a bigger pool of amino acids with glucose generated presence especially in comparison to Panc-1 cells which showed the opposite trend with a smaller pool showing a decreased glucose labelled presence over the experimental period indicates that this is a much less prioritised pathway for glucose utilisation.

This study shows that in addition to the observed diagnostic markers changes the pathways involved in the amino acid metabolite profile differences seen are very unique for different cancers and even different forms of the same cancers.

When analysing glutamine as a carbon source for TCA cycle input the first indications of pathway utilisation possible in this study looks at the proportion of glutamic acid that carried the marker compared to those that do not. The previous section showed glutamine levels were a lot higher in lung cell lines than in pancreatic cells, with an elevated presence in pancreatic cancer cell lines (MiaPaCa-2 cell lines in particular) than non cancerous, whereas lung cancer cell line A549 was slightly lower than the healthy equivalent line Beas-2b.

When considering the sources of this glutamine in this study the amount generated over a fixed time frame from a glucose source was found to be quite limited- showing a less than 30% presence in all the different cell types under evaluation (only considering the initial m+2 isotopologue). This proportion was found to be higher in the cancerous cell lines than their healthy equivalent. When now considering how much of the remaining can be attributed to influx from labelled glutamine resource through measuring the M+5 isotopologue of glutamic acid this is much more prevalent source overall, as would be expected, in all cell lines other than Beas-2b, Figure 2-54. With Beas-2b having the greatest overall pool of glutamine to begin with, this lower percentage is likely a reflection upon how much unlabelled still remains in the cell. This suggests that overall cancer cells are more reliant on glutamine influx and utilisation than healthy cell lines and a way that they enable this is through a great direct influx of glutamine rather than sources through the TCA cycle.

When considering specific cancerous cell lines under evaluation, Mia-PaCa-2 cell lines had such a vast initial pool of glutamine, with levels significantly higher than all other pancreatic cell line the proportion of labelled glutamine that has been taken in during the timeframe of the study is extremely high and illustrates a significantly higher presence of GLUT1 transporters presence in this particular cell type. All TCA intermediates were also found to be up-regulated for the glutamine derived marker in Mia-PaCa-2 cells, adding to the proof of this concept.

Panc-1 cells had comparable net levels of glutamine to healthy cell lines but have shown an increased proportion of glucose derived glutamine accompanied with a decreased glutamine uptake. This further evidences Panc-1 cells extended utilisation

of glycolysis and TCA cycle more for their energy upkeep than healthy and Mia-PaCa cell lines, more so than other pathways and stores available.

A549 had a slightly lower pool of glutamine than healthy counterpart Beas-2b but the proportion of labelled glutamine to unlabelled is higher for both glucose and glutamine derived sources. This shows an increase in glycolysis pathways as well as an overall increased influx required likely to keep up with an accelerated metabolic demand. This case is strengthened through looking at the significantly higher proportion of labelled TCA cycle intermediates derived from glutamine in A549 when compared to A549 (citrate and  $\alpha$ -ketoglutarate)

### **Glucose starvation**

Solid tumours particularly those such as pancreatic cancer due to mass density can have a tendency to experience poor blood and nutritional supply, often experiencing significantly reduced glucose resources, accelerated more by the increase in demands of these highly proliferating cell types and so need to adapt in order to survive these harsher conditions. Glutamine can act as an additional resource alongside the limited glucose to help overcome these circumstances. Through monitoring the alterations of glutamine utilisation during glucose limited conditions a better understanding of these modifications can be achieved. In general cells will struggle to survive and thrive when glucose depletion occurs and so metabolic processes will slow down or stop to reflect this. Cancer cells may have developed a mechanism to better allow survival under such conditions. This study is looking to see if any are demonstrated in the cell lines under study and what these changes might be. The data reflected this theory and showed that when glucose is removed from the system most cancer cell lines showed a decrease in the proportion of labelled glutamine present illustrating a reduced influx over the experimental period despite the removal of another glutamine source. This is most likely due to a decrease in the rate of proliferation and therefore the amount metabolite required to meet energy/anabolic demands. Surprisingly two cell lines, Beas-2b and Panc-1, showed an increase in the presence of the labelled marker when compared to unlabelled possibly showing an adaptation to maintain the demands or reflecting the now



lowered cellular reserve meaning a higher proportion of labelled marker presence. To better understand why these trends are seen a more detailed evaluation of the other metabolites in the TCA cycle is required as well as the further uses of metabolites relating to fatty acid synthesis and pyruvate salvage. First analysing the TCA cycle, when glucose is absent and glutamine is traced the level of feed in remains comparable to when glucose was present in lung cancer cell lines, with pancreatic more of a change is noticeable with Mia-PaCa-2 showing a decreased labelled presence and HDPE and Panc-1 showing an elevated presence, relating to citrate M+4 isotopologue. Expanding to the tracing of other intermediates there is a more clear reduction in metabolite markers in the lung cell lines, most especially Beas-B and a decrease in HDPE and Mia-PaCa-2 but an increase in Panc-1. This finding in Panc-1 cell lines suggests under normal conditions the cell line is more reliant on glucose derived sources for energy cycling whereas under starved conditions it is better adapted to utilising alternative energy sources for its survival than the other cell lines under study. The other cell lines either do not thrive as capably or are utilising other reserved other than glutamine more in order to counteract the lack of glucose.

Pyruvate salvaging was seen to be a significant utilisation of glutamine under glucose depleted conditions with a large proportion of the pyruvate observed being a product of this source in pancreatic cell line HDPE and Mia PaCa-2 only, the others remain extremely low with extremely low detected labelled pyruvate present in the pool. Citrate M+5 production gives another interesting oversight into glutamine utilisation. Citrate isotopologue M+5 is generated through the reductive metabolism of fully labelled  $\alpha$ -ketoglutarate being the source which then becomes heavily utilised in fatty acid synthesis. This source was found to be significantly up-regulated in lung cancer and MiaPaca-2 cell lines with glucose presence though downregulated in Panc-1 when compared to healthy cell lines. Under glucose starvation the healthy lung cells show no detectable traceable markers for this pathway present, whereas the cancerous cells showed a lowered but still active participation. Alternatively pancreatic cells showed an accelerated utilisation of glutamine in this way in healthy cell lines but a lowered proportion in cancerous.

Ultimately this report has shown that alterations in glucose and glutamine metabolism relating to TCA and glycolysis involvement show key differentiating capabilities for identifying cancerous and non-cancerous cell lines from each other but the specifics of these alterations are very unique to different cancer types and cell lines. This study has shown that the overall diagnostic profiles are very distinct even for different forms of the same cancer and whilst SITA has enabled a better mechanistic understanding about why trends like glucose upregulation are seen in some cancers but not others it can offer a more tailored understanding into what the phenotypic profile of the cancers is if these can be better correlated to cancer traits.

The final aim of this chapter sought to compare the different biomarker discovery techniques to assess which would give the best diagnostic markers and show the greatest differences in metabolic patterns whilst also considering which would be the most simplistic and available to deploy.

This study has aided the development of more inclusive markers covering a broad range of different metabolic profiles in pancreatic cancers. This will give a better insight into changes that occur in pancreatic cancer cells that lead to malignancies offering sources to therapeutic targets and of focus provide strong diagnostic biomarkers for earlier and more effective methods of diagnosis than currently are utilised. Overall GCMS and MALDI provided the most effective biomarkers for diagnosis of pancreatic cancers, generating multiple markers with good diagnostic potential given the sample sets under analysis. The caveat of the small sample set does draw into question the robustness of the markers and these will need to be more stringently analysed with increased sample sets. Furthermore, only two different cell-lines were compared during this analysis and whilst this can show that there is big diversity in cancer cell lines that needs to be considered introducing a more varied set of cultures would help to paint a bigger picture of the changes that are occurring and improve the statistical comparisons being drawn for the most important markers and pathway changes that are occurring.

However, these methods require a more invasive sample source, more expensive equipment that requires more training and time to analyse. Though this didn't alleviate the nature of the sample source, it generated a strong number of biomarker candidates capable of diagnosis. The least invasive strategy, looking at nucleoside markers, provided the weakest diagnostic markers of all of the technique employed though it still captured a good enough number of positive cases in a non-invasive manner showing its potential as a screening tool more so than one of diagnostic.

#### Future developments

Whilst this study found multiple potential biomarkers for both diagnosis and screening, there are a number of limitations in this study that can be improved upon.

Firstly, a follow up confirmational study to definitively identify and quantify the markers found to be of interest would enable a more global picture about which pathways are being altered and enable comparisons to be drawn to other similar studies and cancers. Additionally, a big limitation in this study can be found in the limited sample sizes which will impact the robustness of the statistical analysis. Through expanding the number of pancreatic cancer cell line types and therefore sample groups it would allow not only a greater understanding of the variability of pancreatic cancer cells but improve the statistical robustness.

Beyond this, moving into real life samples will enable a better gauge over conserved metabolite markers found in this study and how they are reflected under real conditions. Under artificial environments, where metabolite sources are plentiful in media certain trends may be conveyed that are not reflective of real scenarios. For instance, the pancreatic and lung cells overlapped with each other for metabolic presence during the GCMS analysis proving to be more similar in cancerous status rather than cell source. Though this could be the result of malignancy it could equally be the result of a more abundant environment.

## 2.6- References

Angerer, T. B., Bour, J., Biagi, J. L., Moskovets, E., & Frache, G. (2022). Evaluation of 6 MALDI-Matrices for 10  $\mu$ m Lipid Imaging and On-Tissue MSn with AP-MALDI-Orbitrap. *Journal of the American Society for Mass Spectrometry*, 33(5), 760–771. <https://doi.org/10.1021/jasms.1c00327>

Ballehaninna UK, and Chamberlain RS. (2012) The clinical utility of serum CA 19-9 in the diagnosis, prognosis and management of pancreatic adenocarcinoma: An evidence based appraisal. *J Gastrointest Oncol*.;3(2):105-19.

Bott, A. J., Peng, I. C., Fan, Y., Faubert, B., Zhao, L., Li, J., Neidler, S., Sun, Y., Jaber, N., Krokowski, D., Lu, W., Pan, J. A., Powers, S., Rabinowitz, J., Hatzoglou, M., Murphy, D. J., Jones, R., Wu, S., Girnun, G., & Zong, W. X. (2015). Oncogenic Myc Induces Expression of Glutamine Synthetase through Promoter Demethylation. *Cell metabolism*, 22(6), 1068–1077. <https://doi.org/10.1016/j.cmet.2015.09.025>

Bott, A. J., Shen, J., Tonelli, C., Zhan, L., Sivaram, N., Jiang, Y. P., Yu, X., Bhatt, V., Chiles, E., Zhong, H., Maimouni, S., Dai, W., Velasquez, S., Pan, J. A., Muthalagu, N., Morton, J., Anthony, T. G., Feng, H., Lamers, W. H., Murphy, D. J., ... Zong, W. X. (2019). Glutamine Anabolism Plays a Critical Role in Pancreatic Cancer by Coupling Carbon and Nitrogen Metabolism. *Cell reports*, 29(5), 1287–1298.e6. <https://doi.org/10.1016/j.celrep.2019.09.056>

Berry, K. A., Hankin, J. A., Barkley, R. M., Spraggins, J. M., Caprioli, R. M., & Murphy, R. C. (2011). MALDI imaging of lipid biochemistry in tissues by mass spectrometry. *Chemical reviews*, 111(10), 6491–6512. <https://doi.org/10.1021/cr200280p>

Cai, M., Wan, J., Cai, K., Song, H., Wang, Y., Sun, W., & Hu, J. (2022). Understanding the Contribution of Lactate Metabolism in Cancer Progress: A Perspective from Isomers. *Cancers*, 15(1), 87. <https://doi.org/10.3390/cancers15010087>

Chagovets, V. V., Starodubtseva, N. L., Tokareva, A. O., Frankevich, V. E., Rodionov, V. V., Kometova, V. V., Chingin, K., Kukaev, E. N., Chen, H., & Sukhikh, G. T. (2020). Validation of Breast Cancer Margins by Tissue Spray Mass Spectrometry.

International journal of molecular sciences, 21(12), 4568.  
<https://doi.org/10.3390/ijms21124568>

Chen YJ, Mahieu NG, Huang X, Singh M, Crawford PA, Johnson SL, et al. Lactate metabolism is associated with mammalian mitochondria. *Nat Chem Biol.* (2016) 12:937–43. doi: 10.1038/nchembio.2172

Choi, B. H., & Coloff, J. L. (2019). The Diverse Functions of Non-Essential Amino Acids in Cancer. *Cancers*, 11(5), 675. <https://doi.org/10.3390/cancers11050675>

Cluntun, A. A., Lukey, M. J., Cerione, R. A., & Locasale, J. W. (2017). Glutamine Metabolism in Cancer: Understanding the Heterogeneity. *Trends in cancer*, 3(3), 169–180. <https://doi.org/10.1016/j.trecan.2017.01.005>

Davis-Yadley, A. H., Abbott, A. M., Pimiento, J. M., Chen, D. T., & Malafa, M. P. (2016). Increased Expression of the Glucose Transporter Type 1 Gene Is Associated With Worse Overall Survival in Resected Pancreatic Adenocarcinoma. *Pancreas*, 45(7), 974–979. <https://doi.org/10.1097/MPA.0000000000000580>

de la Cruz-López, K. G., Castro-Muñoz, L. J., Reyes-Hernández, D. O., García-Carrancá, A., & Manzo-Merino, J. (2019). Lactate in the Regulation of Tumor Microenvironment and Therapeutic Approaches. *Frontiers in oncology*, 9, 1143. <https://doi.org/10.3389/fonc.2019.01143>

Desai, T. J., Toombs, J. E., Minna, J. D., Brekken, R. A., & Udugamasooriya, D. G. (2016). Identification of lipid-phosphatidylserine (PS) as the target of unbiasedly selected cancer specific peptide-peptoid hybrid PPS1. *Oncotarget*, 7(21), 30678–30690. <https://doi.org/10.18632/oncotarget.8929>

Desbenoit, N., Walch, A., Spengler, B., Brunelle, A., & Römpf, A. (2018). Correlative mass spectrometry imaging, applying time-of-flight secondary ion mass spectrometry and atmospheric pressure matrix-assisted laser desorption/ionization to a single tissue section. *Rapid communications in mass spectrometry : RCM*, 32(2), 159–166. <https://doi.org/10.1002/rcm.8022>

- Dória, M. L., McKenzie, J. S., Mroz, A., Phelps, D. L., Speller, A., Rosini, F., Strittmatter, N., Golf, O., Veselkov, K., Brown, R., Ghaem-Maghami, S., & Takats, Z. (2016). Epithelial ovarian carcinoma diagnosis by desorption electrospray ionization mass spectrometry imaging. *Scientific reports*, 6, 39219. <https://doi.org/10.1038/srep39219>
- Dudley, E., Lemiere, F., Van Dongen, W., Langridge, J. I., El-Sharkawi, S., Games, D. E., Esmans, E. L., & Newton, R. P. (2003). Analysis of urinary nucleosides. III. Identification of 5'-deoxycytidine in urine of a patient with head and neck cancer. *Rapid communications in mass spectrometry : RCM*, 17(11), 1132–1136. <https://doi.org/10.1002/rcm.1034>
- Fazli, H. R., Moradzadeh, M., Mehrbakhsh, Z., Sharafkhah, M., Masoudi, S., Pourshams, A., & Mohamadkhani, A. (2021). Diagnostic Significance of Serum Fatty Acid Synthase in Patients with Pancreatic Cancer. *Middle East journal of digestive diseases*, 13(2), 115–120. <https://doi.org/10.34172/mejdd.2021.214>
- Fernández, R., Lage, S., Abad-García, B., Barceló-Coblijn, G., Terés, S., López, D. H., Guardiola-Serrano, F., Martín, M. L., Escribá, P. V., & Fernández, J. A. (2014). Analysis of the lipidome of xenografts using MALDI-IMS and UHPLC-ESI-QTOF. *Journal of the American Society for Mass Spectrometry*, 25(7), 1237–1246. <https://doi.org/10.1007/s13361-014-0882-3>
- Fischer, K., Hoffmann, P., Voelkl, S., Meidenbauer, N., Ammer, J., Edinger, M., Gottfried, E., Schwarz, S., Rothe, G., Hoves, S., Renner, K., Timischl, B., Mackensen, A., Kunz-Schughart, L., Andreesen, R., Krause, S. W., & Kreutz, M. (2007). Inhibitory effect of tumor cell-derived lactic acid on human T cells. *Blood*, 109(9), 3812–3819. <https://doi.org/10.1182/blood-2006-07-035972>
- Fu, Y., Zou, T., Shen, X., Nelson, P. J., Li, J., Wu, C., Yang, J., Zheng, Y., Bruns, C., Zhao, Y., Qin, L., & Dong, Q. (2020). Lipid metabolism in cancer progression and therapeutic strategies. *MedComm*, 2(1), 27–59. <https://doi.org/10.1002/mco2.27>
- Gradiz, R., Silva, H. C., Carvalho, L., Botelho, M. F., & Mota-Pinto, A. (2016). MIA PaCa-2 and PANC-1 - pancreas ductal adenocarcinoma cell lines with neuroendocrine

differentiation and somatostatin receptors. *Scientific reports*, 6, 21648. <https://doi.org/10.1038/srep21648>

Ghimire, H., Garlapati, C., Janssen, E. A. M., Krishnamurti, U., Qin, G., Aneja, R., & Perera, A. G. U. (2020). Protein Conformational Changes in Breast Cancer Sera Using Infrared Spectroscopic Analysis. *Cancers*, 12(7), 1708. <https://doi.org/10.3390/cancers12071708>

Garcia-Bermudez, J., Baudrier, L., La, K., Zhu, X. G., Fidelin, J., Sviderskiy, V. O., Papagiannakopoulos, T., Molina, H., Snuderl, M., Lewis, C. A., Possemato, R. L., & Birsoy, K. (2018). Aspartate is a limiting metabolite for cancer cell proliferation under hypoxia and in tumours. *Nature cell biology*, 20(7), 775–781. <https://doi.org/10.1038/s41556-018-0118-z>

Ghimire, H., Garlapati, C., Janssen, E. A. M., Krishnamurti, U., Qin, G., Aneja, R., & Perera, A. G. U. (2020). Protein Conformational Changes in Breast Cancer Sera Using Infrared Spectroscopic Analysis. *Cancers*, 12(7), 1708. <https://doi.org/10.3390/cancers12071708>

Guo, K., Cao, Y., Li, Z., Zhou, X., Ding, R., Chen, K., Liu, Y., Qiu, Y., Wu, Z., & Fang, M. (2020). Glycine metabolomic changes induced by anticancer agents in A549 cells. *Amino acids*, 52(5), 793–809. <https://doi.org/10.1007/s00726-020-02853-0>

Guo, S., Qiu, L., Wang, Y., Qin, X., Liu, H., He, M., Zhang, Y., Li, Z., & Chen, X. (2014). Tissue imaging and serum lipidomic profiling for screening potential biomarkers of thyroid tumors by matrix-assisted laser desorption/ionization-Fourier transform ion cyclotron resonance mass spectrometry. *Analytical and bioanalytical chemistry*, 406(18), 4357–4370. <https://doi.org/10.1007/s00216-014-7846-0>

Hayashi, K., Jutabha, P., Endou, H., & Anzai, N. (2012). c-Myc is crucial for the expression of LAT1 in MIA Paca-2 human pancreatic cancer cells. *Oncology reports*, 28(3), 862–866. <https://doi.org/10.3892/or.2012.1878>

Hillyer, P., Shepard, R., Uehling, M., Krenz, M., Sheikh, F., Thayer, K. R., Huang, L., Yan, L., Panda, D., Luongo, C., Buchholz, U. J., Collins, P. L., Donnelly, R. P., & Rabin, R. L. (2018). Differential Responses by Human Respiratory Epithelial Cell Lines to

Respiratory Syncytial Virus Reflect Distinct Patterns of Infection Control. *Journal of virology*, 92(15), e02202-17. <https://doi.org/10.1128/JVI.02202-17>

Hong, J. H., Kang, J. W., Kim, D. K., Baik, S. H., Kim, K. H., Shanta, S. R., Jung, J. H., Mook-Jung, I., & Kim, K. P. (2016). Global changes of phospholipids identified by MALDI imaging mass spectrometry in a mouse model of Alzheimer's disease. *Journal of lipid research*, 57(1), 36–45. <https://doi.org/10.1194/jlr.M057869>

Jackson, S. N., Barbacci, D., Egan, T., Lewis, E. K., Schultz, J. A., & Woods, A. S. (2014). MALDI-Ion Mobility Mass Spectrometry of Lipids in Negative Ion Mode. *Analytical methods : advancing methods and applications*, 6(14), 5001–5007. <https://doi.org/10.1039/C4AY00320A>

Jiang, L., Chughtai, K., Purvine, S. O., Bhujwalla, Z. M., Raman, V., Paša-Tolić, L., Heeren, R. M. A., & Glunde, K. (2015). MALDI-Mass Spectrometric Imaging Revealing Hypoxia-Driven Lipids and Proteins in a Breast Tumor Model. *Analytical chemistry*, 87(12), 5947–5956. <https://doi.org/10.1021/ac504503x>

Jones, E. E., Dworski, S., Canals, D., Casas, J., Fabrias, G., Schoenling, D., Levade, T., Denlinger, C., Hannun, Y. A., Medin, J. A., & Drake, R. R. (2014). On-tissue localization of ceramides and other sphingolipids by MALDI mass spectrometry imaging. *Analytical chemistry*, 86(16), 8303–8311. <https://doi.org/10.1021/ac501937d>

Jora, M., Lobue, P. A., Ross, R. L., Williams, B., & Addepalli, B. (2019). Detection of ribonucleoside modifications by liquid chromatography coupled with mass spectrometry. *Biochimica et biophysica acta. Gene regulatory mechanisms*, 1862(3), 280–290. <https://doi.org/10.1016/j.bbagr.2018.10.012>

Kaira, K., Oriuchi, N., Imai, H., Shimizu, K., Yanagitani, N., Sunaga, N., Hisada, T., Kawashima, O., Iijima, H., Ishizuka, T., Kanai, Y., Endou, H., Nakajima, T., & Mori, M. (2008). Expression of L-type amino acid transporter 1 (LAT1) in neuroendocrine tumors of the lung. *Pathology, research and practice*, 204(8), 553–561. <https://doi.org/10.1016/j.prp.2008.02.003>

Kaira, K., Sunose, Y., Arakawa, K., Ogawa, T., Sunaga, N., Shimizu, K., Tominaga, H., Oriuchi, N., Itoh, H., Nagamori, S., Kanai, Y., Segawa, A., Furuya, M., Mori, M., Oyama,



T., & Takeyoshi, I. (2012). Prognostic significance of L-type amino-acid transporter 1 expression in surgically resected pancreatic cancer. *British journal of cancer*, 107(4), 632–638. <https://doi.org/10.1038/bjc.2012.310>

Kang J. S. (2020). Dietary restriction of amino acids for Cancer therapy. *Nutrition & metabolism*, 17, 20. <https://doi.org/10.1186/s12986-020-00439-x>

Kamphorst, J. J., Nofal, M., Commisso, C., Hackett, S. R., Lu, W., Grabocka, E., Vander Heiden, M. G., Miller, G., Drebin, J. A., Bar-Sagi, D., Thompson, C. B., & Rabinowitz, J. D. (2015). Human pancreatic cancer tumors are nutrient poor and tumor cells actively scavenge extracellular protein. *Cancer research*, 75(3), 544–553. <https://doi.org/10.1158/0008-5472.CAN-14-2211>

Kaya, I., Jennische, E., Lange, S., & Malmberg, P. (2018). Dual polarity MALDI imaging mass spectrometry on the same pixel points reveals spatial lipid localizations at high-spatial resolutions in rat small intestine. *Analytical methods : advancing methods and applications*, 10(21), 2428–2435. <https://doi.org/10.1039/c8ay00645h>

Kim JE, Lee KT, Lee JK, Paik SW, Rhee JC, and Choi KW. (2004) Clinical usefulness of carbohydrate antigen 19-9 as a screening test for pancreatic cancer in an asymptomatic population. *J Gastroenterol Hepatol*. 19(2):182-6.

Koundouros, N., & Poulogiannis, G. (2020). Reprogramming of fatty acid metabolism in cancer. *British journal of cancer*, 122(1), 4–22. <https://doi.org/10.1038/s41416-019-0650-z>

Lewis, P. D., Lewis, K. E., Ghosal, R., Bayliss, S., Lloyd, A. J., Wills, J., Godfrey, R., Kloer, P., & Mur, L. A. (2010). Evaluation of FTIR spectroscopy as a diagnostic tool for lung cancer using sputum. *BMC cancer*, 10, 640. <https://doi.org/10.1186/1471-2407-10-640>

Li, L., Bi, X., Sun, H., Liu, S., Yu, M., Zhang, Y., Weng, S., Yang, L., Bao, Y., Wu, J., Xu, Y., & Shen, K. (2018). Characterization of ovarian cancer cells and tissues by Fourier transform infrared spectroscopy. *Journal of ovarian research*, 11(1), 64. <https://doi.org/10.1186/s13048-018-0434-8>

- Lima, A. R., Carvalho, M., Aveiro, S. S., Melo, T., Domingues, M. R., Macedo-Silva, C., Coimbra, N., Jerónimo, C., Henrique, R., Bastos, M. L., Guedes de Pinho, P., & Pinto, J. (2022). Comprehensive Metabolomics and Lipidomics Profiling of Prostate Cancer Tissue Reveals Metabolic Dysregulations Associated with Disease Development. *Journal of proteome research*, 21(3), 727–739. <https://doi.org/10.1021/acs.jproteome.1c00754>
- Liberti, M. V., & Locasale, J. W. (2016). The Warburg Effect: How Does it Benefit Cancer Cells?. *Trends in biochemical sciences*, 41(3), 211–218. <https://doi.org/10.1016/j.tibs.2015.12.001>
- Ma, X., Li, B., Liu, J., Fu, Y., & Luo, Y. (2019). Phosphoglycerate dehydrogenase promotes pancreatic cancer development by interacting with eIF4A1 and eIF4E. *Journal of experimental & clinical cancer research : CR*, 38(1), 66. <https://doi.org/10.1186/s13046-019-1053-y>
- Marien, E., Meister, M., Muley, T., Fieuws, S., Bordel, S., Derua, R., Spraggins, J., Van de Plas, R., Dehairs, J., Wouters, J., Bagadi, M., Dienemann, H., Thomas, M., Schnabel, P. A., Caprioli, R. M., Waelkens, E., & Swinnen, J. V. (2015). Non-small cell lung cancer is characterized by dramatic changes in phospholipid profiles. *International journal of cancer*, 137(7), 1539–1548. <https://doi.org/10.1002/ijc.29517>
- Mehrotra, R., Tyagi, G., Jangir, D. K., Dawar, R., & Gupta, N. (2010). Analysis of ovarian tumor pathology by Fourier Transform Infrared Spectroscopy. *Journal of ovarian research*, 3, 27. <https://doi.org/10.1186/1757-2215-3-27>
- Otto A. M. (2020). Metabolic Constants and Plasticity of Cancer Cells in a Limiting Glucose and Glutamine Microenvironment-A Pyruvate Perspective. *Frontiers in oncology*, 10, 596197. <https://doi.org/10.3389/fonc.2020.596197>
- Parker, S. J., Amendola, C. R., Hollinshead, K. E. R., Yu, Q., Yamamoto, K., Encarnación-Rosado, J., Rose, R. E., LaRue, M. M., Sohn, A. S. W., Biancur, D. E., Paulo, J. A., Gygi, S. P., Jones, D. R., Wang, H., Philips, M. R., Bar-Sagi, D., Mancias, J. D., & Kimmelman, A. C. (2020). Selective Alanine Transporter Utilization Creates a Targetable Metabolic

Niche in Pancreatic Cancer. *Cancer discovery*, 10(7), 1018–1037.  
<https://doi.org/10.1158/2159-8290.CD-19-0959>

Pérez-Tomás, R., & Pérez-Guillén, I. (2020). Lactate in the Tumor Microenvironment: An Essential Molecule in Cancer Progression and Treatment. *Cancers*, 12(11), 3244.  
<https://doi.org/10.3390/cancers12113244>

Piubelli, L., Murtas, G., Rabattoni, V., & Pollegioni, L. (2021). The Role of D-Amino Acids in Alzheimer's Disease. *Journal of Alzheimer's disease : JAD*, 80(2), 475–492.  
<https://doi.org/10.3233/JAD-201217>

Pirman, D. A., Efuet, E., Ding, X. P., Pan, Y., Tan, L., Fischer, S. M., DuBois, R. N., & Yang, P. (2013). Changes in cancer cell metabolism revealed by direct sample analysis with MALDI mass spectrometry. *PloS one*, 8(4), e61379.  
<https://doi.org/10.1371/journal.pone.0061379>

Rasmuson, T., & Björk, G. R. (1995). Urinary excretion of pseudouridine and prognosis of patients with malignant lymphoma. *Acta oncologica (Stockholm, Sweden)*, 34(1), 61–67. <https://doi.org/10.3109/02841869509093640>

Seidel, A., Brunner, S., Seidel, P., Fritz, G. I., & Herbarth, O. (2006). Modified nucleosides: an accurate tumour marker for clinical diagnosis of cancer, early detection and therapy control. *British journal of cancer*, 94(11), 1726–1733.  
<https://doi.org/10.1038/sj.bjc.6603164>

Shen, Y., Pu, K., Zheng, K., Ma, X., Qin, J., Jiang, L., & Li, J. (2019). Differentially Expressed microRNAs in MIA PaCa-2 and PANC-1 Pancreas Ductal Adenocarcinoma Cell Lines are Involved in Cancer Stem Cell Regulation. *International journal of molecular sciences*, 20(18), 4473. <https://doi.org/10.3390/ijms20184473>

Shi J., Zong. L., Wang. P., Li. C., Yue. J., and Wang. X., (2022) A study of biomacromolecular changes in dandelion root extract induced PANC-1 cells apoptosis using FTIR microspectroscopy. *Infrared physics and technology*. 120(2022)103995

Sommella, E., Salviati, E., Caponigro, V., Grimaldi, M., Musella, S., Bertamino, A., Cacace, L., Palladino, R., Mauro, G. D., Marini, F., D'Ursi, A. M., & Campiglia, P. (2022).

MALDI Mass Spectrometry Imaging Highlights Specific Metabolome and Lipidome Profiles in Salivary Gland Tumor Tissues. *Metabolites*, 12(6), 530. <https://doi.org/10.3390/metabo12060530>

Sousa, C. M., Biancur, D. E., Wang, X., Halbrook, C. J., Sherman, M. H., Zhang, L., Kremer, D., Hwang, R. F., Witkiewicz, A. K., Ying, H., Asara, J. M., Evans, R. M., Cantley, L. C., Lyssiotis, C. A., & Kimmelman, A. C. (2016). Pancreatic stellate cells support tumour metabolism through autophagic alanine secretion. *Nature*, 536(7617), 479–483. <https://doi.org/10.1038/nature19084>

Sun, T. (2012). Mass Spectrometry Applied to Problems in Lipid Biochemistry: Microchip Based Approach for Lipidomics Profiling and Analysis of Lipid Metabolites by LC-MS/MS (Doctoral dissertation, Duquesne University). Retrieved from <https://dsc.duq.edu/etd/1254>

Swinnen, J. V., Van Veldhoven, P. P., Timmermans, L., De Schrijver, E., Brusselmans, K., Vanderhoydonc, F., Van de Sande, T., Heemers, H., Heyns, W., & Verhoeven, G. (2003). Fatty acid synthase drives the synthesis of phospholipids partitioning into detergent-resistant membrane microdomains. *Biochemical and biophysical research communications*, 302(4), 898–903. [https://doi.org/10.1016/s0006-291x\(03\)00265-1](https://doi.org/10.1016/s0006-291x(03)00265-1)

Tajiri, K., & Shimizu, Y. (2013). Branched-chain amino acids in liver diseases. *World journal of gastroenterology*, 19(43), 7620–7629. <https://doi.org/10.3748/wjg.v19.i43.7620>

Tsai, P. Y., Lee, M. S., Jadhav, U., Naqvi, I., Madha, S., Adler, A., Mistry, M., Naumenko, S., Lewis, C. A., Hitchcock, D. S., Roberts, F. R., DelNero, P., Hank, T., Honselmann, K. C., Morales Oyarvide, V., Mino-Kenudson, M., Clish, C. B., Shivdasani, R. A., & Kalaany, N. Y. (2021). Adaptation of pancreatic cancer cells to nutrient deprivation is reversible and requires glutamine synthetase stabilization by mTORC1. *Proceedings of the National Academy of Sciences of the United States of America*, 118(10), e2003014118. <https://doi.org/10.1073/pnas.2003014118>

van Heerde MJ, Buijs J, Hansen BE, de Waart M, van Eijck CH, Kazemier G, Pek CJ, Poley JW, Bruno MJ, Kuipers EJ, van Buuren HR. (2014) Serum level of Ca 19-9

increases ability of IgG4 test to distinguish patients with autoimmune pancreatitis from those with pancreatic carcinoma.59(6):1322-9

Vander Heiden, M. G., Cantley, L. C., & Thompson, C. B. (2009). Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science (New York, N.Y.)*, 324(5930), 1029–1033. <https://doi.org/10.1126/science.1160809>

Vanhove, K., Graulus, G. J., Mesotten, L., Thomeer, M., Derveaux, E., Noben, J. P., Guedens, W., & Adriaenssens, P. (2019). The Metabolic Landscape of Lung Cancer: New Insights in a Disturbed Glucose Metabolism. *Frontiers in oncology*, 9, 1215. <https://doi.org/10.3389/fonc.2019.01215>

Vettore, L., Westbrook, R. L., & Tennant, D. A. (2020). New aspects of amino acid metabolism in cancer. *British journal of cancer*, 122(2), 150–156. <https://doi.org/10.1038/s41416-019-0620-5>

Wang, S., Chen, X., Luan, H., Gao, D., Lin, S., Cai, Z., Liu, J., Liu, H., & Jiang, Y. (2016). Matrix-assisted laser desorption/ionization mass spectrometry imaging of cell cultures for the lipidomic analysis of potential lipid markers in human breast cancer invasion. *Rapid communications in mass spectrometry : RCM*, 30(4), 533–542. <https://doi.org/10.1002/rcm.7466>

Warburg O, Wind F, Negelein E. The metabolims of tumors in the body. *J Gen Physiol.* (1927) 8:519–30. doi: 10.1085/jgp.8.6.519

Wei, Z., Liu, X., Cheng, C., Yu, W., & Yi, P. (2021). Metabolism of Amino Acids in Cancer. *Frontiers in cell and developmental biology*, 8, 603837. <https://doi.org/10.3389/fcell.2020.603837>

Yang, J., Xu, G., Zheng, Y., Kong, H., Pang, T., Lv, S., & Yang, Q. (2004). Diagnosis of liver cancer using HPLC-based metabonomics avoiding false-positive result from hepatitis and hepatocirrhosis diseases. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences*, 813(1-2), 59–65. <https://doi.org/10.1016/j.jchromb.2004.09.032>

Yang, L., Achreja, A., Yeung, T. L., Mangala, L. S., Jiang, D., Han, C., Baddour, J., Marini, J. C., Ni, J., Nakahara, R., Wahlig, S., Chiba, L., Kim, S. H., Morse, J., Pradeep, S., Nagaraja, A. S., Haemmerle, M., Kyunghye, N., Derichsweiler, M., Plackemeier, T., ... Nagrath, D. (2016). Targeting Stromal Glutamine Synthetase in Tumors Disrupts Tumor Microenvironment-Regulated Cancer Cell Growth. *Cell metabolism*, 24(5), 685–700. <https://doi.org/10.1016/j.cmet.2016.10.011>

Zhang, J., Xu, Y., Li, D., Fu, L., Zhang, X., Bao, Y., & Zheng, L. (2020). Review of the Correlation of LAT1 With Diseases: Mechanism and Treatment. *Frontiers in chemistry*, 8, 564809. <https://doi.org/10.3389/fchem.2020.564809>

Zhang, W. C., Shyh-Chang, N., Yang, H., Rai, A., Umashankar, S., Ma, S., Soh, B. S., Sun, L. L., Tai, B. C., Nga, M. E., Bhakoo, K. K., Jayapal, S. R., Nichane, M., Yu, Q., Ahmed, D. A., Tan, C., Sing, W. P., Tam, J., Thirugananam, A., Noghabi, M. S., ... Lim, B. (2012). Glycine decarboxylase activity drives non-small cell lung cancer tumor-initiating cells and tumorigenesis. *Cell*, 148(1-2), 259–272. <https://doi.org/10.1016/j.cell.2011.11.050>

## Chapter 3 – Human fertility and maternity and metabolic dynamics

### 3.1- Abstract

This study explores serum metabolomic profiles across distinct human subject groups to seek metabolic markers that can best aid further studies including improved understanding of the complex dynamics of fertility, post-partum physiology, and female health conditions such as polycystic ovary syndrome (PCOS) and endometriosis. The study evaluated the serum of 120 candidates (30 males, 30 non-pregnant females, 30 pregnant females and 30 cord serum) using MALDI and GCMS analysis to guide the initial candidate study revealing the best strategy to distinguish each of the different demographical groups from each other. This study identified a majority of markers between pregnant and non-pregnant females to be identified using GCMS analysis to explore amino acid quantification. This strategy was employed to evaluate the serum of healthy (n=6), PCOS (N=5) and endometriosis (n=3) volunteers. Findings reveal significantly lower levels of amino acids in individuals with PCOS and endometriosis compared to healthy cohorts, highlighting potential biomarkers for improved diagnostics and treatment strategies. The next area of the study sought the key variation between cord and maternal serum samples, which was found to be achieved using lipidomic profile changes via MALDI analysis. Adopting this strategy, the analysis of lipidomic profiles serum samples from pregnant women (n=30), cord serum (n=30), and breast milk (n=53) collected at two (n=27) and six (n=26) months post-partum was conducted. The study identifies substantial shifts in lipid composition in breast milk during the post-partum period, underscoring its dynamic nature and identified some potential lipidomic candidates missing from donated breast-milk that are present in cord serum that could be critical for pre-term infants nutritional requirements. Through better understanding each of these elements in the study this chapter will aim to improve diagnostics, guide treatment strategies, and better inform maternal-neonatal healthcare practices.

### 3.2- Introduction

In the pursuit of advancing our understanding of metabolite signatures and their potential ability to distinguish diverse physiological contexts, this chapter embarks on a comprehensive exploration of serum metabolomic profiles among distinct human subject groups, namely males, non-pregnant females, pregnant females, and cord blood samples from pregnant females. The overarching objective is not only to identify differentiating “biomarkers” but also to rigorously evaluate the efficacy of each methodology within its specific instance in order to adapt the best future research use for each.

Furthermore, this chapter delves into the realm of female health differentiation, focusing on the identification of novel biomarkers capable of discerning between serum samples obtained from individuals with Polycystic Ovary Syndrome (PCOS), endometriosis, and healthy female participants. Through an extensive investigation, a deeper understanding of the unique metabolic alterations occurring within each scenario is sought, shedding light on the intricate biochemical underpinnings that define these conditions.

Lastly, the scope of this chapter extends to the complex dynamics of pregnancy and post-partum periods. By scrutinizing metabolite profiles from pregnant women's serum, cord serum, and breast-milk samples collected at varying time points post-partum, the aim is to unravel the intricate alterations in metabolic demands. This holistic examination serves as a means to elucidate the evolving requirements of the foetus and neonate, contributing to a more profound comprehension of developmental processes.

Through better understanding each of these elements in the study this chapter will aim to improve diagnostics, guide treatment strategies, and better inform maternal-neonatal healthcare practices.



### 3.2.1- Serum biomarkers profile in distinct demographics

In recent years, advancements in mass spectrometry- based techniques have revolutionized our understanding of complex biological systems by enabling the comprehensive analysis of metabolites present in various biological fluids, including serum. Serum can provide intricate insights into physiological changes that occur at different life stages or through comparisons of different demographical groups such as males, non-pregnant females, pregnant females and cord serum samples that link pregnant mothers to their infant. Each of these groups represent a different class of physiological states and demands which will then represent diverse metabolic needs and therefore profiles. By utilising mass spectrometry-based techniques, research has begun to better understand these different metabolic intricacies and guide the potential for a better understanding of disease onset, progression and therefore diagnosis strategy.

In this chapter the metabolic profile changes that occur within and between these groups will be explored utilising GCMS, MALDI-TOF and FTIR techniques in order to elucidate these dynamic shifts, their broader implications and offer a tailored and best suited technique for more specific profile comparisons.

### 3.2.2- Female serum fertility exploration

Polycystic ovary syndrome (PCOS) is an endocrinological disorder characterized by an elevated presence of androgens, leading to disruption of ovulation and impacting about 1 in 10 women of reproductive age worldwide (Sadeghi, H. M., et al., 2022). This condition disrupts menstrual regularity, triggers ovarian enlargement, and fosters excessive facial and body hair growth due to an excess of male hormones, culminating in potential fertility complications. The root cause of PCOS remains somewhat elusive, although its hereditary nature and connection to imbalanced hormonal regulation, particularly the high ratio of luteinizing hormone to follicle-stimulating hormone and increased frequency of gonadotropin-releasing hormone, are recognized contributors (Sadeghi, H. M., et al., 2022).

Amino acids play a pivotal role by serving as essential metabolic precursors that maintain biological equilibrium and offer a less essential role as antioxidant capabilities (Unni, S. N., et al., 2015; Xu, N., et al., 2017). This latter feature may be especially important in health as amino acids may exhibit antioxidant potential functions through their abilities in capturing transition metal catalysts. However, during oxidative stress, the presence of reactive oxygen species (ROS) can surpass the available antioxidants, triggering lipid peroxidation and aldehyde formation, ultimately disrupting vital biomolecules. Such disruptions extend to reproductive tissues, impairing their functionality.

Due to the multifaceted nature of PCOS, diagnosis remains challenging, contributing to a high percentage of unidentified cases, reaching up to 75% in the United States (Christ, J. P., & Cedars, M. I. 2023). Presently, a definitive diagnosis cannot be attained through basic methods such as blood tests or biopsies, adding complexity to the diagnostic process. Since 2003, the diagnosis of PCOS has been considered if at least two out of three symptoms; the evidence of oligo-anovulation, hyperandrogenism, and polycystic-appearing ovarian morphology are present, as defined by the Rotterdam criteria (Christ, J. P., & Cedars, M. I. 2023). The prevailing diagnostic strategy in the UK involves evaluating symptomatic presentation, hormonal balance, and follicular presence through ultrasounds. Diagnosis often occurs after excluding other conditions that share similar symptoms, including hyperprolactinemia, thyroid disorders, and Cushing's syndrome (Sadeghi, H. M., et al., 2022; NHSb, 2023).

Apart from fertility challenges, PCOS is associated with several other health issues, such as diabetes due to dysregulated insulin levels, cardiovascular disease, metabolic syndrome, and even depression (Sadeghi, H. M., et al., 2022; Christ, J. P., & Cedars, M. I. 2023). Considering the severe health implications tied to PCOS, coupled with the ongoing delay and uncertainty surrounding its diagnosis, there is a pressing need for innovative diagnostic approaches and a deeper comprehension of its underlying mechanisms to enhance prospects for individuals affected by the syndrome. Certain metabolic markers, such as saturated fatty acids and vitamin D, exhibit connections with PCOS and as such could hold potential for use as diagnostic markers for disease presence. The primary treatment strategy revolves around lifestyle adjustments,

including dietary modifications and certain medication-based interventions (Sadeghi, H. M., et al., 2022).

Endometriosis, an estrogen-dependent reproductive ailment, that affects around 10% of reproductive-aged women and is characterized by the growth of endometrial tissue outside of the uterus. It is a chronic condition causing various symptoms, including dysmenorrhea, dyspareunia, dysuria, chronic abdominal pain, and reproductive issues (Dutta et al., 2018; Parasar et al., 2017; Dutta et al., 2012). Despite its prevalence, the exact cause of endometriosis remains uncertain. Similar to polycystic ovary syndrome (PCOS), endometriosis is linked to inflammation with an increased level of reactive oxygen species (ROS), which play a role in disease development and progression (Clower et al., 2022). ROS influence cell proliferation and enhance the survival and implantation of endometrial cells in abnormal locations. Diagnosing endometriosis is challenging due to the absence of definitive procedures. Currently, diagnosis involves assessing clinical history, evaluating symptoms through differential diagnosis, conducting transvaginal ultrasounds, and performing biopsies on identified masses (Parasar et al., 2017). Not only are these methods of diagnosis extremely invasive and uncomfortable but the lack of a precise diagnostic method often leads to delayed diagnosis, causing needless suffering and a diminished quality of life. Timelier identification could potentially slow disease progression and aid in preserving fertility.

To address these challenges, the development of an affordable, less invasive, and more accurate diagnostic tool could significantly expedite the screening and diagnosis of both endometriosis and PCOS. This study employs exploratory GCMS analysis of serum samples, with a specific focus on amino acid profiles, aiming to identify distinct and unique biomarkers for diagnostic purposes.

### 3.2.3- Lipidomic evaluation of pregnancy dynamics and neonatal requirements

Breast milk is a primary critical food source utilised by mammalian new-borns to establish healthy and optimal development. All current guidance emphasises the importance of exclusive breast feeding being advised for the first 6 months of life.

However, despite this recommendation, less than half of infants under 6 months are exclusively breast fed in this way and are instead moved on to formula alternatives (NHS, 2023; WHO, 2023).

As of 2021, 7.5% of babies were born prematurely in the UK (Office of national statistics, 2023). Pre-term infants tend to suffer health challenges such as incomplete organ development, compromised immune functions and difficulty maintaining adequate nutrition. Being born prior to 37 weeks gestation leaves babies in a vulnerable state, through having poor nutritional stores alongside high nutritional requirements due to a disrupted nutritional supply from the mother and placenta at a critical stage of development. As a result of this, there is a high level of critical growth rate failures noted in premature babies to which breastmilk becomes even more critical as a source of nourishment and protection due to their underdevelopment and vulnerable systems. Breastmilk is considered to be the best form of nutrition for premature babies but there are a number of issues. Firstly, is the availability of donar milk and beyond this, not all breastmilk is made equal. Breastmilk is dynamic, with the composition being specifically tailored to meet an infants needs at a moment in time, providing a blend of nutrients, bioactive compounds and immune factors tailored to support a new-borns growth and protection against infections. The composition of breast milk is known to vary considerably between different species, individuals and even within a single individuals lactation period with changes to neonatal developmental stages. Adjustments can be made, such as fortification with formula, which will account for calorific inadequacies, but the exact nutritional requirements are not considered when doing so.

The composition of breastmilk encompasses an array of carbohydrates, fats and proteins. Lipids hold only 5% total content of breast milk but represent 50% of neonatal energy intake is generated from breast milk fat (Ramiro-Cortijo, D., et al., 2020; Ganeshalingam, M., et al., 2022). Different lipid classes are known to be essential for growth, neurodevelopment, inflammation and infection risk, metabolic and cardiovascular development and disease prevention including intestinal development (Ramiro-Cortijo, D., et al., 2020) making it fundamental to have a better understanding about the requirements at different stages of development. With the

introduction of lipidomic analysis in biomarker discovery there is now a greater push to better understand the presence and specific roles of functional lipids including TAGs, DAGs, FA, plasmalogens and omega based fatty acids (Ganeshalingam, M., et al., 2022). Each component plays a distinct role in supporting the diverse requirements of an infant, contributing to their cognitive development, gut maturation, immune system fortification, and overall well-being. In the case of preterm infants, the challenge lies in comprehending how these components evolve and adapt to best suit the unique nutritional demands and health considerations of these fragile neonates.

This study attempts to analyse breast milk lipidomic profile changes post-partum and draw comparisons to alterations in pregnant women and cord sera lipidomic profiles in order to establish a focus on preterm infants' lipidomic requirements. Such analysis could offer up invaluable insights into potential gaps or imbalances in nutrients to target interventions in order to optimize the nutritional support required by pre-term infants and promote optimal growth and development. This can help to pave the way for personalised nutritional strategies that holds the potential to enhance the overall health and wellbeing of all neonates.

#### 3.2.4- Chapter aims and objectives

This study looks to seek identifying metabolomic differences initially comparing the profiles of one-hundred and twenty serum samples from males, pregnant and non-pregnant females and cord groups with the overall goal to be able to find a technique capable of finding distinct enough differences for later more specific and targeted studies. The profiles of pregnant women and cords will also be scrutinised to assess which metabolic candidates are most required by the foetus for development and whether these are supplied via the cord or other route such as placenta/foetus itself.

All metabolic pathways will be considered in the initial assessment, so use of a non-targeted and global approach is necessary. To achieve the best metabolic coverage the study will utilise several techniques. GCMS will be utilised for focus on amino acid,

lipids and other low molecular weight metabolites. Finally, MALDI-TOF will focus on the analysis of lipidomic analysis of the samples.

#### 3.2.5- The objectives of this chapter are as follows:

- 1) Explore different methodologies that are capable of identifying metabolites or metabolite signatures that can identify changes in serum profile changes in males and non- pregnant females, pregnant females and finally consider the metabolite profile of pregnant females versus cord samples.
- 2) Investigate alterations in metabolite profiles between pregnant women serum, cord serum and breast-milk samples collected at different time points post-partum. Evaluate how fetus/ neo-natal demands change beyond the post-partum period to better understand developmental requirements that may be indicative of nutritional supplementation needs for premature neonates that may be missed by alternative supplementations such as donor milk/ formula.
- 3) Seek novel biomarkers capable of differentiating PCOS, endometriosis and healthy participants serum samples from each other. Provide comprehensive insights into specific metabolic changes occurring in each scenario.

### 3.3. Methodology

#### 3.3.1 Samples

Previously collected serum samples (ethics code: 11/WA/0040) taken from 120 candidates consisting of (n=30) males, (n=30) non-pregnant females at an unknown point of menstrual cycle, (n=30) pregnant females and (n=30) umbilical cord candidates, were frozen at -80°C in small volume (approx. 200ul) aliquots until use. The samples were subject to specific extraction methods for different metabolite compositions before being analysed using GCMS and MALDI technique. These were then analysed using untargeted methodology for significant differences and then critiqued for diagnostic capability. Where possible, the most differing metabolites elucidated will then be identified.

For PCOS and endometriosis analysis, serum previously taken from 14 participants; consisting of healthy (n=6), PCOS (N=5) and endometriosis (n=3) volunteers. The healthy participants were 56% white women, 11% black, 11% mixed and 22% Asian with an average BMI of 26.94, age 25.5 years. Endometreosis were all white women with an average BMI of 24.74 and age of 30 years old. PCOS 61% white, 13% black/mixed and Asian with an average BMI of 27.29 and age of 26.2 years old. All samples were collected with informed consent and ethical approval according to ethics code (SUMS-RESC-2022-023).

Breast milk collected from 53 participants was collected at two (n=27) and six (n=26) months post-partum and before being extracted and analysed using MALDI-TOF analysis.

#### 3.3.2 Metabolite extraction and analysis

##### 3.3.2.1- GCMS

A non-targeted approach was used for the GCMS analysis of the serum samples and aimed to capture the global profiles of the metabolites. Serum was first depleted by taking 100µl of serum, which was then vortexed with 400µl of methanol for thirty minutes before subject to a five minute centrifugation at 300rpm in order to deplete the serum of precipitated protein with the goal of remove the abundant and masking albumin proteins, which would otherwise occlude the spectrum obtained.

Following this, 300µl of the clear depleted layer was then speed vacuum concentrated to dryness before being derivatised as per 2.3.1.2.1 GCMS extraction and derivatisation.

As well as an overall metabolite profile, the data was analysed specifically looking to study changes in amino acid levels and standard curves were utilised for this as outlined in Chapter 2.

All samples in the analysis were run as per 2.3.1.2.2 GCMS analysis. Following analysis, peak areas were extracted from the total ion chromatograms and normalised to the internal standard peak area for inter-sample comparisons and metabolites were tentatively identified by comparison of the mass spectra generated to those of derivatised compounds within the NIST database.

#### *3.3.2.2- MALDI-ToF*

The lipid content of the serum was analysed using MALDI-ToF analysis in both positive and negative reflectron ionisation mode. Ten microlitres of serum was lipid extracted with 100µl of (2:1) chloroform: methanol via 10 second vortex mixing and 3 minutes 3000 rpm centrifugation. After assurance of good phase separation, 50µl of the lower chloroform layer, containing the lipid fraction, was collected and vacuum concentrated to dryness. The dried extract was then reconstituted in matrix. For positive reflectron ionisation mode, the reconstitution was made in 5µl of 10mg/ml of 2,5-Dihydroxybenzoic acid (DHB) in 0.1%TFA (MeOH). Similarly the reconstitution in negative mode was made in 5µl of 10mg/ml para-Nitroalanine (PNA) in 2:1 chloroform: methanol, and a further time with in 5µl of 30 mg/ml of 9-Aminoacridine (9-AA) reconstituted in 3:2 Isopropanol:Acetonitrile was used. In both cases, the dried sample was reconstituted in 5µl of matrix and 1µl of this mixture was then spotted onto a MALDI plate for analysis. Breast-milk samples were extracted by sonication followed by the dilution of the samples 1 in 10000 with water before being mixed 2:1 with the appropriate matrix for positive or negative analysis and then 1µl being spotted directly on a MALDI plate for analysis.



The initial positive reflectron mode analysis using PNA matrix was conducted on a Voyager-DE STR analysed at 20Kv, taking 100 shots per spectra (with the beam moved manually over the spectra) over a 500-2000 Da mass range. The resulting data was analysed using the DataExplorer software package, with a noise filter/ smooth of 0.7 being applied to the data and masses determined in centroid mode. Different spectra were then aligned applying manual calibration using a selection of large reoccurring peaks as calibration peaks, acting as lock masses, thereby ensuring the reproducible masses are defined for each peak, as detailed in Table 3-1.

Table 3-1 Major ions consistently noted in MALDI spectra and used for manual calibration of spectra

Positive mode
518.5093
780.8450
1566.7128

A negative ionisation mode MALDI analysis was conducted on the serum samples using 9-aa matrix as per 2.3.1.2.4 MALDI extraction and analysis. After collection all data was analysed using Microsoft excel and SPSS. The data was first compiled into subsets for each group consisting of triplicate readings for each patient serum, listing the  $m/z$  vs relative intensity values obtained. These were then consolidated into  $m/z$ 's and intensity for all triplicate readings with  $m/z$ 's rounded to the nearest half a decimal place. To reduce the dataset to include only more robust signals,  $m/z$  readings that only occurred in 1 of the triplicate samples were removed from the analysis, before the average relative intensity was consolidated on a new worksheet for each  $m/z$  and patient. Any  $m/z$  values that then only arose in less than 15 global samples was also excluded from the search input.

### 3.3.3- Statistical analysis

Datasets were tested for normality using SPSS analysis before either a student's T-test or Kruskal-Wallis test was applied for pair-wise group comparisons based upon the outcome of the normality test. Any that were then found to be significant were further assessed using ROC analysis for diagnostic marker qualities as outlined 2.3.3. Biomarker evaluation.

### 3.4- Results

The first section of this study profiles the comparative global serum metabolites of four different study groups serum, analysed using different comparative techniques; GCMS and MALDI-ToF. The results will then additionally compare the serum profiles of candidates with endometriosis and PCOS drawing comparisons to healthy volunteers seeking alterations to amino acid abundances using GCMS methodology. Finally, this section will show the evaluation of breastmilk samples using MALDI-TOF techniques drawing comparisons between samples collected at two and six months post-partum.

#### 3.4.1- Comparisons of males, non-pregnant females, pregnant females and cord serum

##### 3.4.1.1- GCMS metabolomic analysis

Over three hundred different metabolites were detected as observable GCMS chromatographic peaks during the initial GCMS analysis, which after refinement provided one hundred and ten different candidate target peaks, which were then further considered. These were initially utilised for significant discriminatory variance between the four different subject groups. Of those deemed to be statistically significant, identifications were elucidated based on NIST 05 library comparisons, from which, any that could be resultant of artifacts rather than metabolite profile variance were excluded. This process generated a total of sixty-two significantly different metabolic markers. Figure 3-1a and b show clear illustrations of some of the visibly different regions of interest on the GCMS spectra through overlapping example chromatograms for one of each of the groups under study.

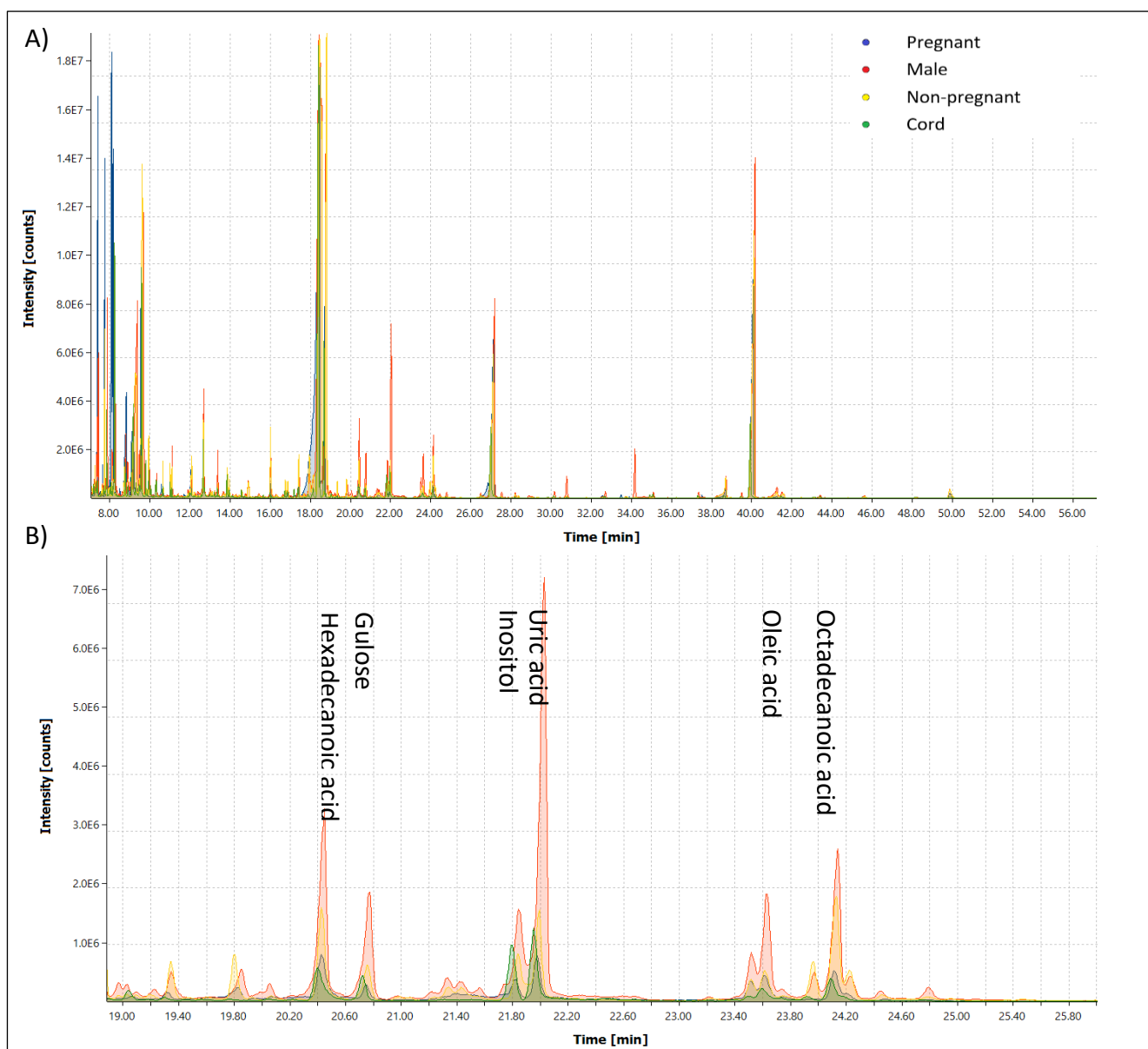


Figure 3-1 Example GCMS chromatogram for the four analysed groups: pregnant females (blue), Males (red), non-pregnant females (yellow) and cord serum samples (green). Figure 3.1B shows an extract of the chromatogram in more detail between 19 and 26 minutes.

The successful untargeted GCMS profiling method developed, utilising the metabolite markers of interest demonstrates an obvious group differentiation and creates a better understanding of the metabolic relationship patterns amongst and between the four groups. This can be visibly appreciated utilising the dendrogram shown in Figure 3-3, which shows the manner by which the metabolomic differences highlighted impact on the way in which the different samples relate to each other, with the most similar samples and groups being illustrated by closer branches and the branches towards the left side of the chart showing the greatest relative variance. The graph illustrates, despite considerable intragroup variation, there is a general good separation of all the four the major groups from one another. Three of the major groups, non-pregnant, cord and males, form generally good and distinct clusters. The lowest branch solely consists of most of the male and non-pregnant samples. The next branch up consists of a group of pregnant samples, which is also spread to the upper end of the chart where the majority of the cord samples are found. The pregnant group consists of generally smaller clusters, which are more spread apart from each other, which can be indicative of the differing and extreme metabolic demands faced during pregnancy term. This is also mirrored in the heatmap generated in Figure 3-2 that looks at the average metabolite level changes amongst the groups for the most significantly different metabolites of interest, illustrating how they fluctuate between the four groups. When using this form of differentiation, it is indicated that the pregnant and cord serums hold the greatest similarity to each other, as do males and non-pregnant females. Pregnant females generally have the lowest levels of all distinguishing metabolites of interest out of all the groups under evaluation, followed by cord serum that holds marginally more. Interestingly there was a lack of steroid hormones noticed in this study, and whilst they should be obvious varying markers noted, the extraction methodology used here was likely not specific enough for steroid hormones (Schmidt. G. and Steinhart, H., 2002).

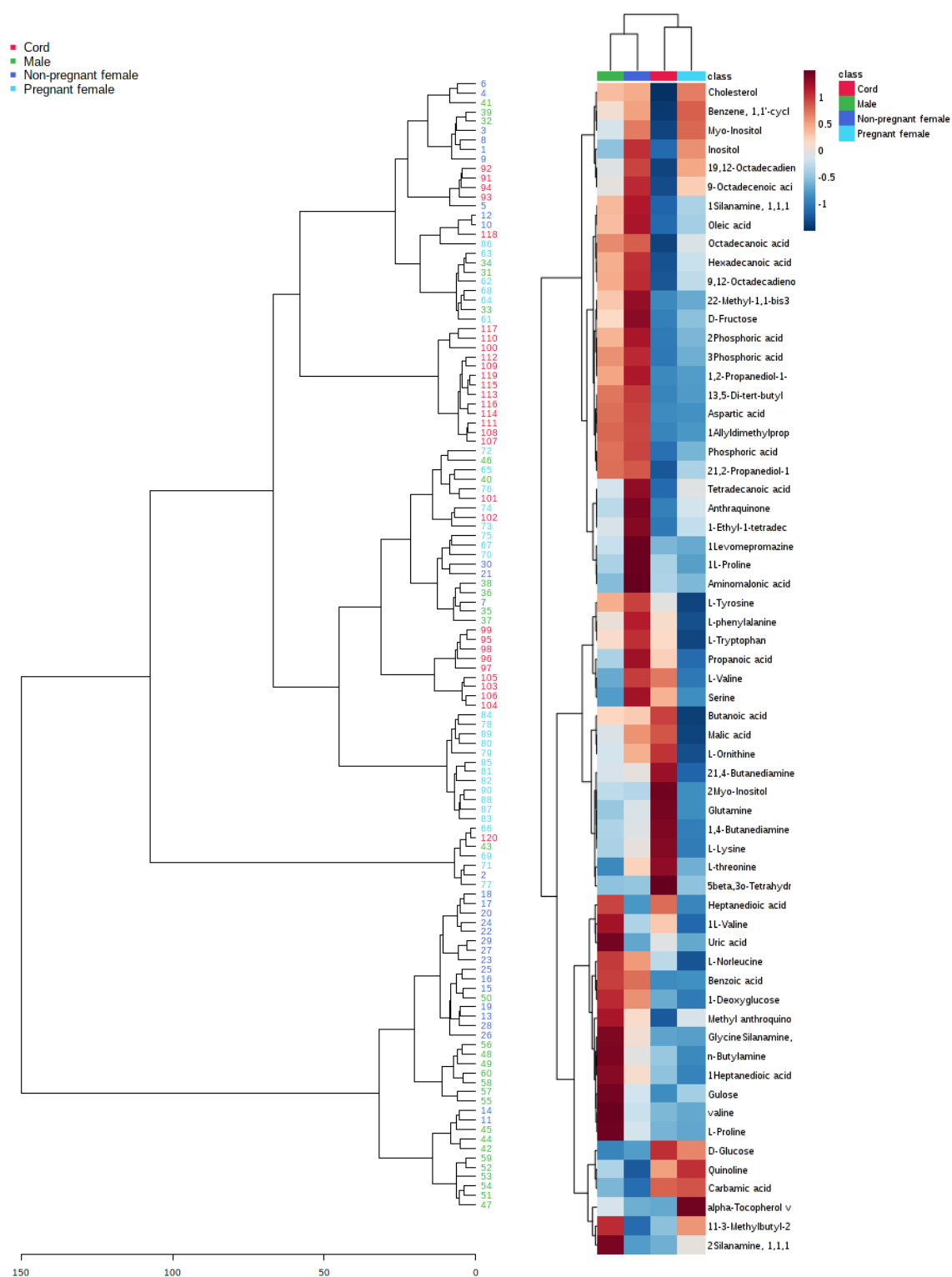


Figure 3-3 Dendrogram illustrating the relationship between the different samples and groups depicting red labels for non-pregnant females, green for males, Blue for pregnant females and red for cord serum samples

Figure 3-2 The average alterations in the significantly different metabolite profiles comparing non-pregnant females, males, pregnant females and cord serum. Relative abundance is shown with high expression in red and low expression in dark blue

#### 3.4.1.1.1- Non-pregnant females vs males

The first comparison to be undertaken was one based solely upon gender in which metabolite profiles from males were compared to similar profiles taken from non-pregnant females. A total of 22 significantly different metabolic features were identified when comparing in this manner using GCMS analysis. A good degree of separation was achieved, Figure 3-4, with a larger spread across the data shown for male serum when compared to non-pregnant serum. A broad range of metabolite classes were found to show significant variance between the two classes of serum samples, the most abundant of which include nine amino acids, six fatty acids and three sugars amongst others.

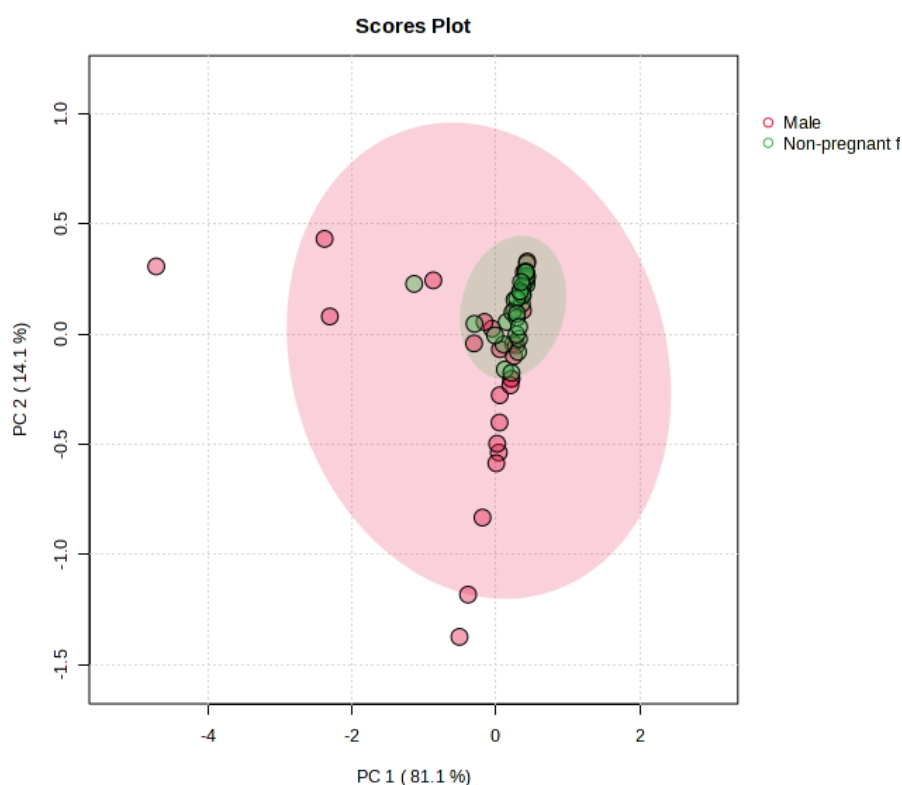


Figure 3-4- PCA chart showing the differentiation of males and non-pregnant women's serum samples from each other using GCMS analysis using biomarkers found to be the most diagnostically capable of distinguishing the groups.

The average abundance changes between the most significantly different metabolites found are displayed in Figure 3.4, with a mix of 15 of the classes being up-regulated in non-pregnant females and 7 in males. Note-worthy all of the lipids with the exception of heptanedioic acid were seen to be up-regulated in non-pregnant females compared to males.

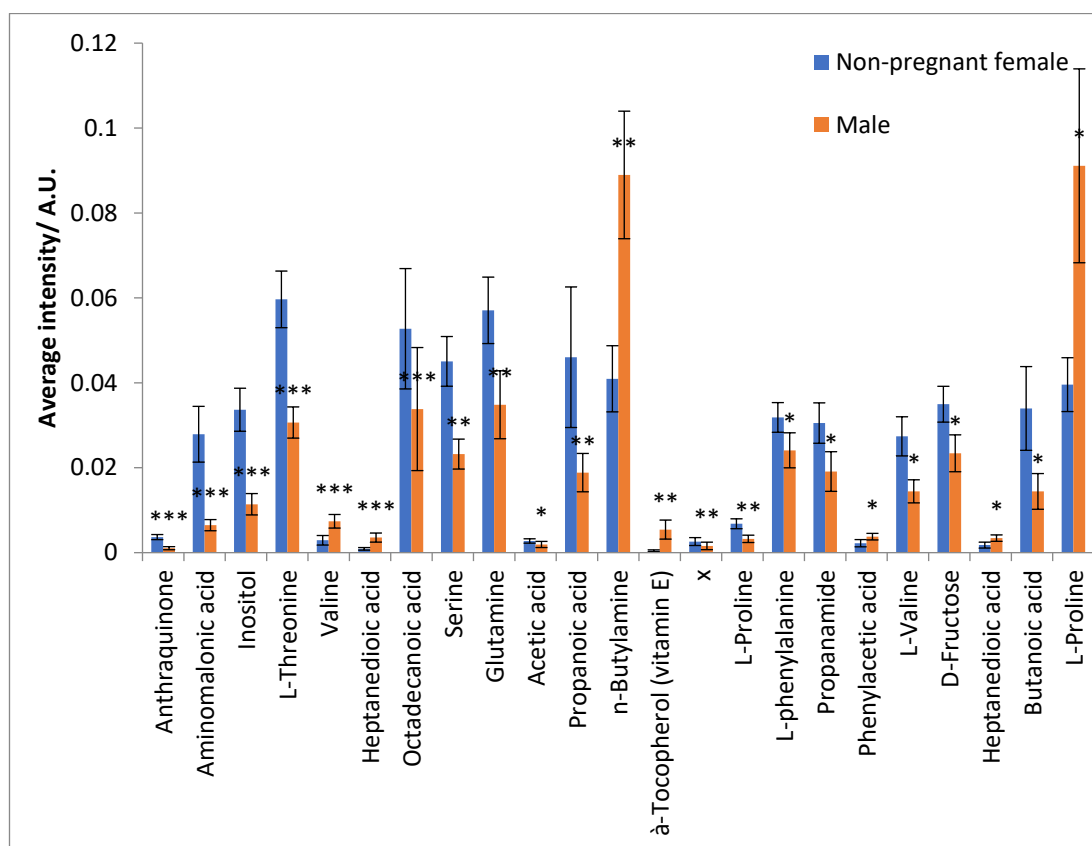


Figure 3-5- The average normalised relative abundance of the 22 most significantly different metabolites identified between males (red) and non-pregnant females (blue), in order of most to least significantly different, with standard error plotted. Bars represent normalised average abundance, with error bars representing the standard error (n=30 for both groups). X- denotes an unidentified metabolite. Significance is shown with asterisk with \*>0.05 \*\*>0.025\*\*\* > 0.005

Though many significant differences were noted in the evaluation, none of these metabolites achieved suitable diagnostic scores for further use in subject identification as solo markers. The changes exhibited in Figure 3-5, are significant, with the most abundant change being found for à-tocopherol (vitamin E) giving a fold change of 11.58, whilst the average fold change excluding this value was that of 2.96. Table 3-2 details the diagnostic qualities of the most significantly different

metabolites between non-pregnant females and males showing that all obtained very good P value scoring. Sensitivity and specificity levels (the inverse of 1-specificity as generated by the ROC curves) were generally very good with many of the metabolites providing values above 70% for both of these and Area Under the Curve (AUC) values ranging from a maximum of 0.749 to 0.654) which provides some evidence of differentiation based upon the metabolite levels. However, despite this, due to poor LHR scores being generated, there is the possibility that a number of false positive and false negative results may be generated and hence none of the metabolites when studied alone would be suitable markers independently for group separation based solely upon gender alone.

Table 3-2- Most significantly different and diagnostically capable metabolites identified between non-pregnant females and males ranked in order of significance. Excellent scorings are shown in green, good in yellow and poor in red. In some instances, i.e. heptanedioic acid, the same identification has been generated multiple times being the identity assigned to multiple different peaks. This can occur for numerous reasons relating to the limitations of GCMS tentatively assigned identifications. Firstly, they can be associated with different isomers of the same amino acids, or they can in fact represent different metabolites but have similar identifications which couldn't be distinguished.

Metabolite	p. value	AUC	Sensitivity	1-Specificity	LHR +	LHR -
Aminomalonic acid	1.16E-03	0.743	75	30	2.50	0.36
Inositol	1.41E-03	0.737	70	30	2.33	0.43
L-threonine	1.48E-03	0.739	75	30	2.50	0.36
Valine	1.74E-03	0.726	75	30	2.50	0.36
Heptanedioic acid	3.22E-03	0.712	75	40	1.88	0.42
Octadecanoic acid	4.97E-03	0.703	75	35	2.14	0.39
Serine	5.44E-03	0.709	75	40	1.88	0.42
Glutamine	6.26E-03	0.705	75	30	2.50	0.36



Acetic acid	9.62E-03	0.684	75	20	3.75	0.31
Propanoic acid	1.08E-02	0.692	75	45	1.67	0.46
n-Butylamine	1.27E-02	0.684	75	45	1.67	0.46
à-Tocopherol (vitamin E)	1.28E-02	0.646	40	15	2.67	0.71
x	1.93E-02	0.663	65	30	2.17	0.50
L-Proline	2.02E-02	0.674	70	40	1.75	0.50
L-phenylalanine	2.41E-02	0.669	75	45	1.67	0.46
Propanamide	2.72E-02	0.666	80	30	2.67	0.29
Phenylacetic acid	3.03E-02	0.653	65	30	2.17	0.50
L-Valine	3.49E-02	0.658	75	45	1.67	0.46
D-Fructose	3.56E-02	0.658	75	40	1.88	0.42
Heptanedioic acid	3.78E-02	0.652	65	40	1.63	0.58
Butanoic acid	3.89E-02	0.654	80	55	1.46	0.44
L-Proline	3.98E-02	0.654	65	35	1.86	0.54

#### 3.4.1.1.2- Non-pregnant females vs pregnant females

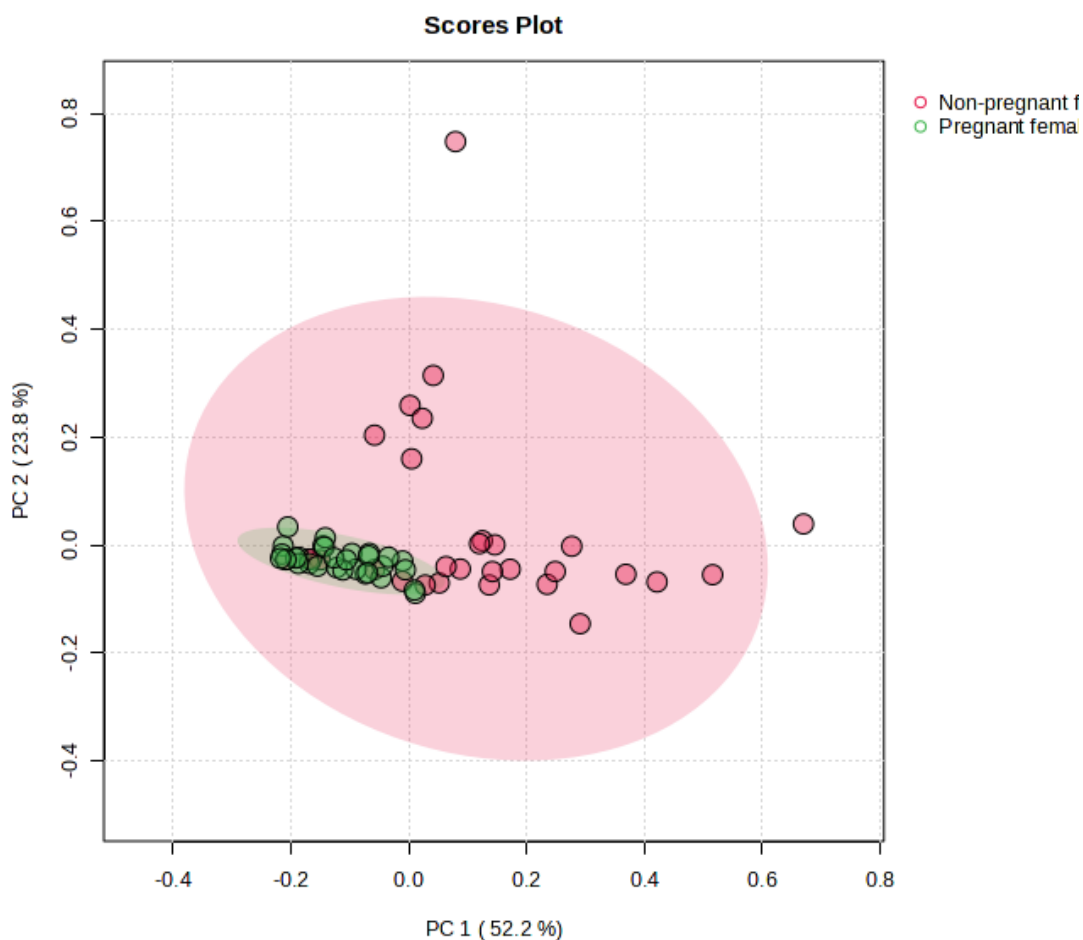


Figure 3-6-PCA chart showing the separation of non-pregnant and pregnant females serum metabolites analysed using GCMS analysis.

Next within the female derived samples taken for analysis, the ability of the metabolite profile to distinguish between the pregnant and non-pregnant cohorts was studied. The analysis generated a good level of separation between pregnant and non-pregnant females samples, Figure 3-6. After significance evaluation of the metabolites between non-pregnant and pregnant serum samples, a total of twenty-one significantly different metabolic characteristics were noted between cohorts (Figure 3-7). Most of the metabolites of difference were found to be amino acids, fatty acids or sugars. Furthermore, it is noteworthy that all but six of the twenty-one significantly different metabolites were shown to be in lower levels in pregnant candidate serum samples compared to the non-pregnant cohort. The range of metabolite levels was quite broad (see glycine versus glyoxylic acid in Figure 3-7),

however unlike the male vs female comparison, the pregnant vs non pregnant female comparison demonstrated a larger fold change in certain metabolites. For example, butanoic acid produced a fold change of 1.31 in this comparison, whereas it was 0.58 between males and females. Similarly, glutamine had a fold change of 1.60 in this evaluation and was only 0.39 between males and non-pregnant females. The largest fold change was seen to be 1-Deoxyglucose with a value of 8.4 and the average overall fold change of the significant markers was 1.17.

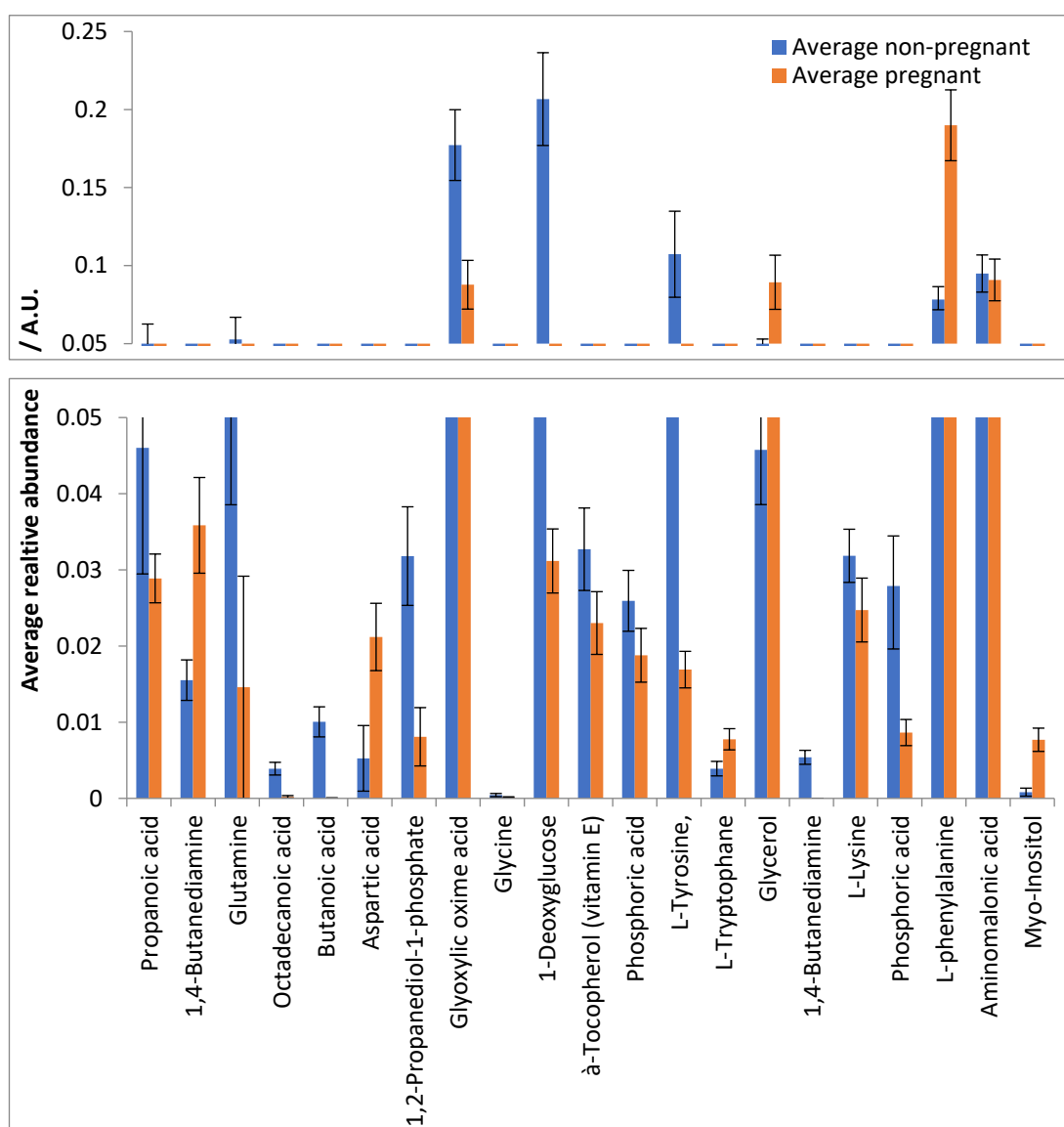


Figure 3-7- The 21 most significantly different metabolites identified between non-pregnant and pregnant females ordered from most significant on the left. Bars represent normalised average abundance, with error bars representing the standard deviation (n=30 for both groups)

Despite several of the metabolites holding good significant values (with P. values ranging from  $3.98 \times 10^{-2}$  to  $3.62 \times 10^{-4}$ ) and with some metabolites being completely absent from one cohort, none were reported to have sufficient diagnostic qualities regarding true positive and true negative diagnostic potential, quantified as likelihood ratios (Table 3-3). Considering the sensitivity and specificity values obtained however, demonstrates a superior separation of the two cohorts compared to the male versus non-pregnant female comparison, therefore suggesting that the metabolomic differences are more significantly robust in the comparison of the two female cohorts and that pregnancy itself has a far greater metabolomic effect in females compared to gender alone.

Table 3-3 Most significantly different and diagnostically capable metabolites identified between non-pregnant and pregnant females. “Excellent” markers are shown in green in the table and “good” and “poor” in red.

Metabolite ID	p. value	AUC	Sensitivity	1-Specificity	LHR+	LHR-
Propanoic acid	9.69E-08	0.901	90	26.7	3.37	0.14
1,4-Butanediamine	9.47E-07	0.866	96.7	36.7	2.64	0.05
Glutamine	2E-06	0.854	80	13.3	6.02	0.23
Octadecanoic acid	1.67E-05	0.776	0.6	0.1	6.00	0.44
Butanoic acid	3.01E-04	0.752	0.633	0.067	9.45	0.39

#### 3.3.1.1.3- Pregnant female vs Cord serum sample

A further comparison was then made between the serum from pregnant females and the serum metabolite profile from cord serum samples. Whilst not matched, the aim was to study the metabolite profile in the parents circulatory system compared to that derived from the cord serum upon returning illustrating metabolites used and generated by the foetus/placenta. This in turn can suggest key metabolites supplied to the foetus that may be essential to premature babies that can be looked at in the

later breast-milk study. When distinguishing the metabolic differences between pregnant women and cord samples, 42 differences in total were found to be significance. Through establishing the most useful metabolic differences for determining the groups from each other a total of 16 metabolites were then established to be the most significant using Bonferroni correction and assigned identities via NIST. Further critique of the 16 metabolites using diagnostic scoring indicated that 5 of these metabolites when used as individual markers achieved better than “good” diagnostic qualities according to AUC, LHR + and LHR- values for group differentiation. The five most diagnostically significant metabolites are shown in Table 3-4.

Table 3-4- The 5 most diagnostically successful metabolites identified between pregnant and cord serum samples. It should be noted that an  $\infty$  symbol is utilised to represent where the value cannot be computed. In this instance LHR+ is calculated by taking the Sensitivity/ 1- Specificity, however in this instance the 1-specificity is zero. “Excellent” markers are shown in green in the table and “good” and “poor” in red.

Metabolite ID	p.value	AUC	LHR+	LHR-
5 $\beta$ ,3 $\alpha$ -Tetrahydro-medroxyprogesteronacetate	1.57E-11	0.967	$\infty$	0.07
Propanoic acid	4.11E-09	0.942	9	0.11
Myo-Inositol	1.05E-08	0.93	5.79	0.04
1,4-Butanediamine (adduct 1)	1.41E-08	0.922	5.59	0.08
1,4-Butanediamine (adduct 2)	1.04E-08	0.908	8.67	0.15

The P. values presented in Table 3-4 indicate that the differences between these 2 samples are significantly more pronounced than previous comparisons and the area under the ROC curve values all meet the generally recognised threshold for being identified as an excellent biomarker or means to differentiate 2 cohorts. The positive likelihood ratio of the progesterone derived metabolite is not calculable as the peak

is absent from the pregnant female sample, but present in sufficient abundance in enough cord derived samples to be determined.

Figure 3-8 shows the 5 most diagnostically valuable metabolites that bared significant variance between the pregnant female and cord serum samples. In addition to these 5 of the other metabolites that proved significant difference between pregnant and cord serums were largely made up of amino acids and fatty acid groups. Two different metabolites that were distinct from each other as GCMS chromatographic peaks but both providing NIST generated identifications of 1,4-Butanediamine showed significant variance between the two different serum profiles under analysis. In both instances the compound was found to be in significantly higher proportions in the cord serum sample compared to the pregnant female. This metabolite was also found to vary in the previous analysis comparing non-pregnant and pregnant females.

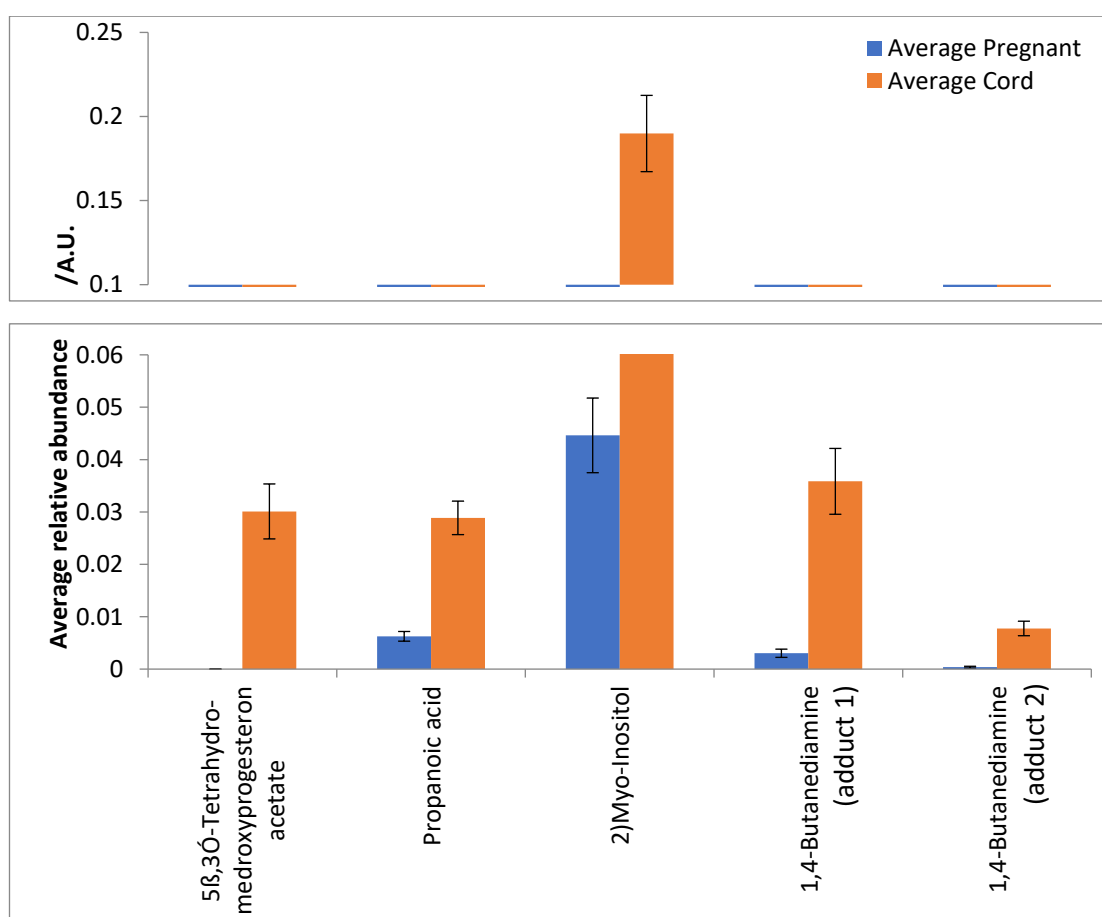


Figure 3-8- The most diagnostically successful metabolites capable of distinguishing pregnant (blue) and cord (red) serum samples from each other in order of significance. The average relative

abundance for the groups is illustrated on a broken scale with standard error bars representing variation in the groups (n=30 for both groups).

A vast majority of the differing metabolites detected using GCMS methodology were found to be amino acids, Figure 3-9 demonstrates the average fluctuations of these amino acids between the four different groups under study.

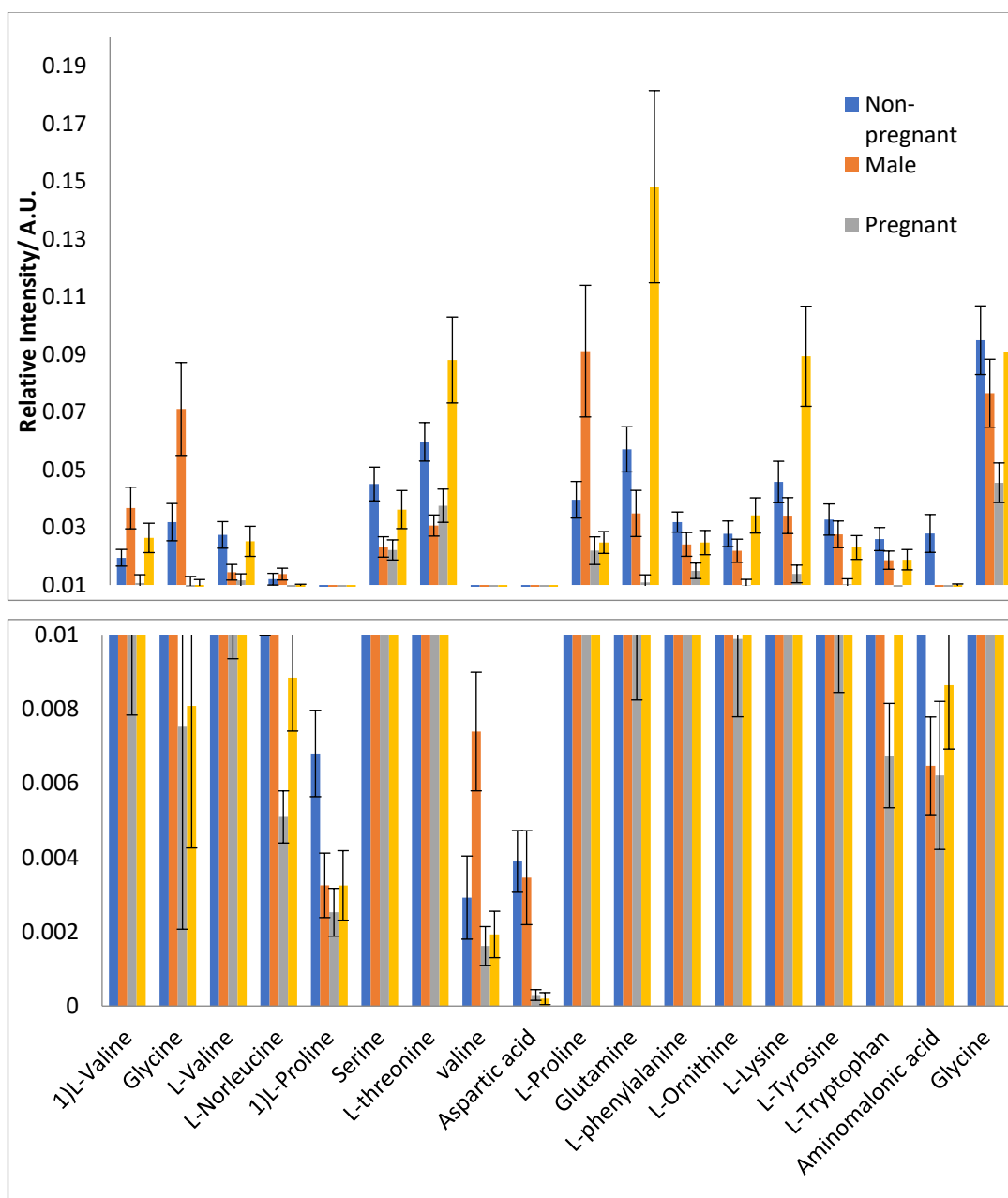


Figure 3-9- Graph illustrating the average abundance of the metabolites indicated as being significantly different between the four different groups with error bars shown (n=30).

### 3.3.1.2- MALDI lipidomic analysis

This section of the study sought, through MALDI derived lipidomic analysis, to identify significant differences between the four different groups; males, non-pregnant females, pregnant females and cords. Different matrices and ionisation modes were trialled to find the most efficacious method for the differentiation of the different



groups. The analysis was split between negative and positive ionisation mode analysis with the negative analysis divided into 2 analyses based upon the use of 2 MALDI matrices.

A number of different lipid classes have been identified as showing significant variance in this study with acronyms for these shown in Table 3-5.

Table 3-5- Acronyms used to denote different lipidomic species during MALDI analysis

L (prefix)	Lyso
PC	Glycerophosphocholines
FA	Fatty Acyl
ST	Sterol
CAR	Acyl carnitine
Cer	Ceramide
PE	Glycerophosphoethanolamines
PS	Glycerophosphoserines
DG	Di(acyl   alkyl)glycerols
PG	Glycerophosphoglycerols
PI	Glycerophosphoinositols

It should also be noted that for identification clarity the specific adduct present was not included in the identifiers name and so there may be instances of multiple identifications of a particular lipid but these might be H<sup>+</sup> or K<sup>+</sup> adducts and thus recorded different mass values.

#### 3.3.1.2.1- Negative mode lipidomic analysis – 9AA MALDI matrix

During the negative mode MALDI analysis of the serum samples, the matrix 9-AA was utilised to detect changes in lipidomic profiles. During this analysis of 360 samples, consisting of triplicate readings of the four (n=30) patient cohorts, 2,506 *m/z* values were initially entered into the analysis that aimed to compare serum samples using negative reflectron mode. This data set was consolidated to 252 *m/z* variables upon

exclusions of those initially with low relative intensities and further then to those that were not consistently found in at least 2 of the three replicate sample sets and not found in at least 15 different patient samples. Figure 3-10 shows example spectra for the 4 cohorts analysed in this study. Individual cohorts were then compared as undertaken in the previous negative MALDI and GCMS analysis.

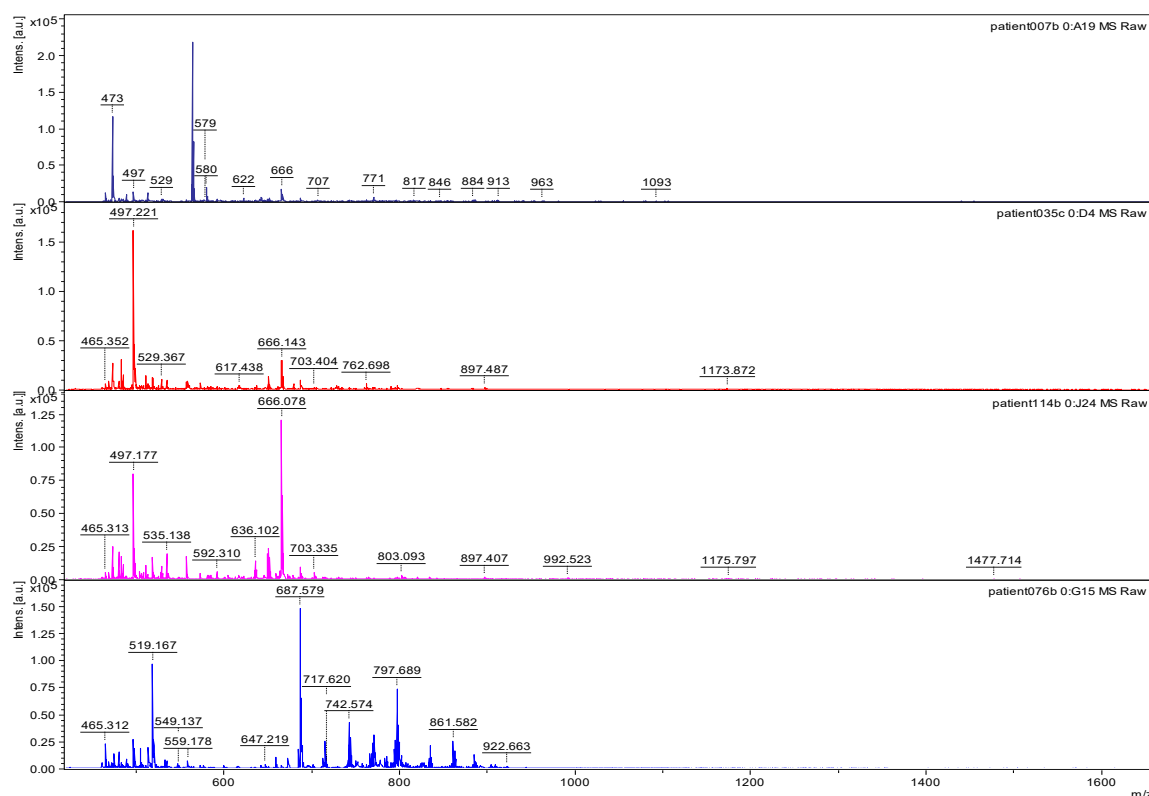


Figure 3-10 Example MALDI spectra for Non-pregnant females (dark blue), Males (red), cord (pink) and pregnant females serum samples (royal blue) obtained using 9-AA matrix in negative reflectron mode.

The common differences were used to generate PCA plots in order to determine how well they could categorise the four different groups in question from each other (). The plot shows a good level of separation of all of the groups in question. The most obvious overlap is seen between non-pregnant and pregnant women, with these groups forming overlapping clusters to the left hand upper quarter of the chart. Male samples are all clustered together in the upper right, whereas cord samples are the most dispersed group but found towards the bottom right corner, quite separate from the other groups in question which illustrates their more unique nature.

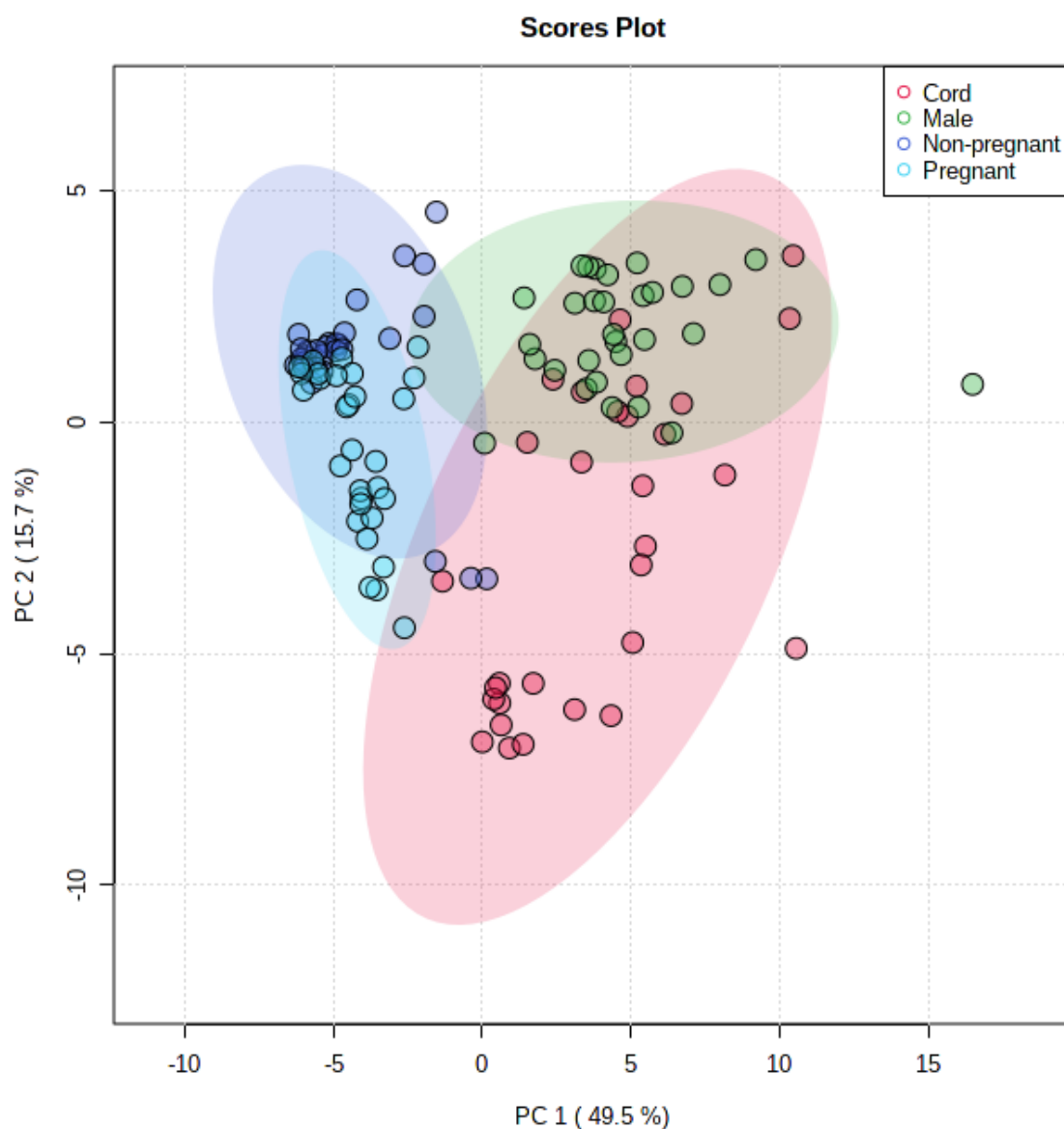


Figure 3-11- PCA chart showing the differentiation of cord, males, pregnant and non-pregnant women's serum samples from each other utilising MALDI-TOF analysis in negative mode using biomarkers found to be the most diagnostically capable at distinguishing multiple groups of interest.

A better understanding behind the basis for the separation can be seen in the heatmap (). This shows that in the vast majority of markers observed the intensity is greatest in male candidates, followed then by cord. Pregnant and non-pregnant women both have the lowest levels of these markers but differ on a small number of other markers that were found to be lowest in male candidates. In these instances,

non-pregnant women had the greatest intensity of the markers than pregnant women. The heatmap also shows that when considering the average intensity of each group males and cord samples are the most similar to each other, as are pregnant and non-pregnant women.

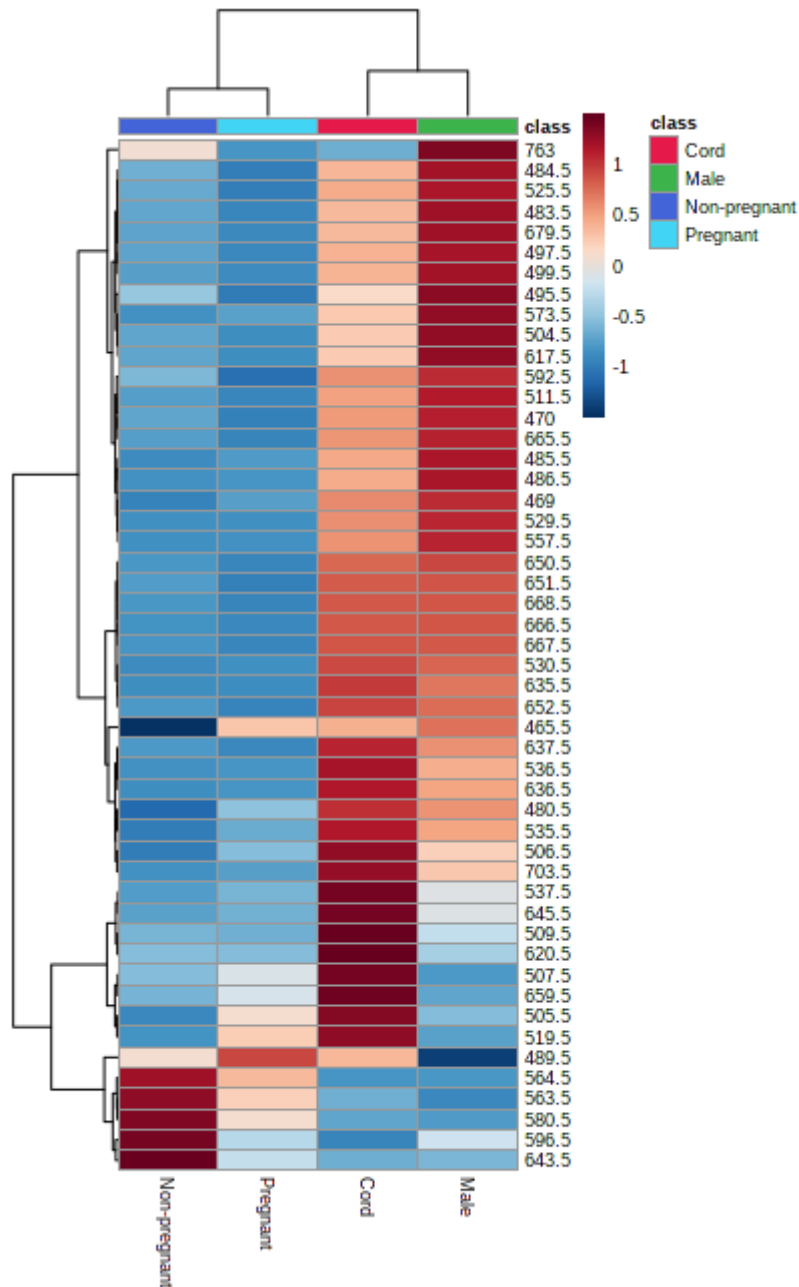


Figure 3-12- Heatmap illustrating the average intensity comparison of the shared markers of interest between the 4 different demographical groups of interest; males, non-pregnant females, pregnant females and cord serum.

Upon achieving a methodology that could separate out the four groups successfully from each other the next step was to identify which markers in the comparisons of interest were responsible for the differences seen.

#### Non-pregnant vs Male

Initial statistical analysis identified a total of one-hundred and forty-five significantly different lipid markers that were found to be capable of distinguishing non-pregnant females from males utilising MALDI analysis in negative mode with 9AA as the matrix. Fifty-six of these held better than good diagnostic capabilities in all regards. As was done in the previous analysis, isotopic variants were then excluded, as example of such is shown in Figure 3-13.

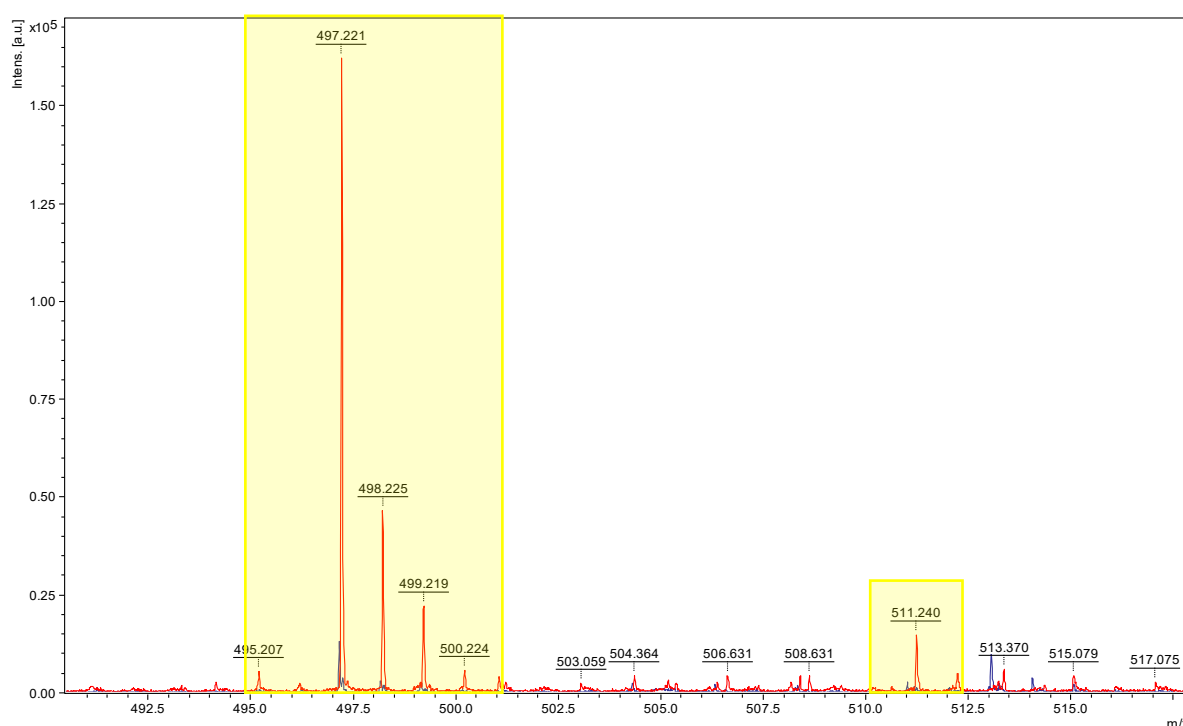


Figure 3-13- Overlaid Non-pregnant (blue) and male (red) MALDI spectra showing two of the significantly different markers identified. One was highlighted as “excellent” in the analysis, 497.5. With a further two 498.5 and 499.5 which were identified as “good” which actually show isotopic variants of the principle 497.5. This is clear from the approximate third decrease in intensity seen

between each peak as it increases in  $m/z$  by 1 unit. These can therefore all be assigned as being the same marker. The other is at 511.5

Following analysis, 41 significantly different  $m/z$  values were rated as being better than good relating to their individual diagnostic potential with accordance to ROC scoring. The fluctuations of which are shown in Figure 3-14, detailing the average relative abundance for species and the variation in signal across the cohort. The majority of the identified lipid markers were found to be more prominent in males when compared to non-pregnant female candidates with the exception of lipid species at  $m/z$  563.5, 564.5 and 580.5.

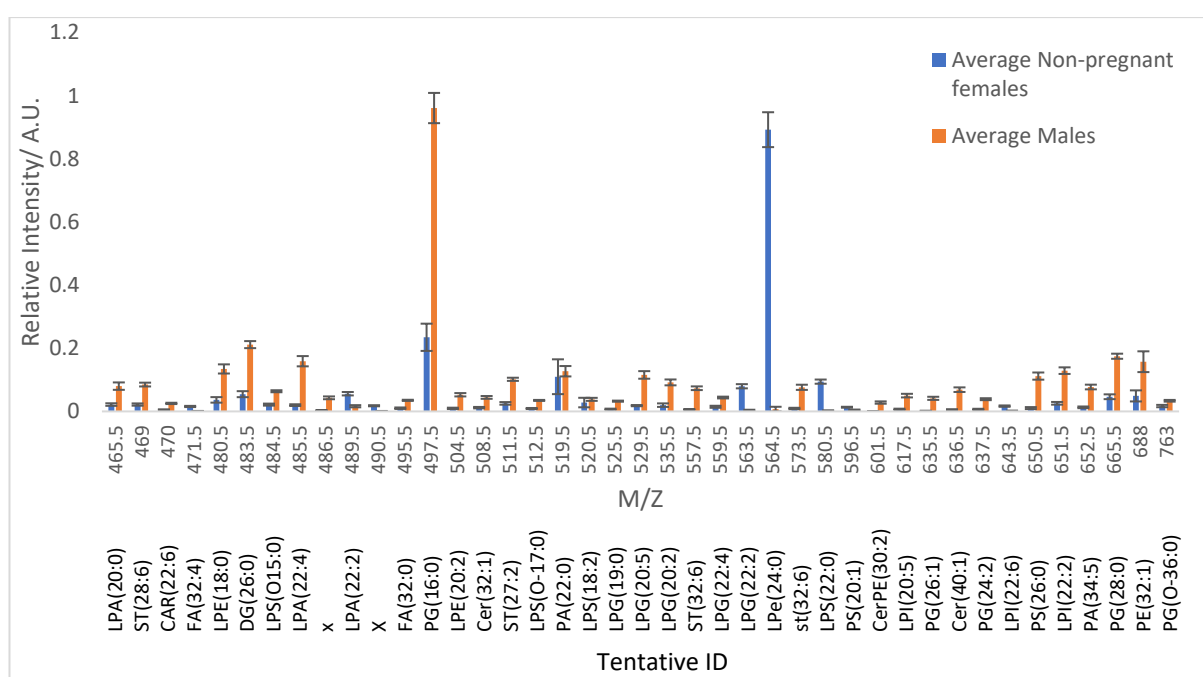


Figure 3-14- Average peak intensities for the significantly different metabolites found to differentiate non-pregnant females from male serum cohorts that were rated better than good diagnostically.

Error bars represent standard deviation. N=30 for both cohorts.

Eighteen of these markers were rated as excellent in all diagnostic properties, Table 3-6. All bar one of these markers were most elevated in males, only  $M/Z$  564.5 (LPE(24:0)) was up-regulated in non-pregnant women during the comparison. This particular marker, 564.5- LPE(24:0), was a particularly prominent discernible difference noted between non-pregnant females and males, Figure 3-15 displays example spectra showing the vast alteration in this marker seen.

Table 3-6- Diagnostic qualities reported for the significantly different metabolites distinguishing males from non-pregnant female cohorts rated as excellent for all diagnostic qualities. “Excellent” markers are shown in green in the table and “good” and “poor” in red.

<i>m/z</i>	p. value	AUC	LHR+	LHR-
471.5	1.00E-03	0.957	29.30	0.03
485.5	1.00E-03	0.993	29.30	0.03
490.5	1.00E-03	0.944	28.27	0.07
497.5	1.00E-03	0.978	13.93	0.07
511.5	1.00E-03	0.982	28.27	0.07
529.5	1.00E-03	0.991	30.30	0.00
530.5	1.00E-03	0.981	∞	0.03
557.5	1.00E-03	0.994	29.30	0.03
564.5	1.00E-03	0.964	28.27	0.07
581.5	1.00E-03	0.962	28.27	0.07
637.5	1.00E-03	0.962	13.93	0.07
650.5	1.00E-03	0.984	30.30	0.00
651.5	1.00E-03	0.984	14.93	0.00
652.5	1.00E-03	0.977	14.43	0.04
666.5	1.00E-03	0.983	30.30	0.00
667.5	1.00E-03	0.983	30.30	0.00
668.5	1.00E-03	0.983	30.30	0.00
679.5	1.00E-03	0.967	13.93	0.07

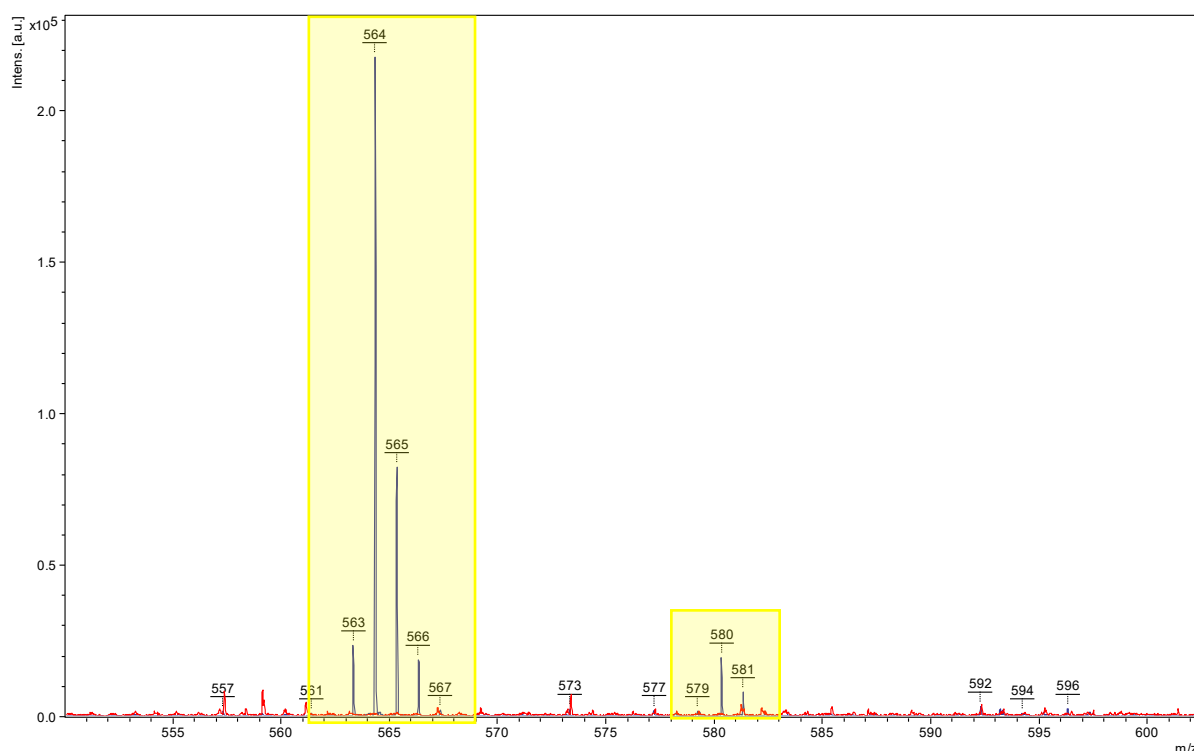


Figure 3-15 Overlaid non-pregnant (blue) and male (red) MALDI spectra showing two of the significantly different markers identified. Both of which were highlighted as diagnostically “excellent” in the analysis. The first of these is sat  $m/z$  564.5, followed by a further 2 isotopic variants at 565.5 and 566.5. The other marker is  $m/z$  of 581.5 and is found to be significantly up-regulated in males when being compared to non-pregnant females.

### *Non-pregnant vs pregnant females*

There were one-hundred and fifty statistically significantly different lipid markers found to differentiate non-pregnant and pregnant females. Of these differences noted, only two scored better than “good” when evaluated for individual biomarker quality relating to ROC scoring, these are shown Figure 3-16 with 539.5 achieving a p.value of 0.001, LHR+ of 6.52 and LHR- of 0.15 and marker 580.5 a p. value of 0.001, LHR+ of 5.19 and LHR- of 0.16. Both of these candidate markers LPG(20:0) and LPS(22:0) were found to be significantly up-regulated in non-pregnant women in comparison to pregnant women. Interestingly, this time far fewer lipid signals were



able to differentiated pregnant and non pregnant females compared to the male vs female comparison using the same matrix.

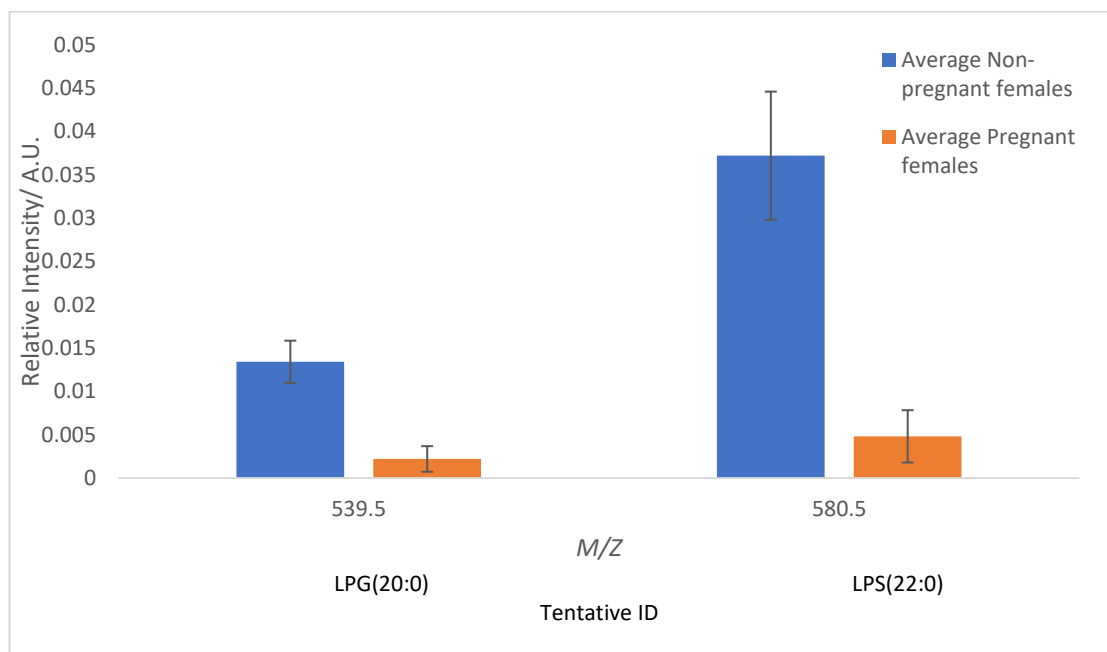


Figure 3-16 Average relative abundance of the metabolites which were major sources of variation between non-pregnant females from pregnant females. Error bars represent standard deviation. N=30 for both cohorts.

#### Pregnant female vs cord serum

One hundred and seventy-four metabolic difference significantly differentiated pregnant females from cord serum samples under analysis. Of these, twenty-six were rated as being better than “good” diagnostically and eleven “excellent” when considering ROC scoring capabilities. All of the metabolites identified in this analysis

showed significant elevation in cord serum samples as opposed to the pregnant female serum cohort (Figure 3-17).

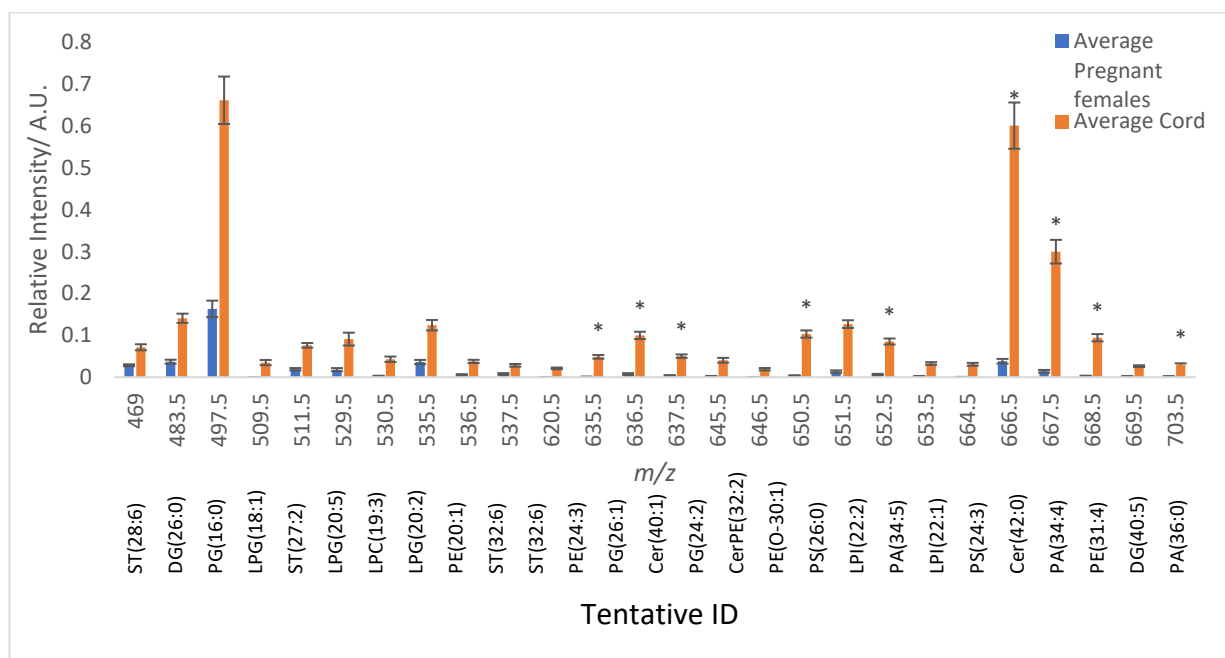


Figure 3-17 The most diagnostically capable biomarkers, rated as “good” found to distinguish Pregnant women and cord serum from each other using MALDI-TOF in negative mode. Error bars represent standard deviation. N=30 for both cohorts. The asterisks illustrate markers rated as better than excellent diagnostically.

The diagnostic qualities of these markers are all rated as being better than good, with quality markers listed in Table 3-7.

Table 3-7- The diagnostic values associated with the significantly different metabolic markers that are capable of distinguishing pregnant females serum and cord serum from each other. “Excellent” markers are shown in green in the table and “good” and “poor” in red.

M/Z	p.value	AUC	LHR+	LHR-
469	0.001	0.951	5.389222	0.120048
483.5	0.001	0.96	7.270677	0.038062
497.5	0.001	0.958	9.33	0.074444
509.5	0.001	0.948	∞	0.1
511.5	0.001	0.953	9.67	0.036667

529.5	0.001	0.939	7.015038	0.077278
530.5	0.001	0.893	25.24242	0.172699
535.5	0.001	0.923	5.586826	0.080432
536.5	0.001	0.911	9	0.111111
537.5	0.001	0.856	6.518797	0.153403
620.5	0.001	0.948	$\infty$	0.1
635.5	0.001	0.982	$\infty$	0.033
636.5	0.001	0.975	29.30303	0.034126
637.5	0.001	0.978	29.30303	0.034126
645.5	0.001	0.928	13.43284	0.107181
646.5	0.001	0.862	$\infty$	0.067
650.5	0.001	0.978	$\infty$	0.033
651.5	0.001	0.999	9.67	0.036667
652.5	0.001	0.976	$\infty$	0.033
653.5	0.001	0.92	$\infty$	0.133
664.5	0.001	0.893	$\infty$	0.2
666.5	0.001	0.986	$\infty$	0.033
667.5	0.001	0.974	$\infty$	0.033
668.5	0.001	0.979	$\infty$	0.033
669.5	0.001	0.94	$\infty$	0.1
703.5	0.001	0.962	$\infty$	0.067

The commonly found markers found to be capable of distinguishing more than one group from each other and additionally maintained a better than good diagnostic potential were amalgamated together and shown in Figure 3-18.

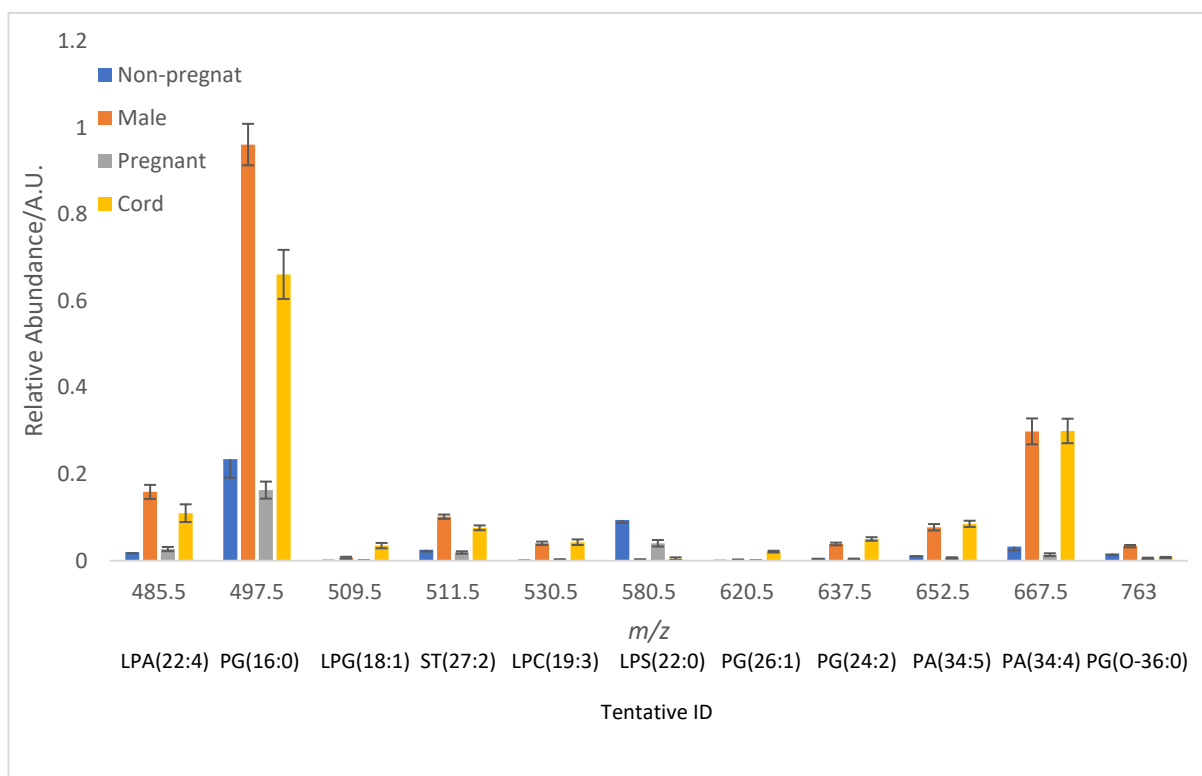
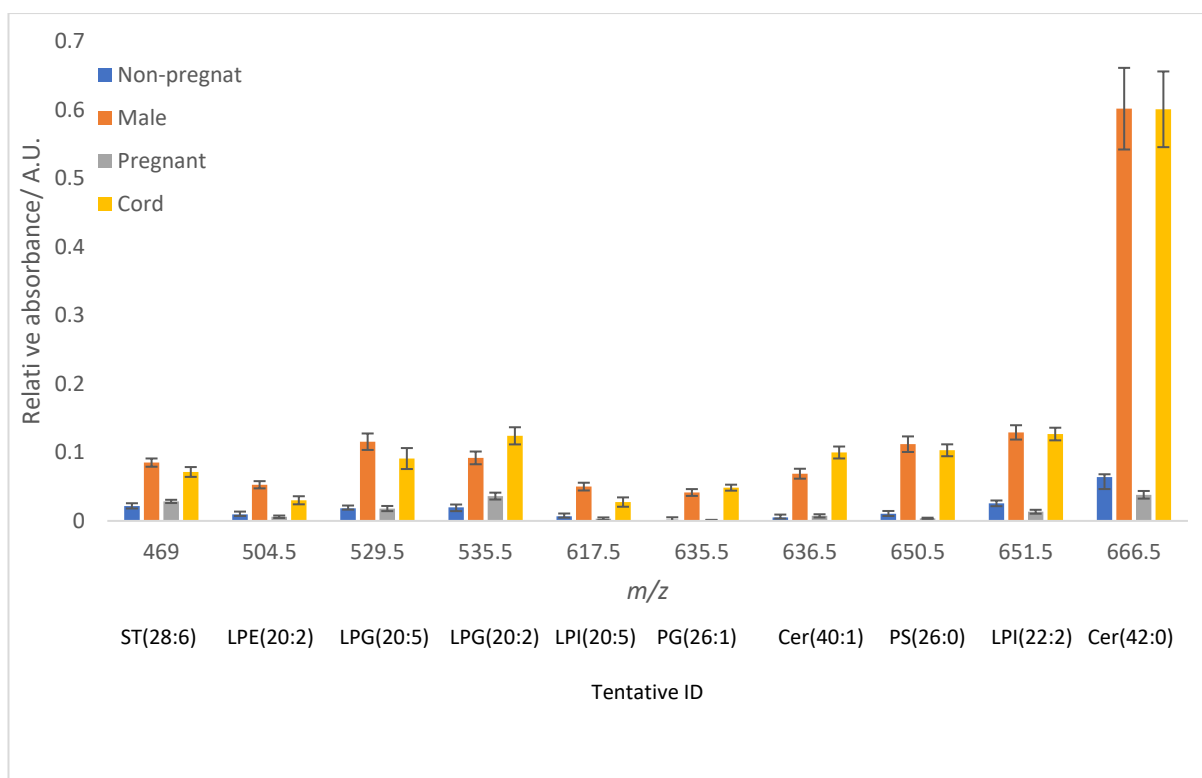
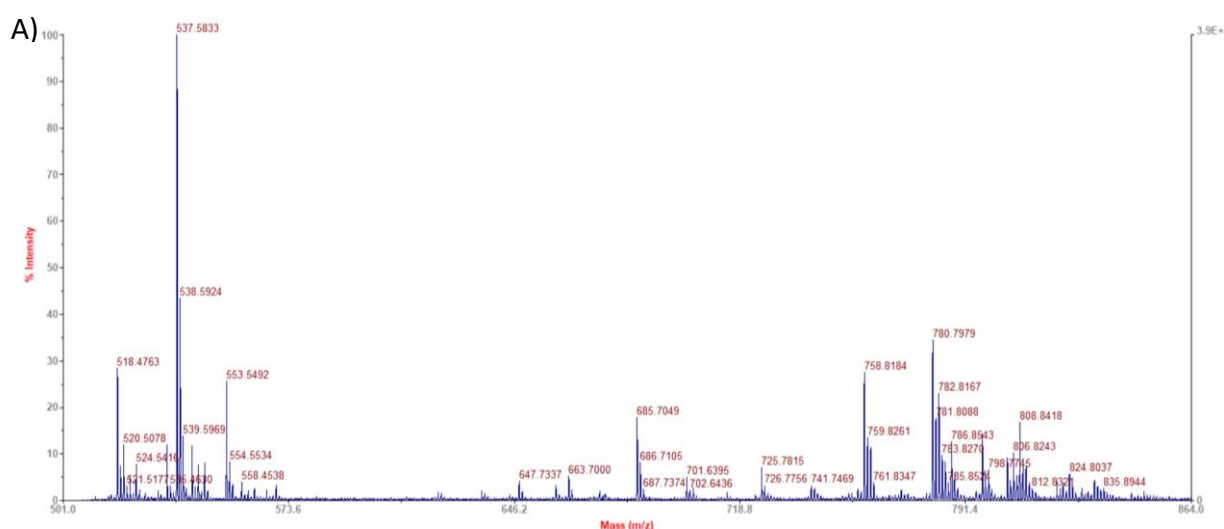


Figure 3-18- The common most biologically capable biomarkers found to be able to distinguish more than one of the demographical groups, male, non-pregnant females, pregnant females and cord serum from each other.

### 3.3.1.2.2- Positive mode lipidomic analysis with DHB matrix

The final MALDI lipidomic analysis utilised positive ionisation mode and DHB as a MALDI matrix and the resulting spectra were predominated by lipidomic species with natural positive charges or an improved ability to attract positively charged species, mainly phosphatidylcholine (PC) and sphingomyelins (SM) lipids (Figure 3-19). Analysis of the cohorts was then undertaken as described previously.



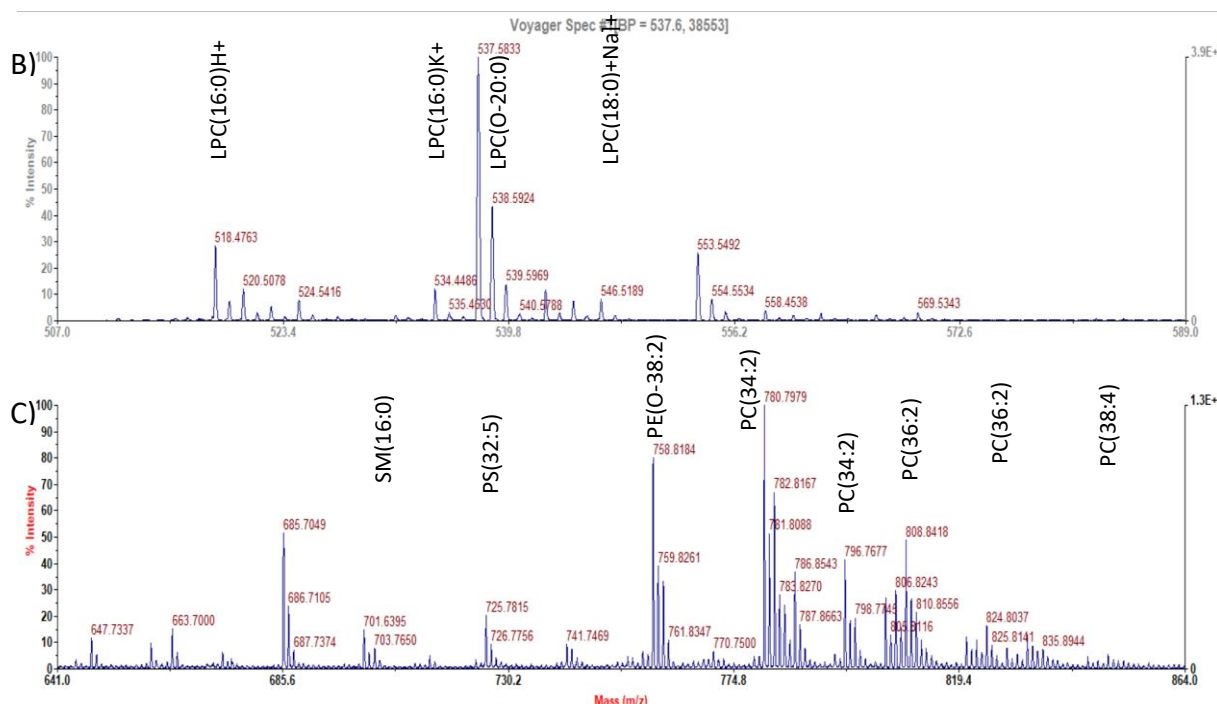


Figure 3-19. Example MALDI spectra of lipid extraction from a non-pregnant female's serum ran in positive ion mode; a) displays a full scan focused on the area of most interest  $m/z$ 's between 500 and 860 b) focused on the LPC content under analysis C) Areas with SM and LPC

To establish the ability of the identified markers to define the individual groups a PCA plot was generated, Figure 3-25 and also a heatmap showing the group fluctuations and how they impact the groups closeness in Figure 3-26.

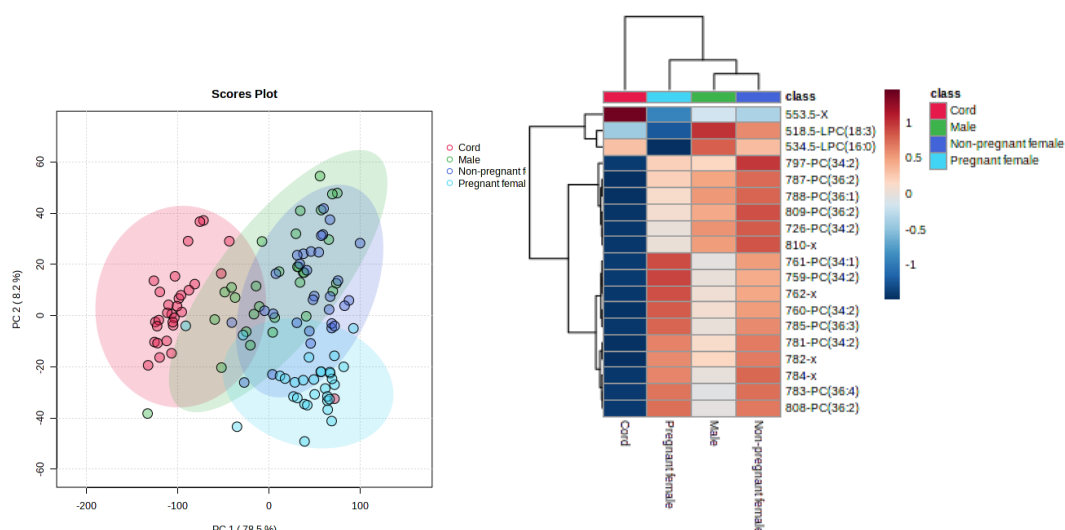


Figure 3-20- PCA plot showing the division of the different groups utilising the most significantly different metabolites identified that appeared to distinguish multiple groups

Figure 3-21- Heatmap showing the average metabolite intensity change of the most distinguishing markers used to generate the PCA plot including an illustration of the these impact the way in which the groups relate to each other

### *Non-pregnant females vs Males*

There were twenty-six significant markers identified which exhibited intensity differences between non-pregnant female and male candidates serum samples when analysed using MALDI-TOF in positive reflectron mode. None of these candidates however scored better than good in all diagnostic qualities when analysed using ROC curves. The significantly different features did not achieve better than good in all areas of diagnostic potential relating to AUC, LHR+ and LHR- scores across the board (Table 3-8) and as seen in GCMS metabolite analysis and initial negative MALDI analysis (but contradicted in the second negative analysis), there is limited differentiation based solely upon gender.

Table 3-8 The most significant differences diagnostically capable of distinguishing non-pregnant females from male serum cohorts using MALDI in positive mode with corresponding AUC, LHR- and

LHR+ scores represented in yellow for good and green for excellent with regards to diagnostic potential.

Tentative ID	MZ	p.value	AUC	Sens	1-Spec	LHR+	LHR-
LPC(18:1)	522.5	5.76E-05	0.761	0.633	0.100	6.33	0.41
LPC(O-20:0)	538.5	6.19E-04	0.757	0.700	0.267	2.62	0.41
PC(34:2)	781	4.66E-04	0.76	0.667	0.200	3.34	0.42

### *Non-pregnant vs Pregnant females*

Thirty-nine significantly different lipid markers were found to differentiate non-pregnant females from pregnant females under analysis positive reflectron mode. Two of these were evaluated as better than good using diagnostic capabilities. These lipids were tentatively identified both as ions derived from LPC(16:0) at  $m/z$  518.5 which provided an AUC of 0.96, sensitivity of 0.867 and a LHR+/- of 26.27 and 0.14 respectively ( $P= 9.02E-11$ ) and both of these markers of interest were found to be significantly more prominent in non-pregnant females than in males (Figure 3.23).



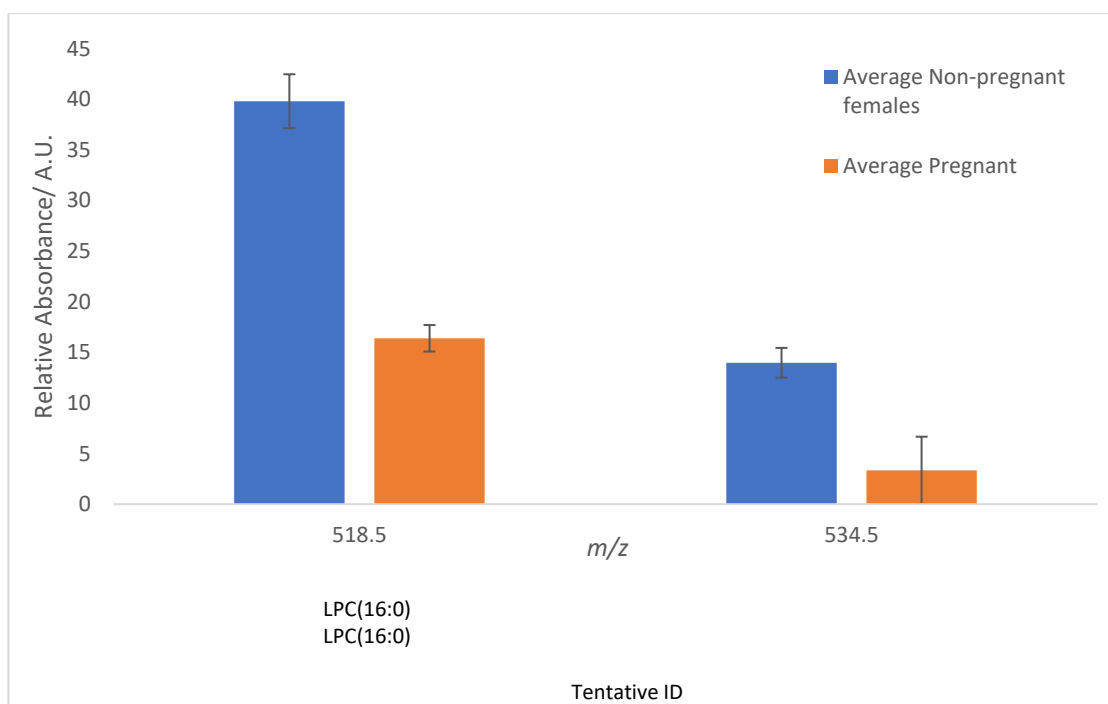


Figure 3-22- Average intensities of the lipids capable of distinguishing non-pregnant females from pregnant females with corresponding standard error bars shown. N=30 for both.

### Pregnant female vs Cord

The comparison of pregnant female and cord serum lipid profiles identified 40 potential distinguishing features, with 19 showing better than good diagnostic capability as described in Table 3-9 and shown in Figure 3-23.

Table 3-9 The most significant differences diagnostically capable of distinguishing pregnant females from cord serum cohorts using MALDI in positive mode with corresponding AUC, LHR- and LHR+ scores represented in yellow for good and green for excellent with regards to diagnostic potential.

MZ	Tentative ID	p. value	AUC	LHR+	LHR-
534.5	LPC(16:0)	1.35E-08	0.886	25.24	0.17
553.5	LPG(20:5)	2.4E-08	0.919	8.67	0.15
726	PS(32:5)	3.98E-07	0.859	5.19	0.16
759	PE(O-38:2)	4.66E-10	0.967	29.30	0.03

760	PC(34:2)	9.06E-11	0.964	29.30	0.03
761	PC(34:1)	6.81E-09	0.936	28.27	0.07
762	PE(O-36:5)	2.17E-09	0.931	28.27	0.07
781	PC(34:2)	7.96E-11	0.986	29.30	0.03
782	PC(34:1)&PC(36:4)	4.69E-11	0.982	29.30	0.03
783	PC (34:2)	1.93E-08	0.922	13.43	0.11
784	x	1.63E-08	0.924	9.33	0.07
785	PC(36:3)	2.86E-08	0.916	7.27	0.04
787	PC(36:2)	8.27E-10	0.944	13.93	0.07
788	PC(36:1)	1.35E-08	0.886	25.24	0.17
797	PC(34:2)	2.22E-10	0.965	13.93	0.07
805	PC(36:2)/	7.18E-05	0.798	27.27	0.10
808	PC(36:2)/	3.5E-09	0.929	9.67	0.04
809	PC(36:2)	1.34E-09	0.953	27.27	0.10
810	PC(38:4)	6.88E-10	0.937	28.27	0.07

The vast majority of the metabolites of interest were found to be significantly more intense in pregnant female serum samples than they were in cord serum with the only exception to this being lipids exhibiting an value of 534.5 and 553.5.

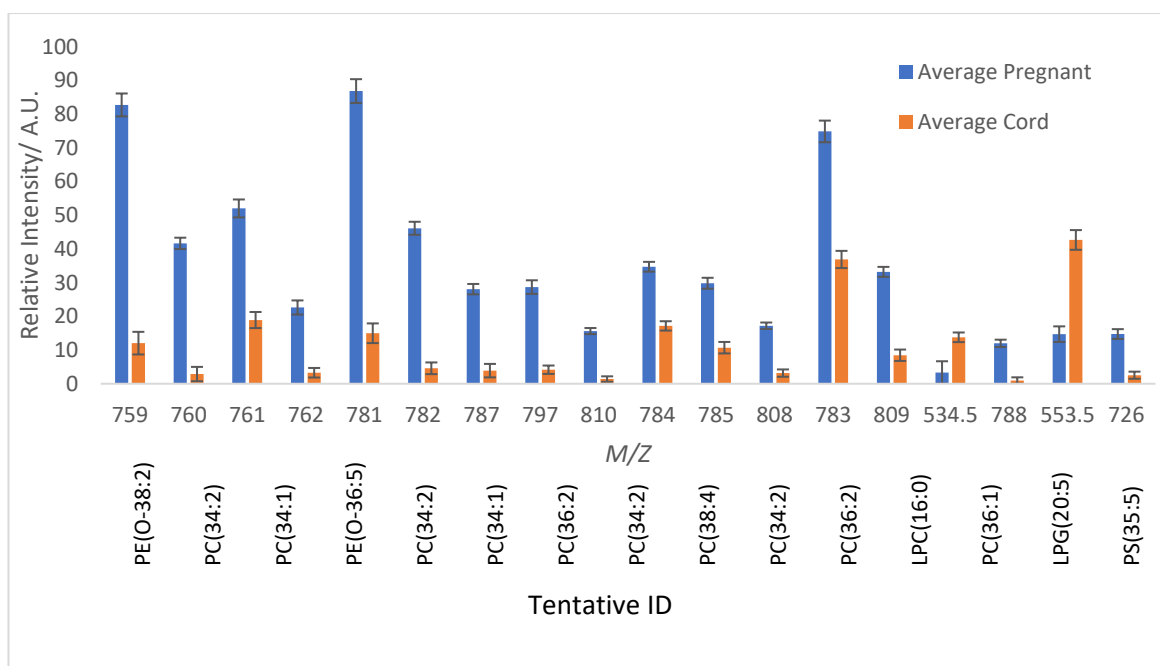


Figure 3-23 Average intensities of the most significantly different  $m/z$  diagnostically capable of distinguishing pregnant females (red) from cord (orange) serum samples with corresponding standard error bars shown (N=30).

There were a number of commonly found markers that were capable of distinguishing multiple different groups from one and other during the analysis whilst maintaining a good diagnostic score in each case. These markers have been amalgamated together to establish their combined capability of group differentiation. The variation of these markers for the group averages is shown in Figure 3-24.

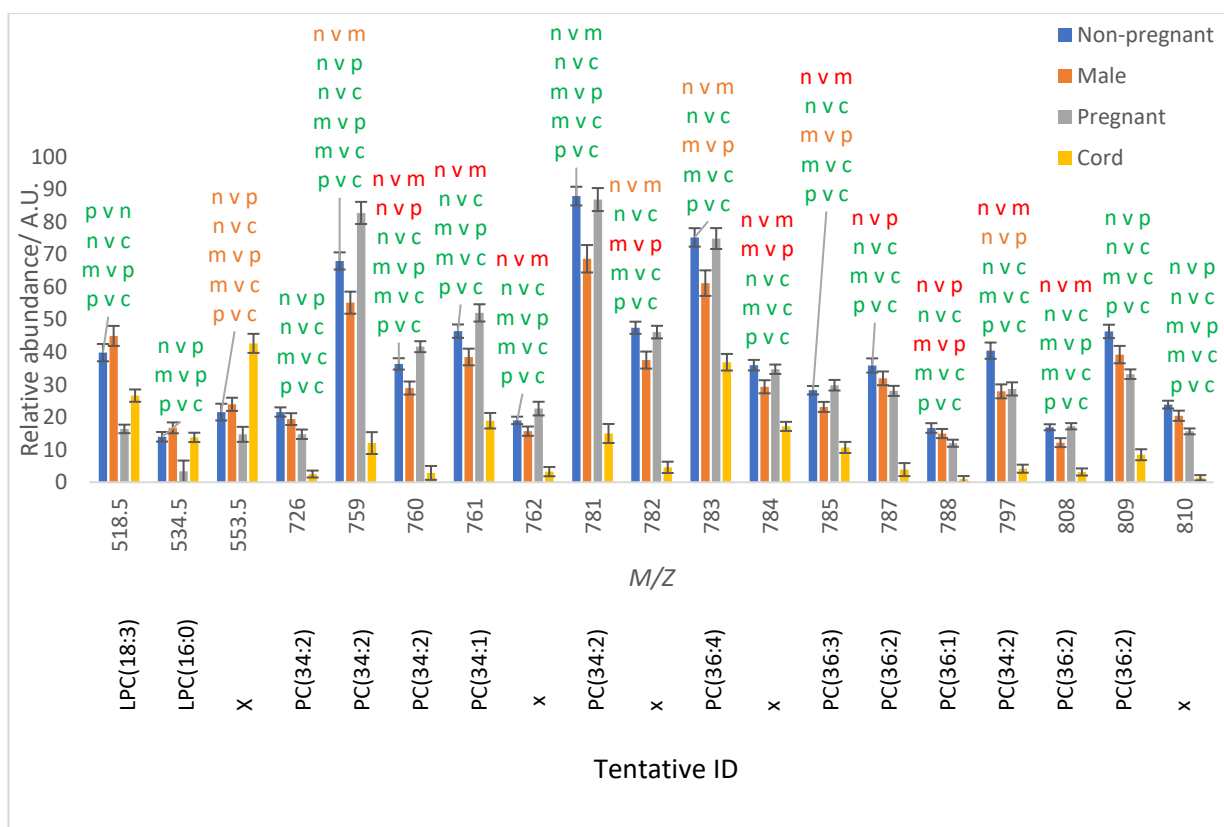


Figure 3-24 Average intensities of the most significantly different  $m/z$  diagnostically capable of distinguishing all of the different groups under analysis with corresponding standard error bars shown (N=30 per group, 120 total). Significance is shown with symbols whereby a comparison has been made between the bar denoted and n being non- pregnant females, m- males, p- pregnant females and c being cord. Red symbol shows a significance of less than 0.5, orange symbolises less than 0.01 and green less than 0.001.

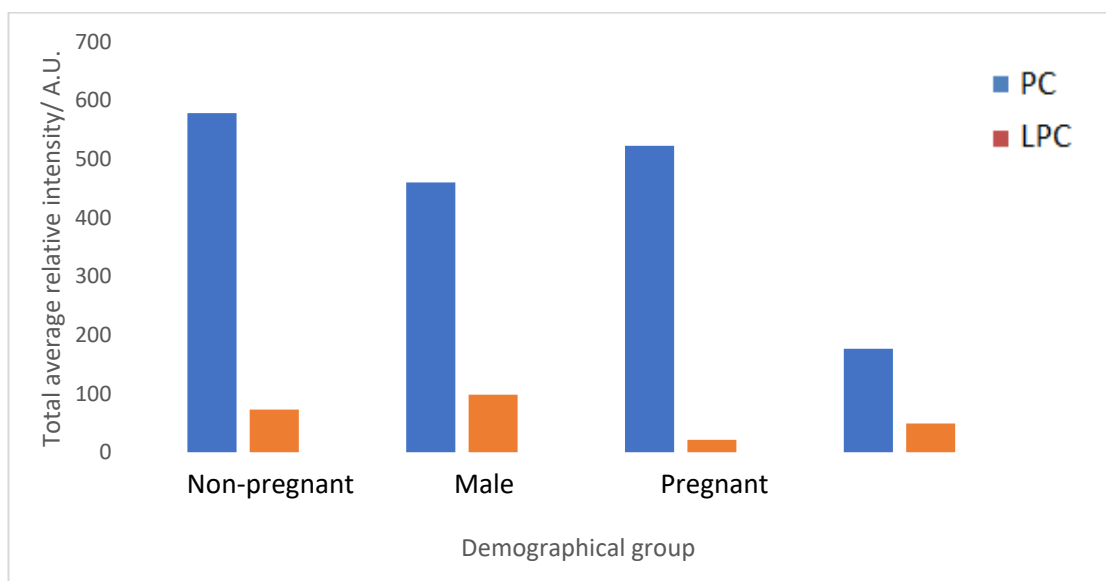


Figure 3-27- Average total presence of phosphatidylcholine (PC) and lysophosphatidylcholine (LPC) in each different demographical group

When considering the ratio of total measured amount of PC compared to the total amount of LPC measured there is a notable threefold elevation, Figure 3-28, in this proportion in pregnant women when compared to all other demographical groups. Not only can this signify that pregnant women's biological status is closest to resembling a state of inflammation, but it can also show the extra demand on key metabolic resource LPC by the foetus during development as well as being critical for the maternal body development for supporting foetus/ baby in later stages (Klatt K.C., et al., 2020). This is further reflected by the cord showing the lowest levels in each instance.

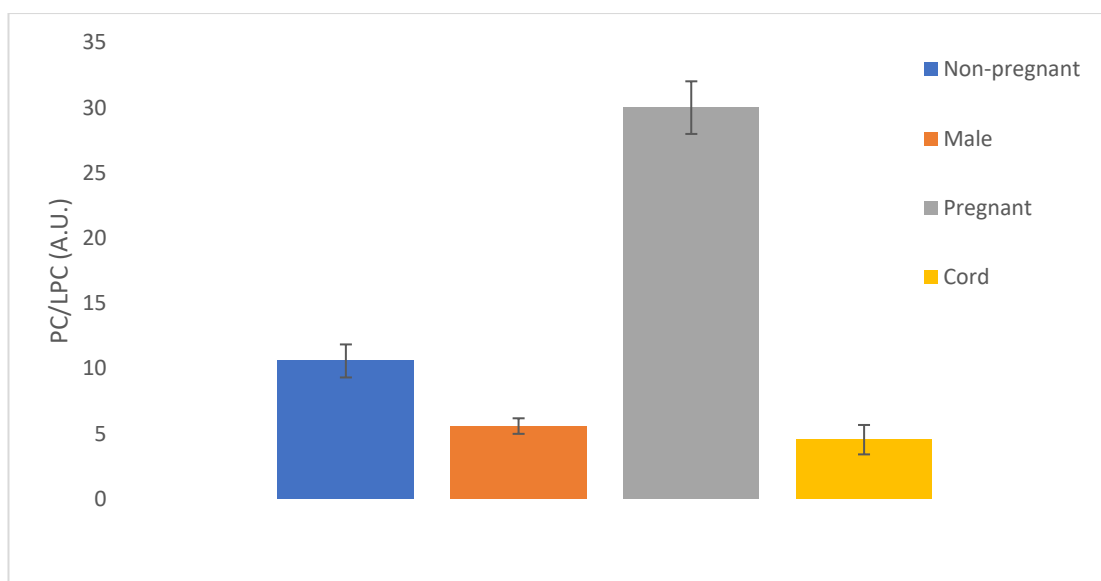


Figure 3-28- Graphical representation of the average variation of phosphatidylcholine (PC)/Lysophosphatidylcholine (LPC) in each demographical group

### 3.3.2- Foetal lipidomic demands Breast-milk

After conducting the prior analysis on demographical group alterations, it was decided that the most discerning differences between pregnant women and cord serum were captured in the lipidomic analysis. It was therefore decided to use this technique when comparing profiles between pregnant female serum, cord serum and breastmilk samples. Additionally, with such a heavy reliance on fat content this area did seem a principal point to consider.

#### 3.3.2.1- Positive mode lipidomic analysis

After positive mode analysis, there were 529 significantly different markers found to distinguish breastmilk collected at two and six months post-partum, with 16 of these scoring better than good discriminatory value according to ROC analysis as shown in Figure 3-29. This data suggests that all significantly different lipidomic markers that differentiate the two timepoints are up-regulated at 6-months when compared to 2.

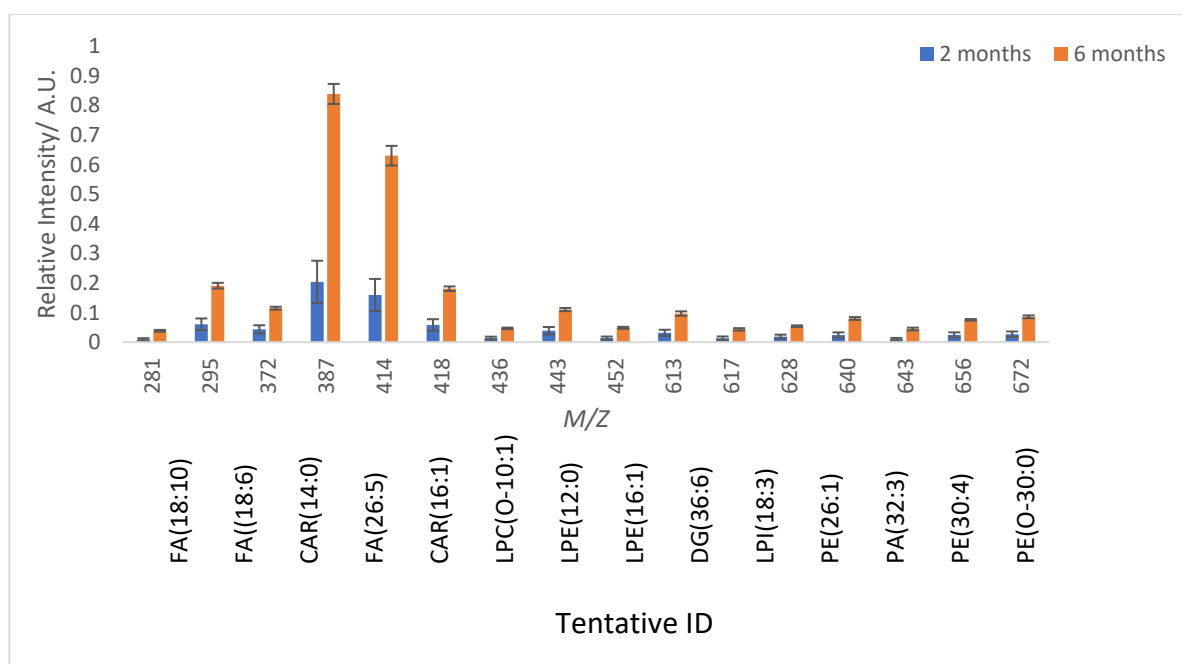


Figure 3-29. Significantly different biomarkers found to differentiate breastmilk collected at 2 and 6 months post-partum analysed in positive mode analysis. All of the markers achieved better than

“good” ROC scoring showing high quality and capability of the metabolites as individual markers. All markers achieved a significance score p. value < 0.001.

#### 3.3.2.2- Negative mode lipidomic analysis

Two hundred and sixty-nine significantly different markers were seen to differentiate 2 and 6 month post-partum breastmilk from each other. However, none provided discriminatory value when analysed using ROC curve analysis.

#### *Comparison between pregnant female and cord serum and breastmilk lipid profiles*

When comparing cord serum and pregnant serum the majority of metabolites had been found to be up-regulated in cord when compared to pregnant samples. Of those that were higher in pregnant women’s serum, possibly signifying a greater usage by the foetus, these were compared to the metabolites that were seen to be significantly altered in the breastmilk samples (between 2 and 6 months).



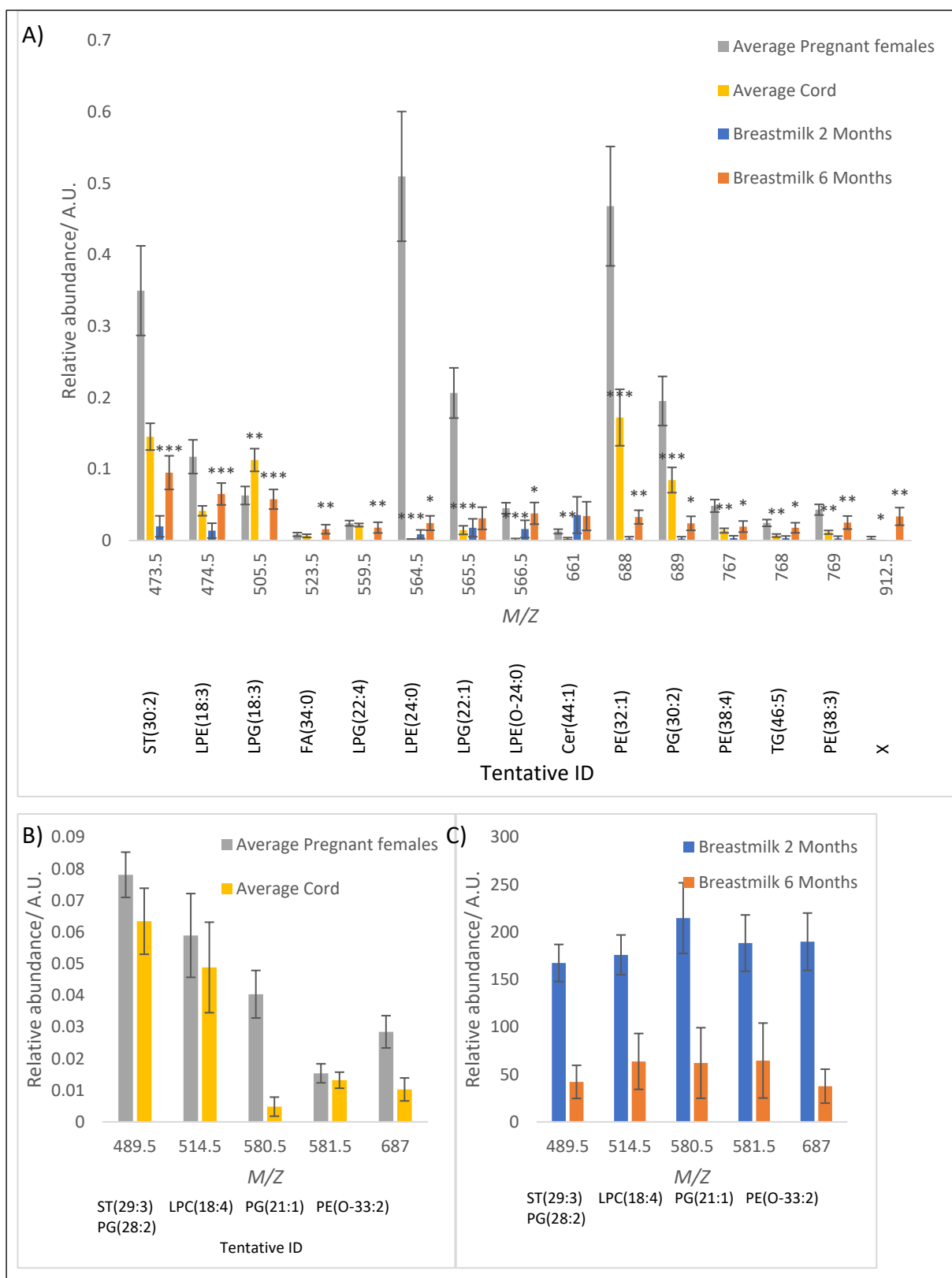


Figure 3-30. Markers found to be significantly up-regulated in pregnant serum samples when compared to cord serum with additional notation of alterations in breastmilk samples between 2 and 6 months. Significant variance between pregnant/ cord and 2 and 6 month breast-milk is

illustrated with asterisks whereby \* $<0.05$ , \*\* $<0.01$  & \*\*\*  $<0.001$ . Tentative ID's are shown where possible, where these could not be generated they have been denoted with an x.

The markers that showed these changes and were additionally consistently seen in the breastmilk sample sets were reported Figure 3-30. In the case of almost these markers, with the exception of  $m/z$  661 where it was mostly the same, the abundance was greater 6 months post-partum than it was at 2 months. These metabolite markers are of particular interest as this gives a good suggestion that they are most required by foetus during development. Consideration was also given to the metabolites most altered between cord and pregnant women's serum found previously, all of which had been found up-regulated in cord serum. These markers are shown in Figure 3-31. Mirroring previous findings almost all of these markers were found to be elevated in later stages on lactation with the exception of 469, 483.5, 497.5, 511.5, 536.5, 645.5 and 664.5.

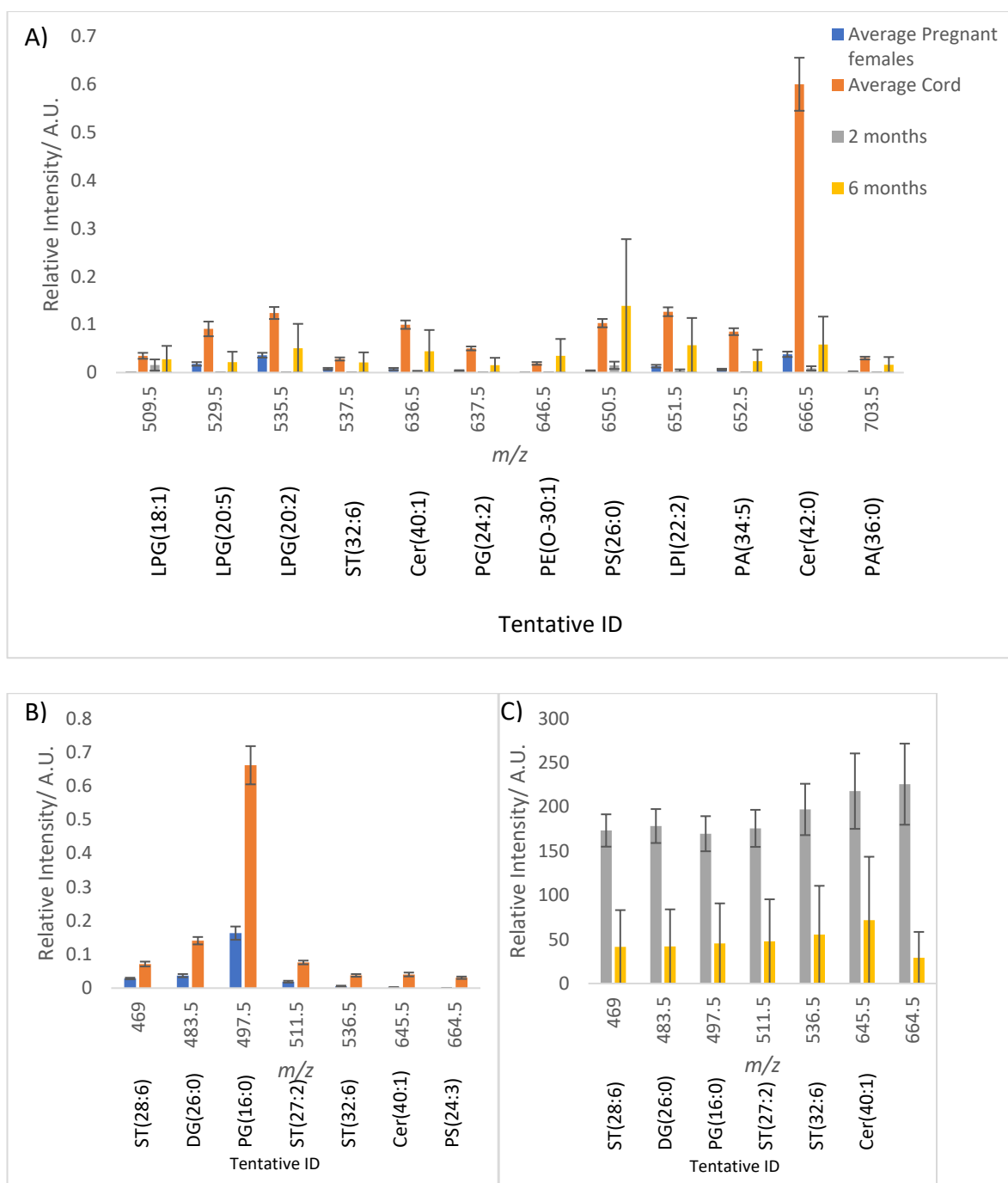


Figure 3-31. Comparison of the most distinguishing biomarkers found to differentiate pregnant and cord serum, to those that were also seen in breastmilk samples comparing 2 and 6 months post-partum. See text for details.

### 3.3.3- Amino/ organic acid biomarker exploration PCOS and Endometriosis

After considering the results from the previous data set, 3.4.1- Comparisons of males, non-pregnant females, pregnant females and cord serum, it was decided that the most suitable analytical technique to employ to seek profile differences capable of distinguishing healthy candidates from PCOS/ endometriosis sufferers would be a targeted GCMS analysis with focus on amino acid presence due to this previous study finding that these were the best quality serum biomarkers that were capable of distinguishing males, pregnant and non-pregnant females from each other and so should provide the greatest capability of finding more intricate profile changes. Related metabolites, such as citric acid and other organic acids were also analysed as potential metabolites by products of these species.

When evaluating the global profile changes of these amino acids and how they are capable of distinguishing the three different groups from each other, PCA plots generate a reasonable separation between all of the groups of interest. This is somewhat improved when only considering the significantly varied amino acids ().

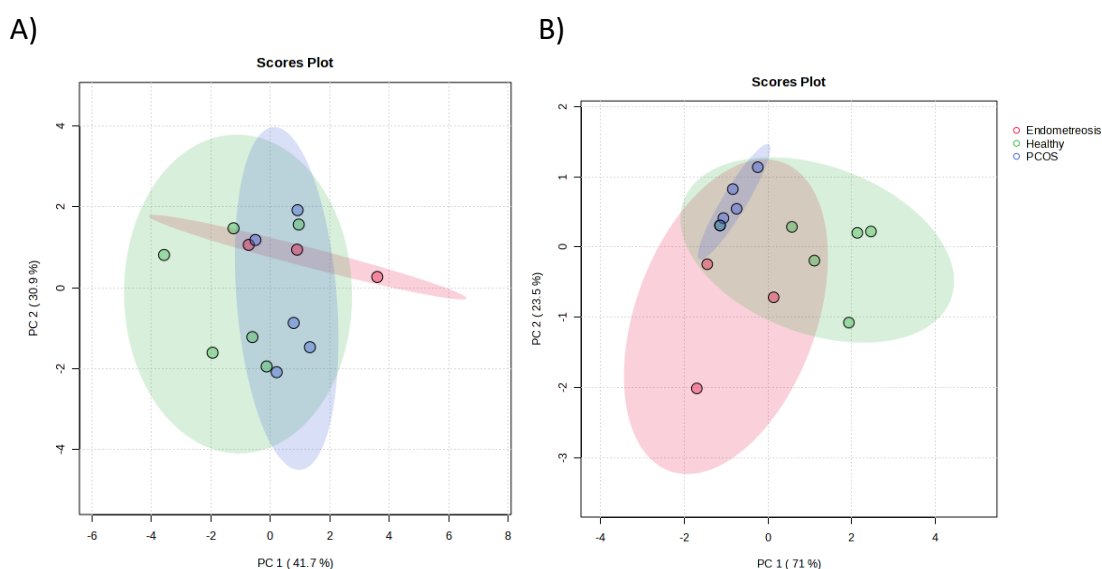


Figure 3-32 PCA analysis comparing the amino acid profiles of healthy, endometriosis and polycystic ovary syndrome (PCOS) candidate's serum. A) Compares the profiles when considering all amino acids under evaluation. There is a division between healthy and PCOS candidates with the former

largely being congregated to the left of the chart. The confidence intervals do overlap (shown by the ovals) B) Considers only the significantly different metabolites. This shows a slightly better differentiation between healthy and endometriosis samples where the healthy candidates are found towards the top right-hand corner of the chart and the endometriosis candidates towards the bottom left. PCOS samples in this instance are congregated towards the top left of the chart, with overlapping confidence intervals.

In order to analyse the relationship between the groups under study heatmaps were generated to depict the concentration fluctuations within and between groups under study. The best differentiation, similarly, to the PCA analysis, was achieved when considering only the statistically significantly different markers. The heatmaps showed the clear upregulation of all of these candidate markers in the diseased states, most prevalently the endometriosis candidate sera. When considering all groups, Figure 3.31, the healthy groups was very clearly identifiable from the other candidate groups with all bar one healthy candidate coming off at the first major branch. The other major branch consisted of all of the PCOS and endometriosis candidates that subdivided over three branches, the most major holding only one endometriosis member and the other subdividing into clearly defined PCOS and remaining endometriosis candidate groups. When considering how the groups averages affect the metabolic profiles and group relationships, , in this instance endometriosis candidates were the most distinct grouping with the most reduced presence of all amino acids under evaluation. The other major branch then subdivided into the clearly highly abundant amino acid bearing healthy control group and the moderately down-regulated PCOS.

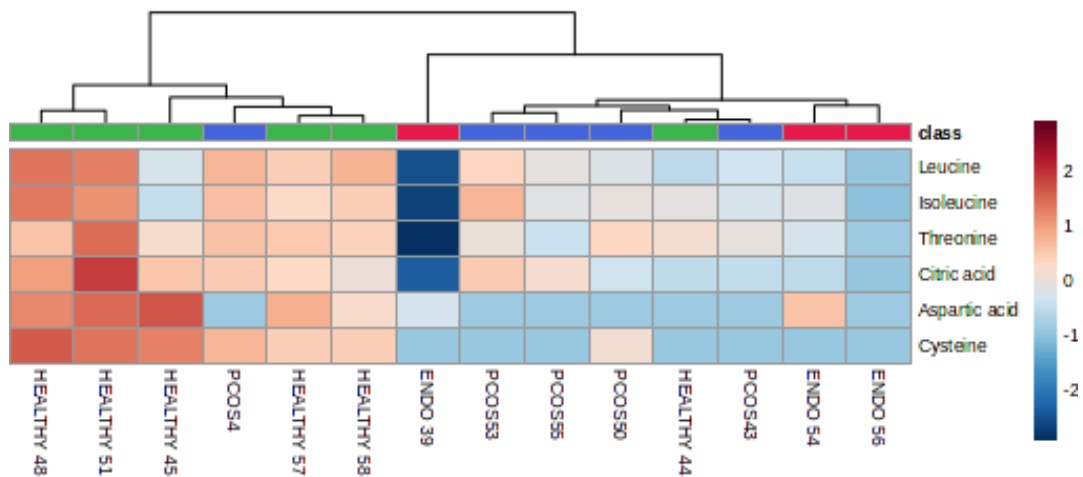


Figure 3-33 heatmaps showing the intensity variation of the significantly different amino acid markers within and between different groups of interest. Shows the individual groups fluctuation with 83% of the healthy group branching off together at the first major arm with one polycystic ovary syndrome (PCOS) candidate group. the other major branch consists of two minor arms, one consisting of one of the endometriosis groups and the other further sub-dividing into one arm consisting of all of the PCOS members and the other the remaining two endometriosis samples.

PCA charts showed a poor level of separation and so in order to generate better group separations PLSDA charts were generated, . In both instances a very good degree of separation was achieved between the healthy and other groups under evaluation with all group members being clearly defined and very minimal confidence interval overlap. This shows that a good level of group definition can be achieved using these biomarker candidates. Another attempt to generate improved group variance was to compare the total relative abundance of all amino acid markers between the three different groups under study. This was conducted as though not all amino acids were found to vary significantly between different groups, all showed the same trend. Figure 3-36 shows the variance of the average total relative intensity for each group under question with a very clear drop in abundance seen in each disease state when compared to healthy control groups. This variance was only found to be significant when comparing endometriosis profiles to healthy. Using this as a biomarker achieved good ROC scoring with an AUC of 0.944, a sensitivity of 83.3 and specificity of 100 being achieved.

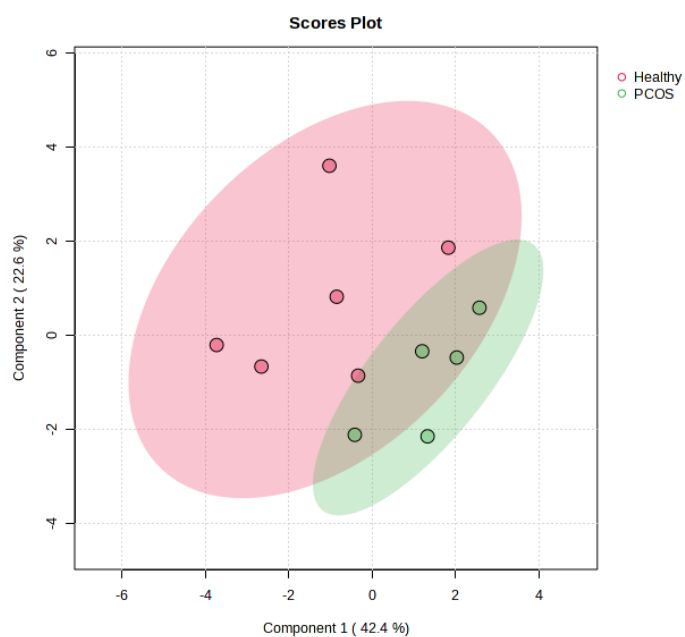
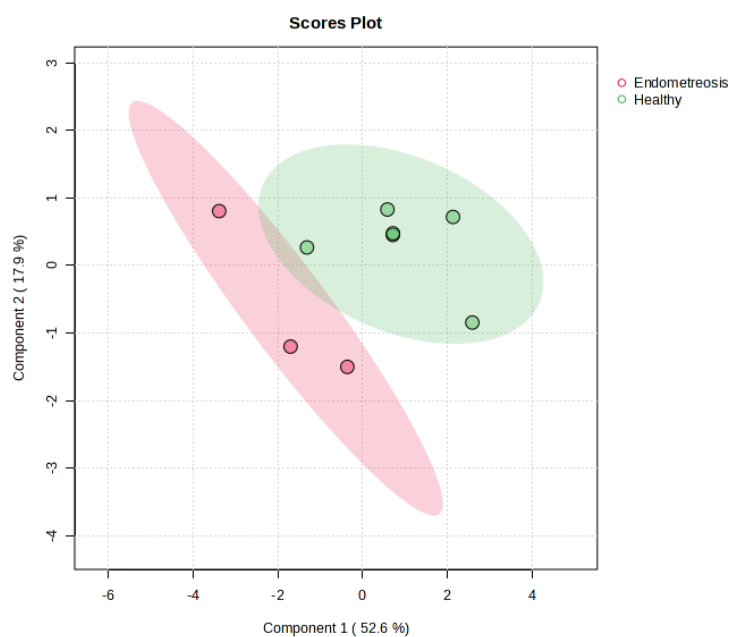


Figure 3-34. PLSDA analysis comparing the significantly different amino acids relationship between healthy and A) endometriosis B) Polycystic ovary syndrome (PCOS) candidate serums. Both charts generate a very good group separation with all group candidates clearly defined from each other and minimal confidence interval overlap depicting good group separation using these markers.

Several significant alterations were seen to differentiate both PCOS and endometriosis candidates from healthy as well as each other when looking at individual amino acid profiles in serum samples. The majority of amino acids showed a decrease in presence between each condition and healthy cohorts with the exception of tryptophan, which instead was upregulated in endometriosis candidates. The trend seen was almost always most obvious between endometriosis sufferers and healthy cohorts than PCOS against healthy. Of the differences seen Leucine, Isoleucine, Threonine, Citric acid and Cysteine were all seen to be significantly different in endometriosis candidates whereas PCOS this was true only for Leucine and Citric acid (Figure 3-35).

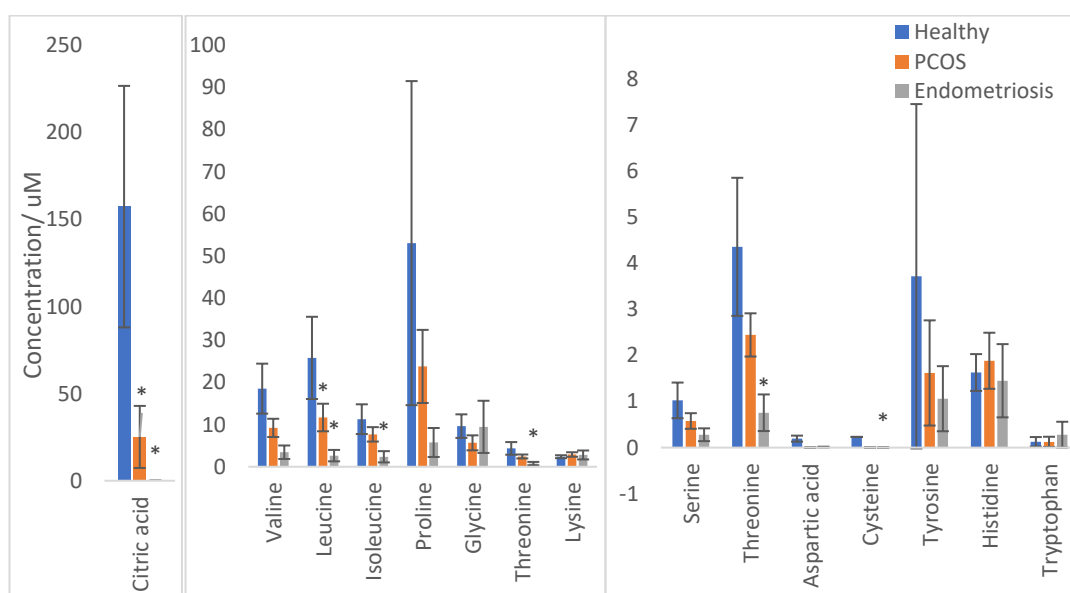


Figure 3-35 Amino acid concentrations observed in serum samples of healthy (blue), polycystic ovary syndrome (PCOS) (red) and endometriosis (green) candidates with statistical significance between healthy and PCOS/endometriosis indicated with asterisks whereby \* $<0.05$ , \*\* $<0.01$ , \*\*\* $<0.001$ .

When considering the diagnostic potential of these amino acid biomarkers, comparisons between healthy and endometriosis candidates showed apt individual biomarker potential with Threonine and Aspartic acid achieving excellent ROC scoring in all elements with an AUC scoring of 1. Leucine, Isoleucine and Cysteine all achieved better than “good” scoring and still scored high AUC values of 0.944, 0.944 and 0.917 respectively. None of the markers identified as significantly varying between PCOS



and the healthy achieved better than good diagnostic quality relating to ROC scoring in any instance when being considered as individual biomarker candidates.

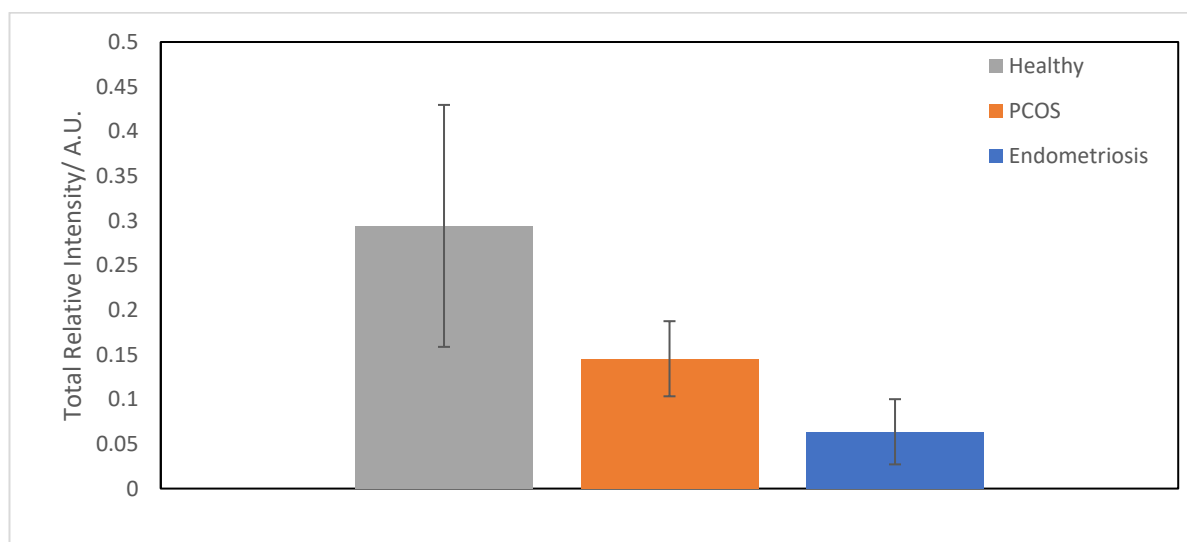


Figure 3-36 Average total relative intensity of all amino acids present in serum of healthy, polycystic ovary syndrome (PCOS) and endometriosis candidates.

### 3.4. Discussion

In the first instance the aim was to identify metabolic markers that were capable of distinguishing different demographical groups; males, non- pregnant females, pregnant females and cord samples from each other. The next aim was to evaluate the effectiveness of each method trialled and better understand the suitability of each. The results of this analysis would then provide the technical basis for the next aim, which sought to identify novel diagnostic biomarkers for PCOS and endometriosis in candidate's serum samples through comparing metabolic profiles. As well as providing a better system of diagnostic this chapter aims to better understand the metabolic changes that occur in these diseases. The final aim then sought metabolic profile changes that are indicative of key metabolites required during neonatal development. By identifying these metabolic changes, researchers can potentially offer improved guidance and care for pre-term infants to ensure that their nutritional and developmental needs are met.

#### 3.4.1- Discussion of serum markers that can distinguish male, pregnant and non-pregnant female and cord candidates.

For the initial analysis seeking profile differences between diverse groups, all metabolic processes and pathways were under scrutiny in the evaluation using global untargeted profiling methods to create the most encompassing picture of the bodies status as possible in order to establish the most suitable analytical process for different forms of more specific later analysis. Three different analytical methods were trialled; GCMS and MALDI-TOF a number of subclasses of metabolites were identified.

##### Fatty acids:-

Fatty acids are a major class of metabolite that showed significant variance between serum sub-groups, which is unsurprising when noting their established role in many biological processes within the body including structurally in cell membranes, hormone production, fat stores, immune responses, cell signalling and energy. The changes were consistently identified when analysing the samples using both of the analytical techniques GCMS and MALDI-TOF.

GCMS saw an overall significant increase in the total levels of fatty acid present in pregnant women, when compared to non-pregnant and cord samples (this wasn't possible in MALDI as relative lipid intensities were used in analysis- effectively normalising the total lipid content extracted). The picture is more complex however when more detailed evaluation of individual fatty acids and even certain groups of fatty acids are scrutinised more closely. The consistent up-regulation of the mothers serum compared to cord, infers immediately that the key supply for foetal development is from the mother for the majority of lipids under analysis.

Cholesterol is a key steroid lipid identified for bearing significant variance using GCMS (Figure 3-2) (Figure 3-24 and Table 3-1). Cholesterol is the only lipid noted in the study that was found to be significantly upregulated in the serum of pregnant women when

compared to any other group under evaluation. Cholesterol is known to become upregulated during pregnancy largely due to its additionally involvement in sex steroid hormone levels and aiding to maintain the pregnancy (Bartels, Ä., & O'Donoghue, K. 2011). There are prerequisites for a foetus to have a source of cholesterol for similar functionality to normal mammalian requirement, through its role in cell membranes, hormonal and use as a cell signalling moiety (Palinski W., 2009). A significant amount of the cholesterol reserved by the foetus will be generated past the placenta by the foetus itself and some by the placenta. There is also heavy debate regarding the potential for some cholesterol to be transported from the mother to the foetus via the serum itself, which could also attribute to the difference seen in maternal levels when being compared to non-pregnant females (Woollett, L. A., & Heubi, J. E. 2020).

Mis-regulation of cholesterol levels in the foetus have been known to have detrimental effects, some even catastrophic for foetal survival. There is a lack of information regarding healthy cholesterol levels in pregnant women besides the acknowledgement of them becoming upregulated. There have been associations with higher levels of cholesterol in the mother and later development of atherosclerosis in the carried child in later life of recent (Leiva, A., et al., 2013. And Woollett, L. A., & Heubi, J. E. 2020). A better understanding of foetal sources of cholesterol and what constitutes healthy maternal circulating levels of cholesterol could hold a target for identification of healthy foetal development.

The majority of the lipids seen during GCMS analysis follow a trend of showing the highest abundance in non-pregnant females, followed by males, pregnant females and then being significantly lower in cord serum samples. This is true of hexadecanoic acid, Octadecanoic acid (Stearic acid), Oleic acid 9,12-octadecanoic acid (Linoleic acid) and Tetradecanoic acid (Myristic acid) as can be seen in Figure 3-2 The trends seen in these fatty acids generally follow previously published literature (Crocker, I., et al., 1999; Meher, A., et al., 2016). Alterations in the lipid levels in serum can have a multitude of severe consequences for both the mother and foetus including those stemming from gestational diabetes mellitus (Zhu, Y., et al., 2018), this includes a number of these lipids which were found to show significant variance during

pregnancy. The lower abundance noted in arterial cord serum illustrates a heavier usage and uptake of these markers by the foetus.

Cord serum exceeded pregnant women's serum for the lipids Heptadecanoic acid, Heptanedioic acid (a marker that previous studies maternally has a strong association with gestational diabetes mellitus), Butanoic acid (this particular lipid exceeds all other groups under study), Propanoic acid and Octadecanoic acid. Whilst the vein umbilical cord supplies the foetus with nutritional requirements from the mother, as previously discussed this will then typically illustrate a lower level of the used substrates in the arterial cord than in the mothers serum with the addition of waste products to be removed. As such the up-regulation of these metabolites may infer a product of the foetus or placenta either generated in excess or as a metabolic waste (Meher, A., et al., 2016; Zhu, Y., et al., 2018). Short chain fatty acids are prominent in gut microbiota and therefore could suggest a microbiological component accountable for the changes noted, (Ren S.J., et al., 2024). These observations would be better observed in serum samples where mother and cord serums were matched to observe specifically how these vary. Though this wasn't possible for this study a future study looking at this would be highly beneficial to better understand these trends.

MALDI analysis provided improved detail on lipidomic changes that arose between the different cohorts under analysis. In a similar fashion to the lipids marked as significantly different during the GCMS evaluation, the vast majority of lipids were seen to be significantly lower in cord serum than compared to the other groups under evaluation. The main differences to this trend occurred in the significantly different LPC lipids, including those reported at 534.5 and 553.5, Figure 3-27 . These all saw the cord serum supersede the pregnant serum samples by significant amounts, and the lipid analysed at 553.5 showed significantly higher levels than all of the other groups. Cord serum shows lower levels, though not significantly, than non-pregnant women in 518.5 and 534.5, Figure 3-24 and whereas the changes are shown to be significant between non-pregnant and pregnant in these lipids, Figure 3-22. Conversely to the trend seen in LPC, the PC levels are all significantly lower in the aortic cord serum than all other groups serum, again signalling towards a heavy foetal

utilisation of these lipids in particular. Non-pregnant and pregnant women's serum were significantly different in all of the PC lipids bar- 761,762,781,783,784,785 and 808, Figure 3-24. Of those that are significantly different between pregnant and non-pregnant, pregnant serum tends to be down-regulated in the presence of all PC lipids evaluated.

A possible reason for the trends in PC and LPC lipids distinguishing foetal serum when compared to the maternal has been suggested by Ferchaud-Roucher, V., et al., 2019. During the study the same trends were observed, with LPC being lower in Maternal than foetal and additionally showed a much higher level present in the placenta than either maternal or foetal. In addition, it was found that PC was higher in maternal than cord and also much higher again in the placenta. In the placenta PC is converted into LPC using Land's pathway via PLA2 enzymes where the enzyme is highly expressed as both an intrinsic and secretory enzyme. LPC is then actively transported to foetal circulation by major facilitator superfamily domain containing 2A (MFSD2a) so it can be utilised for foetal brain development and weight accretion. This pathway is also particularly important for PUFA accumulation products. Additional explanations may lie in LPC's involvement in inflammation through its activation of multiple signalling pathways that are involved in oxidative stress and inflammatory responses, these can be triggered during pregnancy. Further possibilities for PC alterations when considered in conjunction with the previous observation regarding foetal usage could be due to choline being a heavily required lipid during pregnancy due to it having use in the synthesis of PC, sphingomyelin and acetylcholine, DNA/RNA methylation as well as further uses again. It is actively transported across the placenta and as such is seen in higher levels in the cord serum than the maternal system in this study. This is specifically the PC derived from the PUFA pathway containing arachidonic acid (20:4n-6) during the third trimester.

PCA analysis showed that overall MALDI depicted cord serum as having the most similarity to male serum under the analysis used, especially when considering the most biologically capable biomarkers found to distinguish all of the groups from each other. In almost each circumstance both cord and males were found to be significantly higher in the lipid levels than the other groups under study.

## Proteins and Amino acids-

FTIR analysis saw a global upregulation of proteins in the cord serum when compared to all other groups under between  $3000\text{-}3600\text{cm}^{-1}$ , with an additional noted lower level in pregnant women, . This level was not however significantly lower than males. This indicates that the foetus actively transports or synthesis's its own proteins to meet requirements. This trend was similarly reflected when looking at the total levels of amino acids present using GCMS analysis, however this time the pregnant females overall amino acid presence was significantly less than all other cohorts tested, Figure 3-37

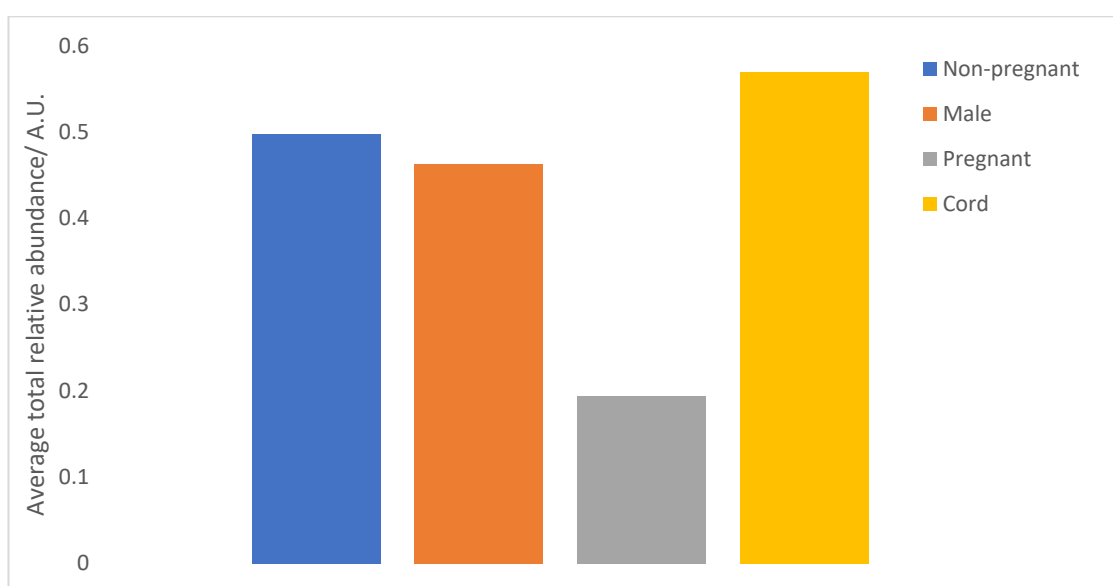


Figure 3-37- Average total intensity of all amino acids measured using GCMS analysis in the four different study groups serum; non-pregnant (blue), male (red), pregnant (green) and cord (purple).

A number of amino acids have been highlighted as bearing significant difference in levels between the groups as indicated Figure 3-9. The most obvious difference in amino acids is the alterations in pregnant women's serum concentrations, which bare significantly lower levels of all the amino acids highlighted. This is consistent with literature, which have noted similar patterns amongst groups of amino acids (pregnancy term dependent) due to the higher demands of the foetus and as such the essential amino acids are carried through the placenta to the foetus via active transport (Kalhan SC et al., 2003 and Di Giulio AM et al., 2004). The significance of this elevation in foetus amino acid levels when compared to maternal is the source/

mode of transfer across the placenta to the foetus to sustain requirements. If it is a mode of transfer, it must be active.

The significant drop in amino acid concentrations between non-pregnant females and pregnant females may illustrate a requirement to better monitor pregnant women's supplementation and fully understand the biological requirements to assure that not only the mother but the foetal needs are being fully met and also indicates a good system for differentiating different fertility related demands metabolically. These can also be indicative of certain fertility markers that change through-out female cycling/ pregnancy.

Research by Evans R. W., et al., 2003 agreed with all of the findings for alterations between amino acid levels in cord and pregnant women's serum; and additionally, found associations between lowered availability of a number of these amino acids and pre-eclampsia and determinantal effects on infant development.

When analysing the differences in serum amino acid levels between males and females there were several amino acids shown to display significant differences. Of these Valine, Glycine, Proline, Aspartic acid, Glutamine, L-Ornithine, L-Lysine, L-Tyrosine and L-Tryptophan were all shown to be significantly greater in males than in pregnant females. However, when comparing males serum to non-pregnant females the number of significantly different amino acids dropped and showed different trends. Valine and proline were the only amino acids found to be significantly higher in males whereas serine threonine and malonic acid were found to be significantly higher in non-pregnant females.

### Myo-inositol

Myo-inositol is a carbohydrate that is essential for numerous metabolic pathways and biological functions and is considered non-essential as it can be intrinsically synthesised from glucose. Myo-inositol is a precursor for various inositol phosphate signalling molecules which are fundamental for nutrient sensing and energy homeostatic control. Its metabolic role is multifaceted, with biological functions

include growth factor signalling and regulation, hormone secretion, neurotransmitter signal transduction, insulin signal transduction, cytoskeleton assembly, lipid metabolism and signalling and membrane formation antioxidant pathways and cellular protection. In this manner it is critical to reproduction and fertility as well as foetal and child growth and development (Wang, x., ET AL., 2020; Tu-Sekine, B., & Kim, S. F. 2022). With the variety of essential roles of myo-inositol it was unsurprisingly found to show one of the greatest group to group variations including one of the largest up-regulations in foetal cord serum when compared to maternal (with over a 4x increase shown, Figure 3-8). Free sugar-alcohol moiety is formed from glucose in the serum. The data in this study shows an up-regulation exceeding all other groups in cord serum and a lower level than all other groups in pregnant women. This observation infers that a large amount is generated by the foetus and/or the placenta. One possible reasoning for this finding it that studies have found that inositol levels are highest in slow dividing tissues in adult life such as skeletal and cardiac muscles. After birth as gestational age increases the levels of myo-inositol drop in the neonates serum to a stable point (Brion L., et al., 2021; (Brusati, V., et al., 2005).

#### 3.4.2. Exploring metabolite requirements of foetus and neonates.

This section aimed to conclusively characterise key metabolic changes that were notable in pregnancy and vary in cord serum that may divulge important lipidomic factors required during foetal development and indicate those that may be produced by the placenta/ baby. A further aim was to note how these requirements are maintained postnatally via breastmilk and observe how these change at different stages after birth. Evaluate how these are indicative of requirements of pre-term infants.

This study attempted to analyse breast milk lipidomic profile changes post-partum and draw comparisons to alterations in pregnant women and cord serums lipidomic profiles in order to establish a focus on preterm infants' lipidomic requirements. Such



analysis can offer up invaluable insights into potential gaps or imbalances in nutrients in order to target interventions in that could help to optimize the nutritional support required by pre-term infants and promote optimal growth and development. This can help to pave the way for personalised nutritional strategies that holds the potential to enhance the overall health and wellbeing of all neonates.

This study began by considering the differences between pregnant and cord sera lipidomic profiles seeking out the most significant and notable changes which would illustrate factors most utilised by the foetus and some that are generated by the foetus as waste products or by the placenta and provided to the foetus rather than via the mothers serum. During negative mode MALDI lipidomic analysis, the majority of markers that were found to vary most significantly were found to be up-regulated in cord serum when compared to maternal serum. This observed trend is most likely to be indicative of factors that are being supplied via the mothers serum to the foetus and are actively taken up and utilised by the foetus. For the markers that were found to significantly vary in this way and which were subsequently also found to be present in breastmilk samples almost all were found to be significantly up-regulated at 6 months postpartum and 2 months, Figure 3-31a. This evaluation suggests that these markers are more heavily utilised by the foetus as their requirement increases as they develop. A large number of these lipids were found to be sterols and cholesterol, which are known to be critical lipids for embryo, foetal and neonatal development through their heavy involvement in growth as well as use in steroid hormone production. Pre-term a large amount of the cholesterol and sterols are supplied to the foetus via maternal blood supplies, as was suggested in this study. There is some generation of cholesterol by the foetus itself in addition. Postnatally without the serum source there is a high demand on food sources to maintain the cholesterol uptake required during this period of intense growth and development, and this is typically largely sourced through the mother's breastmilk. Cholesterol hold a range of uses within infants including use neurologically, bile acid and lipoprotein synthesis and hormones. Without correct sustenance of sterol based factors numerous health factors are attributed such as heart disease, metabolic diseases (Woollett LA & Shah AS., 2023; Kelishadi, R., et al., 2012). Another large proportion of these varies lipids

is attributed to the presence of glycerophospholipids. These lipids are utilised in membranes and are found heavily concentrated in brain and muscular structures.

There were additionally a number of lipid species that were found to be significantly up-regulated in pregnant women when compared to cord serum samples. These lipids can show a heavy utilisation in the foetus after supply by the mother. Of those found to be up-regulated in the mother and found to be maintained in the breastmilk supply to foetus all again were found to be upregulated with an increased foetal demand at 6 months post-partum when compared to 2 months, Figure 3-30. This time the majority of the lipids consisted of phosphatidylethanolamines and their lyso counter-parts. Phosphatidylethanolamine (PE) is the second most abundant lipid present in mammalian cells with heavy utilisation in membrane structures as well as brain and muscular developments.

During the negative mode analysis, comparisons were made between the fatty acids that were found to be significantly higher in maternal serum when compared to cord serum, which was taken to be indicative of high usage by the foetus. Of these differences ST(29:3), LPC(18:4), PG(21:1), PE(O-33:2) and PG(28:2), Figure 3-31C were found to be higher at 2 months than 6 months. Furthermore, a marked increase was observed in almost all lipids between two and six months lactation. This is likely due to the growing nutritional and energy requirements of a larger growing infant.

Overall this study found a number of key lipidomic changes that occurred during the developmental stages of foetus and neonates that highlighted a number of key fatty acid markers that should be further considered as critical to maintain adequate neonatal support.

### 3.4.3.- Seeking novel fertility biomarkers for endometriosis and PCOS diagnosis.

A comprehensive analysis of the amino acid profiles in a cohort of PCOS and endometriosis patients with age-matched healthy controls was also conducted within this current study. Serum samples were collected and subjected to targeted GCMS amino acid profiling techniques to quantify the concentrations of individual amino acids. The analysis revealed a number of significantly altered key amino acids that are individually capable of being purposed as diagnostic biomarkers for both conditions presented.

Polycystic ovary syndrome (PCOS) is a complex endocrine disorder that affects reproductive-aged individuals, characterized by hormonal imbalances, and metabolic disturbances. Endometriosis is a chronic condition, characterised by the inflammation and abnormal growth of tissue outside the uterine cavity. These alterations to normal tissue growth can result in the presentation of novel metabolic biomarkers due to cell processing alterations. One of the prominent discoveries in this research was the notable reduction in the concentration of amino acids in the serum of individuals with various diseases, with the most distinct impact observed in cases of endometriosis. Upon statistical assessment of the variance in the amino acid profiles between each condition and those of healthy individuals, it became evident that specific amino acids were significantly altered, Figure 3-35. In the context of endometriosis, the amino acids Leucine, Isoleucine, Threonine, Citric acid, and Cysteine showed significant deviations, while in the case of PCOS, the alterations were primarily seen in Leucine and Citric acid. The summed global amino acid profiles intensity was also considered and again was found to be decreased in diseased candidates serum, significantly so in endometriosis sufferers.

There is increased evidence that oxidative stress through the presence of reactive oxygen species (ROS) is a contributing factor to the pathogenesis behind PCOS. PCOS has associations with low-grade systemic inflammation through its associated elevation of multiple markers including C-reactive proteins, interleukin-18, monocyte chemoattractant protein-1 and white blood cell count. Whilst there are known

associations of an increase in oxidative stress in women with PCOS and acknowledgment of a decreased total antioxidant status through reduced glutathione and haptoglobin, no further markers or evaluations have been made to date nor utilisation of these processes as potential diagnostic markers (Duleba, A. J., & Dokras, A. 2012; Rudnicka, E., et al., 2020). In addition to the possible onset triggers caused by ROS presence, there is also an increased risk of these ROS elevations enhancing associated long-term health complications associated with PCOS including diabetes, ovarian tissue damage, hypertension and endometrial cancers as well as further metabolic disorders and DNA damage (Cheng, X., & He, B. 2022; Duleba, A. J., & Dokras, A. 2012; Surapaneni, K. M., & Venkataramana, G. 2007).

A large number of the amino acids studied in the presented data have known antioxidant and therefore anti-inflammatory properties, helping to remove or suppress the actions of oxidative radicals (Xu, N., et al., 2017). In addition to directly contributing the antioxidant potential, some of the varied amino acids also hold the capacity to upregulate the activity of numerous antioxidant enzymes including glutathione-S-transferase and catalase, further contributing to the issue, which have also been shown in studies to be significantly decreased in PCOS patients (Unni, S. N., et al., 2015; Surapaneni, K. M., & Venkataramana, G., 2007).

Endometriosis is also characterised by inflammation and again, studies have indicated markers of inflammation and lowered presence of antioxidant enzymes including GSH, superoxide dismutase and catalase (Ortiz, C. N., et al., 2021). A similar study mirrored a large number of the trends noted in this study and additionally noted decreases in proline bearing associations with increased ROS. This is due to prolines degradation via dehydrogenase being associated with ROS generation (Dutta, M., et al., 2018) Additional potential explanations for the decreased amino acid presence reported in this study include an increase in metabolic demand due to alterations to cellular proliferation increasing. The result of such would be a higher usage of energy and cellular resources including amino acids.

In conclusion, our study provides novel insights into the amino acid alterations associated with both PCOS and endometriosis, highlighting new potential roles and

implications in inflammation and ROS presence. Further research is warranted to elucidate the exact underlying molecular mechanisms linking amino acid alterations and antioxidant activity in diseased states. These findings hold promise for the identification of amino acids as a new diagnostic target for both conditions.

### 3.5. Conclusion

Overall, this chapter has provided details of key metabolic changes using multiple mass spectrometry based tools and provided disease characterisations as well as a better understanding of metabolism changes under different bodily states. Such markers can act as both a diagnostic biomarker and provide detail regarding key profile changes that occur at developmental time points.

The primary objective was to best profile metabolic markers that could distinctly differentiate between various physiological states, such as males, non-pregnant females, pregnant females, and cord samples.

Through different forms of analysis, and methodological evaluation tailored to each scenario, a better understanding of the advantages and disadvantages of a range of analytical platforms was generated for the subsequent pursuit of novel diagnostic biomarkers for conditions like PCOS and endometriosis. This latter study then provided key amino acid biomarkers that not only enhance our understanding of these conditions but also provided potential new avenues for early diagnosis and treatments.

The chapter's final aim extended its focus to the critical phase of neonatal development. By identifying shifts in lipidomic metabolite profiles, we have highlighted potential key metabolites crucial for neonatal growth and development, the study contributes invaluable insights into guiding care strategies, particularly for pre-term infants.

### 3.6. References

- Angelini, R., Vortmeier, G., Corcelli, A., & Fuchs, B. (2014). A fast method for the determination of the PC/LPC ratio in intact serum by MALDI-TOF MS: an easy-to-follow lipid biomarker of inflammation. *Chemistry and physics of lipids*, 183, 169–175. <https://doi.org/10.1016/j.chemphyslip.2014.07.001>
- Cheng, X., & He, B. (2022). Clinical and Biochemical Potential of Antioxidants in Treating Polycystic Ovary Syndrome. *International journal of women's health*, 14, 467–479. <https://doi.org/10.2147/IJWH.S345853>
- Christ, J. P., & Cedars, M. I. (2023). Current Guidelines for Diagnosing PCOS. *Diagnostics* (Basel, Switzerland), 13(6), 1113. <https://doi.org/10.3390/diagnostics13061113>
- Clower, L., Fleshman, T., Geldenhuys, W. J., & Santanam, N. (2022). Targeting Oxidative Stress Involved in Endometriosis and Its Pain. *Biomolecules*, 12(8), 1055. <https://doi.org/10.3390/biom12081055>
- Di Giulio AM, Carelli S, Castoldi RE, Gorio A, Taricco E, and Cetin I. (2004) Plasma amino acid concentrations throughout normal pregnancy and early stages of intrauterine growth restricted pregnancy. *J Matern Fetal Neonatal Med.*;15(6):356-62.
- Duleba, A. J., & Dokras, A. (2012). Is PCOS an inflammatory process?. *Fertility and sterility*, 97(1), 7–12. <https://doi.org/10.1016/j.fertnstert.2011.11.023>
- Dutta, M., Joshi, M., Srivastava, S., Lodh, I., Chakravarty, B., & Chaudhury, K. (2012). A metabonomics approach as a means for identification of potential biomarkers for early diagnosis of endometriosis. *Molecular bioSystems*, 8(12), 3281–3287. <https://doi.org/10.1039/c2mb25353d>
- Dutta, M., Singh, B., Joshi, M., Das, D., Subramani, E., Maan, M., Jana, S. K., Sharma, U., Das, S., Dasgupta, S., Ray, C. D., Chakravarty, B., & Chaudhury, K. (2018). Metabolomics reveals perturbations in endometrium and serum of minimal and mild

endometriosis. Scientific reports, 8(1), 6466. <https://doi.org/10.1038/s41598-018-23954-7>

Elmi, F., Movaghar, A. F., Elmi, M. M., Alinezhad, H., & Nikbakhsh, N. (2017). Application of FT-IR spectroscopy on breast cancer serum analysis. *Spectrochimica acta. Part A, Molecular and biomolecular spectroscopy*, 187, 87–91.

Ferchaud-Roucher, V., Kramer, A., Silva, E., Pantham, P., Weintraub, S. T., Jansson, T., & Powell, T. L. (2019). A potential role for lysophosphatidylcholine in the delivery of long chain polyunsaturated fatty acids to the fetal circulation. *Biochimica et biophysica acta. Molecular and cell biology of lipids*, 1864(3), 394–402. <https://doi.org/10.1016/j.bbalip.2018.12.007>

Fuchs, B., Schiller, J., Wagner, U., Häntzschel, H., & Arnold, K. (2005). The phosphatidylcholine/lysophosphatidylcholine ratio in human plasma is an indicator of the severity of rheumatoid arthritis: investigations by <sup>31</sup>P NMR and MALDI-TOF MS. *Clinical biochemistry*, 38(10), 925–933. <https://doi.org/10.1016/j.clinbiochem.2005.06.006>

Ganeshalingam, M., Enstad, S., Sen, S., Cheema, S., Esposito, F., & Thomas, R. (2022). Role of lipidomics in assessing the functional lipid composition in breast milk. *Frontiers in nutrition*, 9, 899401. <https://doi.org/10.3389/fnut.2022.899401>

Kalhan SC, Gruca LL, Parimi PS, O'Brien A, Dierker L, and Burkett E. (2003) Serine metabolism in human pregnancy. *Am J Physiol Endocrinol Metab.*; 284(4):E733-40

Katagiri R, Goto A, Budhathoki S, Yamaji T, Yamamoto H, Kato Y, Iwasaki M, and Tsugane S. (2018) Association between plasma concentrations of branched-chain amino acids and adipokines in Japanese adults without diabetes. *Sci Rep.* ;8(1):1043.

Kim, KJ., Kim, HJ., Park, HG. et al. A MALDI-MS-based quantitative analytical method for endogenous

estrone in human breast cancer cells. *Sci Rep* 6, 24489 (2016). <https://doi.org/10.1038/srep24489>

Lee, J. H., Kim, Y. H., Kim, K. H., Cho, J. Y., Woo, S. M., Yoo, B. C., & Kim, S. C. (2018). Profiling of Serum Metabolites Using MALDI-TOF and Triple-TOF Mass Spectrometry to Develop a Screen for Ovarian Cancer. *Cancer research and treatment*, 50(3), 883–893. <https://doi.org/10.4143/crt.2017.275>

NHS (2023) [Your breastfeeding questions answered - NHS \(www.nhs.uk\)](https://www.nhs.uk)

NHSb (2023) [Polycystic ovary syndrome - Diagnosis - NHS \(www.nhs.uk\)](https://www.nhs.uk)

Ortiz, C. N., Torres-Reverón, A., & Appleyard, C. B. (2021). Metabolomics in endometriosis: challenges and perspectives for future studies. *Reproduction & fertility*, 2(2), R35–R50. <https://doi.org/10.1530/RAF-20-0047>

Parasar, P., Ozcan, P., & Terry, K. L. (2017). Endometriosis: Epidemiology, Diagnosis and Clinical Management. *Current obstetrics and gynecology reports*, 6(1), 34–41. <https://doi.org/10.1007/s13669-017-0187-1>

Ramiro-Cortijo, D., Singh, P., Liu, Y., Medina-Morales, E., Yakah, W., Freedman, S. D., & Martin, C. R. (2020). Breast Milk Lipids and Fatty Acids in Regulating Neonatal Intestinal Development and Protecting against Intestinal Injury. *Nutrients*, 12(2), 534. <https://doi.org/10.3390/nu12020534>

Ren, S. J., Feng, J. T., Xiang, T., Liao, C. L., Zhou, Y. P., & Xuan, R. R. (2024). Expression and clinical significance of short-chain fatty acids in patients with intrahepatic cholestasis of pregnancy. *World journal of hepatology*, 16(4), 601–611. <https://doi.org/10.4254/wjh.v16.i4.601>

Rudnicka, E., Kunicki, M., Suchta, K., Machura, P., Grymowicz, M., & Smolarczyk, R. (2020). Inflammatory Markers in Women with Polycystic Ovary Syndrome. *BioMed research international*, 2020, 4092470. <https://doi.org/10.1155/2020/4092470>

Sadeghi, H. M., Adeli, I., Calina, D., Docea, A. O., Mousavi, T., Daniali, M., Nikfar, S., Tsatsakis, A., & Abdollahi, M. (2022). Polycystic Ovary Syndrome: A Comprehensive Review of Pathogenesis, Management, and Drug Repurposing. *International journal of molecular sciences*, 23(2), 583. <https://doi.org/10.3390/ijms23020583>



Schiller J., Zschornig O, Petkovic' M, Muller M, Arnhold J, and Arnold K (2001) Lipid analysis of human HDL and LDL by MALDI-TOF mass spectrometry and <sup>31</sup>P-NMR. *Journal of lipid research.*; 42(9)1501-1508

Schmidt. G. and Steinhart, H., 2002. Impact of extraction solvents on steroid contents determined in beef. *Food Chemistry.* 76(1) 83-88

Shimizu F, Ishii Y, Ogawa M, Takao T, Matsuoka K, Kato K and Takada A (2017) Plasma levels of various amino acids and their changes upon protein uptakes in Japanese young and old men and women. *Integr Food Nutr Metab*; 4(6)1-5

Surapaneni, K. M., & Venkataramana, G. (2007). Status of lipid peroxidation, glutathione, ascorbic acid, vitamin E and antioxidant enzymes in patients with osteoarthritis. *Indian journal of medical sciences*, 61(1), 9–14.

Tu-Sekine, B., & Kim, S. F. (2022). The Inositol Phosphate System-A Coordinator of Metabolic Adaptability. *International journal of molecular sciences*, 23(12), 6747. <https://doi.org/10.3390/ijms23126747>

Unni, S. N., Lakshman, L. R., Vaidyanathan, K., Subhakumari, K. N., & Menon, N. L. (2015). Alterations in the levels of plasma amino acids in polycystic ovary syndrome--A pilot study. *The Indian journal of medical research*, 142(5), 549–554. <https://doi.org/10.4103/0971-5916.171281>

van den Akker CH, Schierbeek H, Minderman G, Vermes A, Schoonderwaldt EM, Duvekot JJ, Steegers EA, and van Goudoever JB. (2011) Amino acid metabolism in the human fetus at term: leucine, valine, and methionine kinetics. *Pediatr Res.*;70(6):566-71. doi: 10.1203/PDR.0b013e31823214d1.

Wang, X., Yue, H., Li, S., Guo, J., Guan, Z., Qin, J., Zhu, Z., Niu, B., Cui, M., & Wang, J. (2020). The Effects of Inositol Metabolism in Pregnant Women on Offspring in the North and South of China. *Medical science monitor : international medical journal of experimental and clinical research*, 26, e921088. <https://doi.org/10.12659/MSM.921088>

WHO (2023) [Breastfeeding \(who.int\)](https://www.who.int/breastfeeding)

Woollett LA, Shah AS. Fetal and Neonatal Sterol Metabolism. [Updated 2023 Mar 25]. In: Feingold KR, Anawalt B, Blackman MR, et al., editors. Endotext [Internet]. South Dartmouth (MA): MDText.com, Inc.; 2000-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK395580/>

Wu G (2010) Functional amino acids in growth, reproduction, and health. *Adv Nutr.*;1(1):31-7. doi: 10.3945/an.110.1008. Epub 2010 Nov 16.

Wu G. (2013) Functional amino acids in nutrition and health. *Amino Acids.*;45(3):407-11. doi: 10.1007/s00726-013-1500-6. Epub 2013 Apr 18.

Xu, N., Chen, G., & Liu, H. (2017). Antioxidative Categorization of Twenty Amino Acids Based on Experimental Evaluation. *Molecules (Basel, Switzerland)*, 22(12), 2066. <https://doi.org/10.3390/molecules22122066>

Yan, L., Yi, J., Huang, C., Zhang, J., Fu, S., Li, Z., Lyu, Q., Xu, Y., Wang, K., Yang, H., Ma, Q., Cui, X., Qiao, L., Sun, W., & Liao, P. (2021). Rapid Detection of COVID-19 Using MALDI-TOF-Based Serum Peptidome Profiling. *Analytical chemistry*, 93(11), 4782–4787. <https://doi.org/10.1021/acs.analchem.0c04590>

## Chapter 4 – Elephant faecal metabolite analysis for the determination of reproductive status

### 4.1.1- Abstract

Asian elephants face significant breeding challenges in captivity, with the underlying causes of infertility largely unknown. This study employs the global metabolomic analysis of faecal samples to investigate potential biomarkers associated with infertility. Sixty faecal samples were collected from two elephants, divided equally across three groups: cycling, non-cycling, and pregnant. The collected samples were methanol extracted and derivitised before being analysed using GC-MS. The study revealed a notable upregulation of a number of different markers, a number of which being associated with markers of stress, in the faeces of non-cycling elephants when compared to the other groups under study. These findings suggest that stress may play a role in the observed fertility issues. Expanding the study to include a larger population of elephants and additional sample time points could provide deeper insights into the causative factors behind infertility in captive Asian elephants.

### 4.1.2- Introduction

Free-roaming elephants reproduce with great efficiency, even leading to overpopulation of elephants in certain areas, despite overall endangerment levels being true (Allen, W.R., 2006). In captivity however, despite holding such a large proportion of the global number of elephants, there is great difficulty in breeding, noted with reported low birth rates and high calf mortality rates. The importance of having a self-sustaining captive breeding programme is more prevalent than ever in order to reduce numbers of elephants needed to be brought in from the wild. In order to achieve such a programme an improved understanding of captive elephant's reproductive developments and complications is needed (Rees, P.A., 2003).

In animal studies, stool sampling is a popular source choice for metabolomic analysis principally for its ease of collection, being both non-invasive and additionally for its less time-critical collection and with minimal storage needs. Both of these qualities enable sampling to be done with minimal training requirements and protects the subjects from stress that could not only impact upon the resulting metabolic profile but further contribute to the condition under study. Despite faeces not offering as much detail as blood samples, nor a snap shot of the current internal status of the body, it can offer useful insights into the overall downstream profiles that can signify upstream variations in metabolism. Metabolomic analysis of elephant dung may allow for the identification of any metabolic changes, including potential hormonal alterations, leading up to and during pregnancy that may be accountable for a different breeding success rates among captive elephants.

In addition to this, there will be a vast reflection of the gut microbiome displayed in the analysis as well, which would be missed during a blood analysis. Whilst analysis of a microbiome rich source will occlude some internal metabolic contributions, it will provide an important factor that can impact the hosts health and wellbeing through influences on the integrity of the mucosal barrier, pathogen protection, immune functioning, and hormone regulation with proven associations between bacterial relative abundances and hormone concentrations associated with fertility markers, though the latter is not yet well established (Klinhom, S., et al., 2023; Keady, M. M., et al., 2021). The microbiota of Asian elephants reaches maturity between 10 and 14 years of age and remains stable in healthy circumstances until reaching advanced ages and as such fluctuations in such can have associations with numerous ailments including obesity, poor diet, gastrointestinal complications, hormonal alterations and stress (Keady, M. M., et al., 2021).

Currently, steroid level patterns, including circulating progesterone, prolactin or conjugated estrogens are commonly used to track animal pregnancy, but these often involve numerous invasive blood samples and difficult collection being taken (Kajaysri, J., and Nokkaew, W., 2014; Niemuller, C.A., et al 1998). Progress has been made in developing less invasive faecal sample testing strategy including the tracking of faecal progestogen concentrations that show a general increase towards 12

months and then a decrease to its lowest concentration at 22 months; when taken into consideration with stress hormone glucocorticoid levels it can give an indication of imminent calving. However, this pattern could be seen in other scenarios and unless tracked over the duration of the pregnancy (both costly and time consuming) the strategy will fail to predict as efficiently (Kajaysri, J., and Nokkaew, W., 2014).

Enzyme immunoassays have already successfully highlighted key hormonal alterations during the estrous period of female elephants, with some holding up to 10 times elevated glucocorticoid levels due to captivity stresses which can affect reproductive cycling and lack of social interactions with fellow elephants due to separation via chains/ cages (Kumar, V., et al 2014) The 2014 study was also able to link faecal androgen and glucocorticoid metabolite profiles with male elephants Musth, period of elevated sexual activity in free roaming Asian elephants (Kumar, V., et al 2014). A separate study has also provided evidence of a pheromone presence in female elephant dung, that enabled male elephants to establish the reproductive phase of females, however, no identification of the actual pheromones was undertaken, nor comparison between wild and captive elephants made to show the occurrence in both being critical (Ghosal R., et al 2021)

Success has been found using such methodology when looking at other species of animals their fluctuations in faecal hormones and associated metabolites. Wild black rhinoceros were found to have faecal  $20\alpha$ -progestagens levels up to five times elevated during pregnancy when compared to cycling rhinoceros detected by enzyme immunoassay (Garnier, J. N., et al 1998). However, when a study was conducted tracking corticosterone levels in white rhinoceros, limited success was found distinguishing pregnant and cycling animals from each other (Metrione, L. C., & Harder, J. D., 2011). Instead the fluctuations were noted between captive and wild rhinoceros showing the importance of animal specific testing strategy. Wasser S.K., et al., (1991) developed a rapid method of monitoring pregnancy status in wild yellow baboons utilising radioimmunoassays to monitor the significant fluctuations in oestradiol and progesterone in faecal samples able to detect pregnancy efficiently after 25 days from conception, whilst captive armadillos have also been successfully screened for pregnancies using enzyme immunoassays to monitor the ten fold

increase in progestogens extracted and detected in faeces (Howell-Stephens, J., et al 2013). It is important to note that different species of animals have been found to have altered metabolisms of steroid hormones, altering the profiles seen in steroid by products in the faecal matter. In general steroid hormones are metabolised by the liver before being excreted via urine or bile which is then excreted as faeces (Schwarzenberger F., 2007) and as such a specific method is needed for specific reflection of captive elephants.

#### 4.1.3- Aims

The aim of this present study is to utilise a less targeted mass spectrometry analysis of faeces in order to highlight and identify potential markers of fertility issue in the elephants and to further understand the altered metabolite level in relation to reproductive cycles with consideration to faecal microbiome. Similar global faecal metabolomic studies conducted on white rhinoceros in both the wild and captivity captured metabolic profile changes indicative of fertility issues such as lowered phytoestrogen levels and general lower overall metabolite presence and gut microbiome profiles (Williams, C. L., et al, 2019). However, this is the first study to apply this approach to elephant reproductive status in a non-invasive manner.

1. Optimise a suitable extractions method for capturing metabolic profiles from elephant faeces
2. Compare the profiles of pregnant, cycling and non-cycling elephants faecal samples to note any alterations
3. Assess if any of these profile changes could be attributed to fertility issues in non-cycling elephants.

## 4.2- Methods and Materials

### 4.2.1 Samples

Faecal samples were collected from three different reproductive stages of female elephants; cycling (collected outside of pregnancy term), non-cycling (collected from elephants unable to conceive for unknown reasons) and pregnant (collected during term). There were no significant differences in food provisions noted besides increased volume around the time of collection. The samples were collected at Chester Zoo and were immediately stored at -80°C until later processing. A total of sixty samples were collected from two elephants; Maya and Sundara. Twenty of these were from Maya, a non-cycling female Asian elephant, 53 years old at the time of study. Samples were collected every 2 days between 1<sup>st</sup> July 2017 and the 6<sup>th</sup> August 2017. The remaining forty samples were collected from Sundara, a cycling female Asian elephant, 15 years old at the time of study. The cycling samples were collected in two batches of 10 (with 2 days between each sample) with batches taken 13 weeks apart. Though the exact point of the cycle is unknown, the sampling time gap enables the capture of two very different time points of the elephants cycle. This is essential to making the metabolic variations discovered as universal as possible and not specific to the cycling stage, hence all of the cycling samples were used as a global cycling group and then further analysed as two distinct cycling groups against each other and against the other groups so that any noted metabolic changes between non-cycling/ pregnant groups and cycling were not also attributed to the stage of the cycle. The final set of samples, termed “pregnant” were collected from Sundara’s 2016 pregnancy during the mid-late stage, at approximately 10 and 14 months. The samples were collected in batches of 10 samples at two different time points.

### 4.2.2 Metabolite extraction optimisation

Trial extractions were first undertaken and broadly based upon previous studies utilising eight different extraction solvents with different strengths to establish which

holds the best global metabolite extraction profile. For the optimisation of the metabolite extraction protocol a pooled faecal sample, consisting of equal parts of one of each different sample group in the study, was used and multiple different metabolite extraction methods applied and compared to one and other based upon their consistency, metabolite numbers and metabolite class range before a full analysis was undertaken using the most effective protocol. Protocols were adapted from ELISA methodology currently utilised by Chester zoo as well as Hadinger U., et al., 2015 and Palme R., 2005 whereby the best recovery of steroid related hormones was achieved through simplified extraction procedures, whereby wet elephant faecal samples were pooled and divided into eight separate 100mg aliquots. The aliquots were then extracted using 500µl of a number of different solvents including; acetone, methanol, ACN, DCM, chloroform: methanol, DCM:methanol (1:1), Ethyl acetate and water. The samples were vortexed for 30 minutes in the respective solvent before being centrifuged at 300rpm for five minutes. The supernatant was collected and transferred into fresh reaction tubes before being re-centrifuged until all solid matter had been successfully removed. For the water extraction an additional 10kDa filtration step was applied, using a Milipore Microcon-10kDa centrifugal filter units being span at 4000g for 15 minutes. 400 µl of the resultant supernatant was then transferred to a fresh tube before being speed vacuum concentrated to dryness and then derivitised and analysed via GCMS.

The three most effective methods from this pre-trial were then re-analysed in further detail whereby three different, separate randomised samples were used for the solvent trial extraction. Triplicate 100mg aliquots of each sample were prepared in three separate reaction tubes. 500µL of methanol (MeOH), ACN (Acetonitrile) and Deionised water (MILI-Q) was added to one of the three sample tubes. The samples were then vortexed for thirty minutes before being centrifuged at 300rpm for five minutes. The supernatant was collected and transferred into fresh reaction tubes before being re-centrifuged until all solid matter had been successfully removed. The deionised water extracted samples were then re-centrifuged through a 10 kDa filter unit. 400µL of the prepared supernatant was then speed vacuum concentrated to dryness before being derivitised and analysed by GCMS as outlined below.



#### 4.2.3 Metabolite analysis by animal group

The sixty study samples were then analysed by the optimal methanol extraction protocol. For metabolite extraction 100mg of dried, thawed elephant faeces was vortexed for 30 minutes in 500µL of methanol before centrifugation at 300rpm for 5 minutes. The supernatant was then collected and re-spun to ensure all solid material was successfully removed before being speed vacuum concentrated to dryness. The samples were then derivitised and analysed by GCMS as outlined in 2.3.1.2.1 GCMS extraction and derivatisation.

#### 4.2.4 Metabolite data processing

A number of data processing steps were required to analyse the large datasets produced during the untargeted analysis. All data generated through GCMS analysis was first converted from .raw files into .csv format using Openchrom software. This enabled AMDIS to then be used to deconvolute and process the spectra generating data reports of identified peaks utilising NIST08 and corresponding elution (.elu) files noting the peak details (retention times, intensity and area details and corresponding identification). In order to correct for any peak drift in the GC system seen between different samples, the free online system Spectconnect was then used to process the generated elu. files of all of the data sets in order to produce data matrices of aligned RT and RI and mass data using the following parameters: Elution threshold: medium (1 min), Support threshold: low ( $\geq 50\%$  samples), Similarity threshold: low (70% match). These settings also help to reduce the data-sets by removing any inconsistent features within a group. The samples were treated as 5 different groups for this process, consisting of Maya non-cycling (n=20), Sundara cycling June (n=10), Sundara cycling March (n=10), Sundara pregnant 10 month (n=10) and Sundara pregnant 14 month (n=10), this was done to ensure that any subtle inter-group alterations would not be excluded from the analysis. The generated matrices were used to manually confirm correct alignment using the RT matrix, which contains the original retention times matched to a new corresponding retention time. This matrix was also used for

checking for duplicate data entries and any errors made during alignment that were required to be amended before analysis. The RI matrix was then used for data analysis after first being normalised according to the internal standard, tetracosane. Biomarker evaluation was carried out as described previously in Chapter 2 and in the introduction section on biomarker evaluation methods.

## 4.3- Results

### 4.3.1 Extraction methods and their evaluation

Prior to the analysis of the faecal samples for comparison between elephants reproductive status an initial analysis was undertaken in order to determine the most appropriate extraction method to utilise for these samples. The mass spectrometric ion chromatogram from the initial optimisation analysis is shown in Figure 4-1. The different solvents were selected for trial due to their different chemical properties, having an impact upon the ability that they have to interact with different solutes of interest and therefore the types of metabolites they will be most suited to extraction. The polarity index being a good indicator of this quality, with a higher value indicative of a higher polarity; DCM < Chloroform < Ethyl acetate < Methanol = Acetone < ACN < Water for the solvents used in the present study. Another key consideration for this analysis is the selectivity of the solvents, as this study seeks to analyse in a non-targeted manner, a less selective solvent such as water or methanol was suspected to likely to be preferable for achieving the greatest metabolic coverage (Mushtaq, M. Y., et al, 2014).

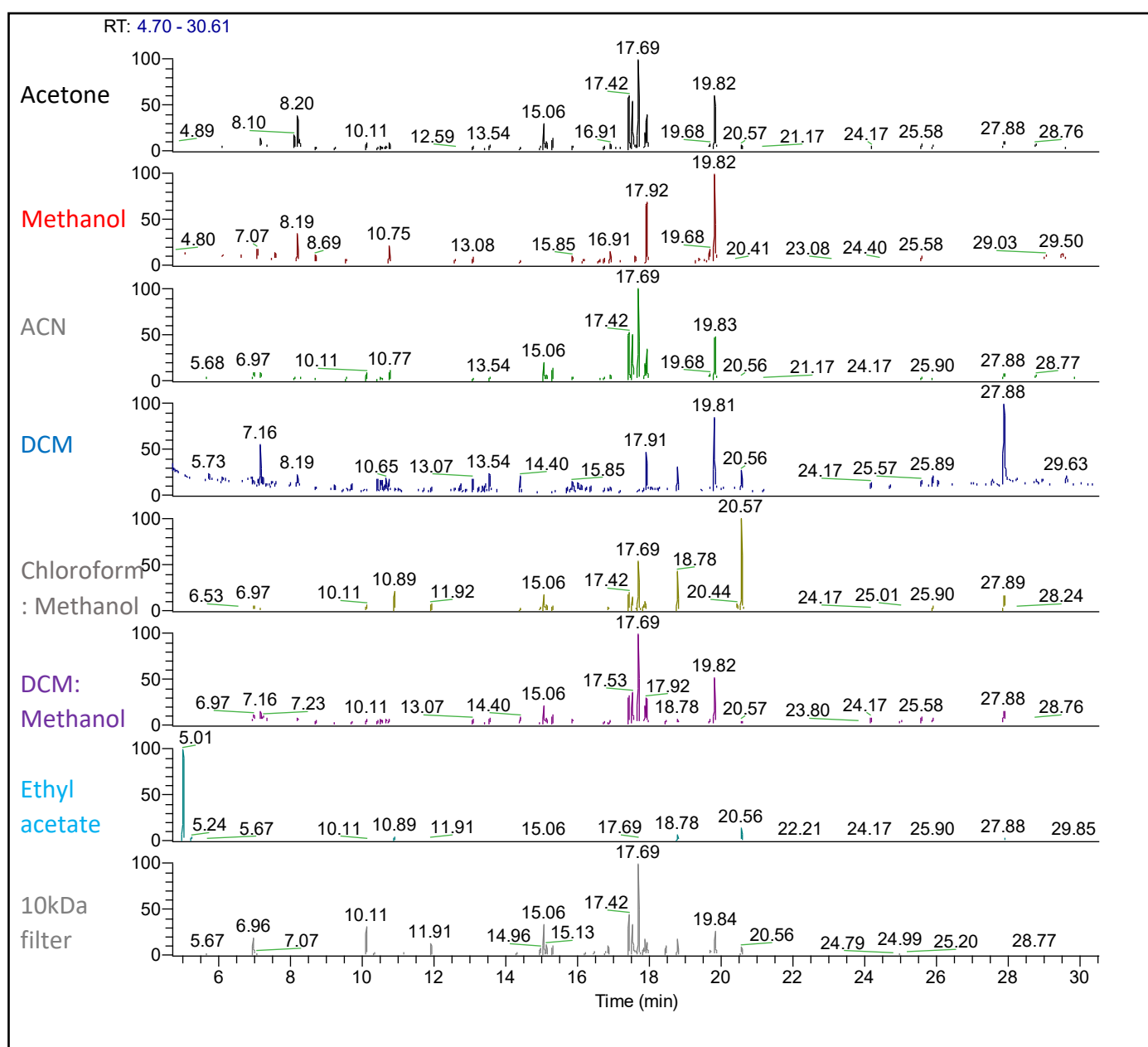


Figure 4-1- Comparison of raw GCMS spectra from the same pooled elephant sample extracted using acetone, methanol, acetonitrile, dichloromethane, chloroform: methanol, dichloromethane: methanol, Ethyl acetate and water with a 10kDa filter respectively.

Methanol is a popular choice as it is able to extract a wide range of metabolites including sugars, amino acids, organic acids, alkaloids and phenolic compounds and has been used successfully in other related studies before (Cheng, K., et al, 2020; Cesbron, N., et al, 2017; Lopez-Bascon, M. A., et al, 2019). This can be seen as the methanol extraction of faecal metabolites produced the most consistently prominent peaks compared to the spectra. A common cluster of metabolites associated with all the extraction methods was visible in all of the different traces including a set of

peaks between 7 and 8 minutes, a cluster at 15 and two larger clusters at 17 and 19 minutes. These are present in almost all of the scans except Ethyl acetate which lacks the majority and is clearly the least capable solvent for this analysis having the fewest number of peaks present. Acetone, ACN and the aqueous 10 kDa filter extraction method share the most similarities amongst the scans, with ACN and water having more peaks in the lower retention time region than acetone. Comparing the traces, it can clearly be seen that the largest number of unique metabolites generated with the most consistently intense peaks were extracted using methanol, water/ 10kDa filtration and ACN based extractions and as such these were looked at in more detail for a trial extraction to determine the best method for the main sample cohort.

A further comparison of these three extraction methods was conducted in more detail to decide the best solvent for the extraction of the faecal metabolites based not only on metabolite range and intensity but also the reproducibility. During this evaluation, the methanol extraction was clearly the best methodology. A more detailed depiction of the number of metabolites that were extracted by each extraction method is illustrated in Figure 4-2. In total, methanol extraction pulled out 177 different identifiable metabolites, 137 were extracted using aqueous filtering and only 121 using ACN. It can also be seen that there are only 80 common metabolites extracted using each of the three different methods and when considering which extraction method best allows for capture of the greatest number of unique metabolites, methanol proves to be the more optimal solvent achieving 60 unique metabolites, compared to filter and ACN scoring 34 and 9 respectively. Further to this after normalisation to the internal standard tetracosane, the peak intensities achieved using the methanol extraction are also generally greater for each distinguished peak than the other extraction methods, hence suggesting a more efficient extraction (See

Figure 4-3).

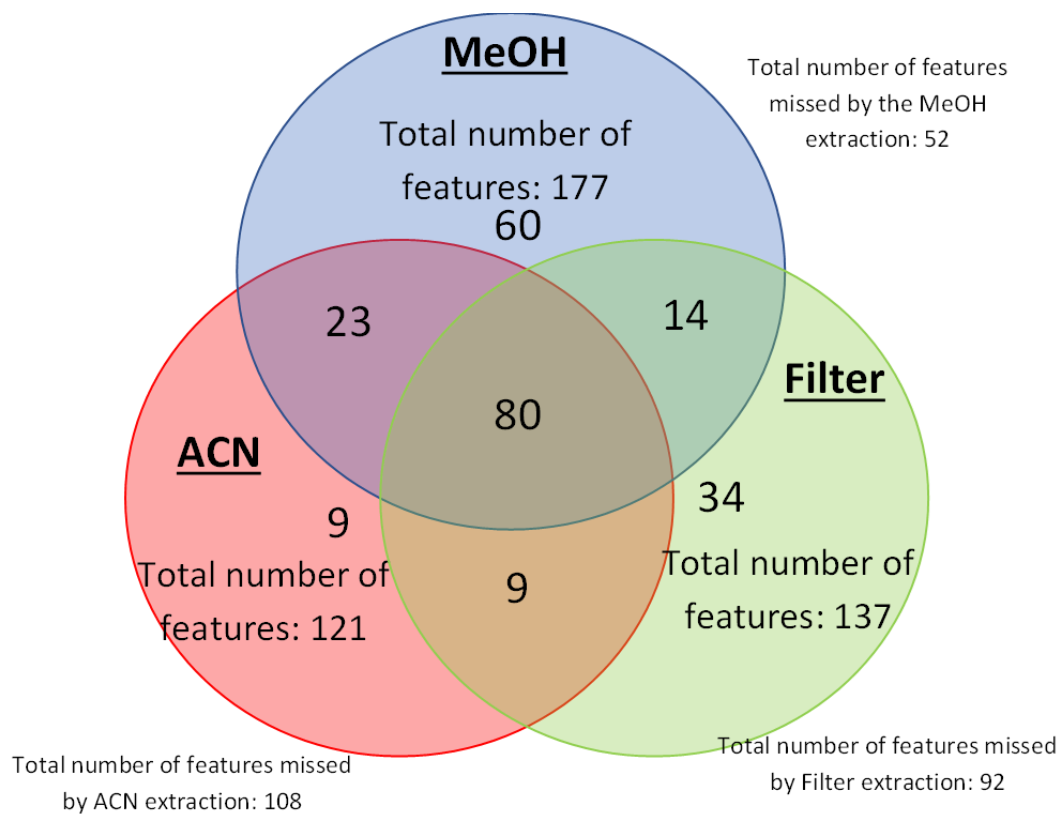


Figure 4-2- The numbers of metabolites captured by the three different extraction methods used

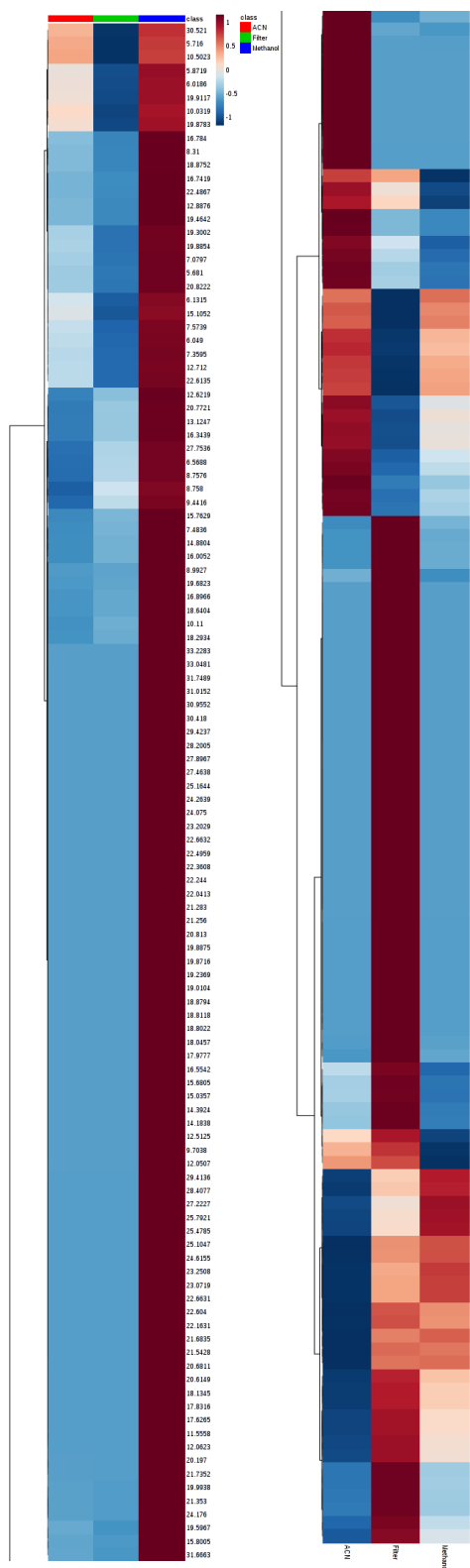


Figure 4-3- Heat map showing the intensities of the 25 most distinguishing features between the three different extraction methods used- illustrating the majority of these to be generated by methanol extraction

A final consideration when deciding the most effective extraction method was the reproducibility of the extraction. This consistently of chromatographic area for each peak was determined for the triplicate extraction. This was achieved through calculating the average number of times unique peaks were successfully reported. Methanol achieved an average extraction score of 2.60, water filtration 2.40 and ACN 2.14; which, when taken into account alongside the lower overall number of metabolites that were identified with ACN and aqueous filtration extraction identifies methanol as the most optimal extraction solvent. The number of unique features that were successfully extracted in all three triplicate extractions was also determined and methanol produced a total of 124 metabolites which was significantly greater than the number of consistently extracted metabolites for the other approaches (aqueous filtration: 74 metabolites and with ACN extraction there were 51 metabolites). Additionally, the consistency of the peak intensity of the metabolites was analysed using standard error, with a reduced standard deviation for the average of replicate extractions indicating a more reliable recording of intensity. Again methanol proved to be the most successful in this regard ( $SD= 6.5e-5-0.333$  normalised abundance units), followed by the aqueous filtered extraction ( $0.0002-0.383$  NAU) and ACN extraction showed a slightly poorer reproducibility ( $SD= 0.0003-0.298$  NAU).

#### 4.3.2. Elephant reproductive status biomarker analysis

The methanol metabolite extraction was next applied to the elephant faecal samples. After spectral alignment, normalisation and correction procedures were applied to the 60 samples in the dataset, 117 metabolites were consistently identified in order to undertake further comparative analysis. The first stage of this analysis involved comparing the non-cycling, cycling and pregnant groups against each other in order to highlight any statistically significant differences. These selected metabolites of interest were then further studied in order to determine which metabolite levels best distinguished the three groups of elephants, before being assigned a metabolite identification. Following the statistical analysis conducted between the twenty non-

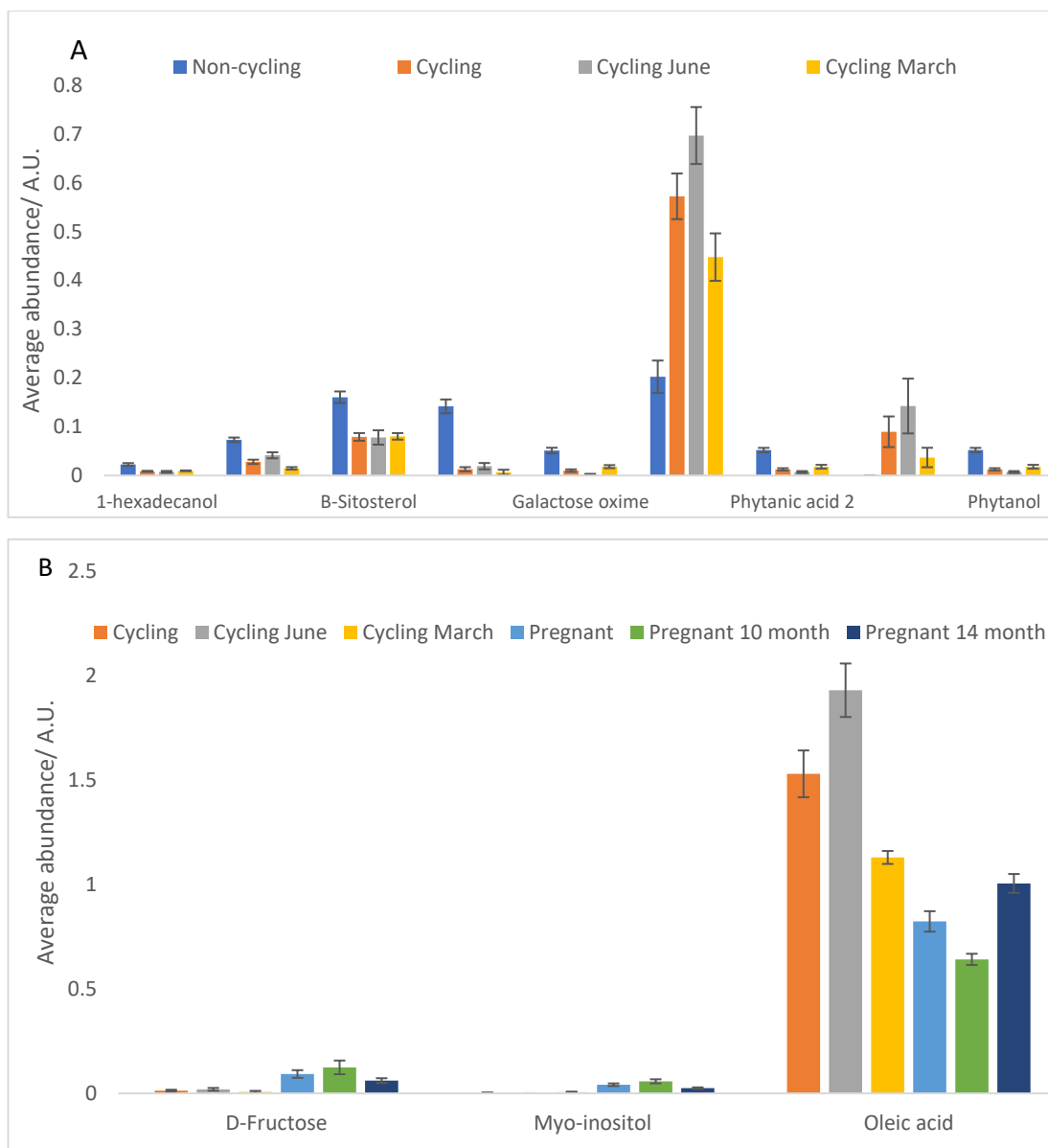


cycling and twenty cycling elephant samples, a total of fifty metabolites were found to exhibit a significant differentiation between the groups. Of these, nineteen metabolites were selected as being statistically significant following the application of a Bonferroni correction. Of these nineteen metabolites, there was no clearly definitive pattern with regards to metabolite level changes, but there was a tendency for up-regulation to be seen in the non-cycling group, with a total of thirteen out of the nineteen metabolites exhibiting this trend. Of these nineteen metabolites, nine scored a better than “good” diagnostic biomarker value as detailed later and the levels of these metabolites are shown in Figure 4-4a. Due to the fact that cycling samples were collected across 2 time periods within a cycle, intra-group variations were sought out to ensure that the metabolites indicated as being significant between the groups could not be attributed to the time at which the sample was collected during a cycle. Six of the metabolites that exhibited a significant difference within the cycling group also showed significant variance between non-cycling and cycling elephants; namely sebatic acid, phytanol, galactose oxime, glycerol and benzene acetic acid. Despite the significant intra-group variations within these metabolic profiles, the trend between this group when compared to the non-cycling group always displayed a consistent metabolite level change showing that regardless of when the sample had been taken during the cycling term, a significant difference from the non-cycling group is always observed. Cycling groups were always found to be significantly upregulated in glycerol and sebatic acid when compared to non-cycling groups, the trend being most distinguishable during the later period (June) samples. Galactose oxime, benzeneacetic acid, and phytanol were always found to be up-regulated in non-cycling groups than in cycling. With galactose oxime and phytanol the trend was more prominently observed in the later (June) samples, whereas with benzeneacetic acid the trend was more prominent in the earlier (March) sample (Figure 4-4a).

When comparing cycling and pregnant elephants, a total of fifty-five metabolites were found to initially show a statistically significant difference, with only 10 of these remaining significantly different following the Bonferroni correction. Half of the metabolites exhibited up-regulation in pregnant elephants and half in cycling

elephants and only three of the metabolites of interest achieved good diagnostic properties in later analysis and these are shown in Figure 4-4b. Of interest when looking at variations in metabolite levels between the two cycling period groups and the two pregnant period groups (each groups total samples representing collection at two time periods with their different cycles), only oleic acid was found to differ significantly between the intra-group variation of cycling elephant samples of those metabolites highlighted as distinguishing between cycling and pregnant elephants, hence suggesting that oleic acid may be a less reliable diagnostic indicator across all timepoints for these two groups. However it should be noted that the trend in oleic acid levels always represented an increase in cycling elephants compared to the pregnant cohort (Figure 4-4b).

Finally, when comparing non-cycling and pregnant elephants, a total of thirty-eight metabolites were initially found to show a significant difference, ten retaining significance following a Bonferroni correction, all of which were significantly up-regulated in cycling elephants when compared to pregnant elephants. Three of these metabolites were identified as being the most significant, achieving good diagnostic properties and the comparative levels of these are shown in Figure 4-4c, remaining significant regardless of the timepoint samples were taken for the pregnant elephants. It is interesting to identify metabolites which express significantly altered levels in non cycling elephants compared to the other two groups and this study identified four metabolites which show significant variance between every group of interest. Phytanol, Phytanic acid, Benzene acetic acid and D-Fructose exhibited a significant elevation in non-cycling elephants with regard to Phytanol and Phytanic acid, Pregnant elephants represent the elephant group with metabolite levels closest to the non cycling group whereas cycling elephants are the closest comparison group for levels of Benzene acetic acid and D-Fructose (Figure 4-4d).



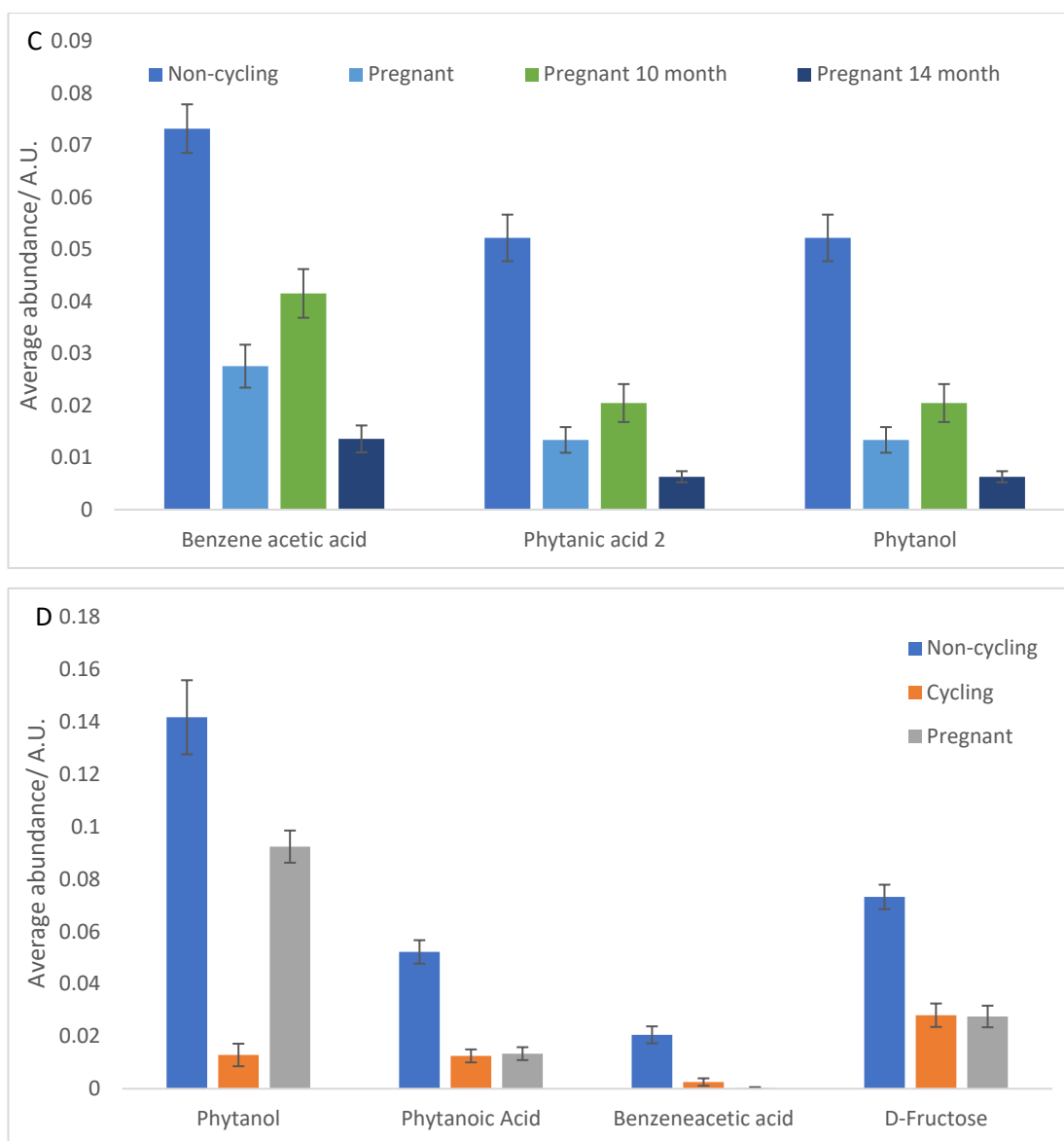


Figure 4-4- Metabolites with differences between A) non cycling and cycling elephants, B) cycling and pregnant elephants and C) non cycling and pregnant elephants. D) metabolites shown to be elevated in non-cycling elephants when compared to the other two groups. Error bars= SD. Relevant statistical P values summarised in Table 4-1.

Table 4-1- A summary of the metabolites that distinguish different groups of elephants and their comparative levels of distinction ordered by significance. Groups were N= non cycling, C= cycling, P= pregnant. n.c.= likelihood ratios that could not be determined as the metabolite level was never elevated in the group exhibiting a decreased level of the metabolite.

Tentative metabolite identification	Comparison groups	Fold change (group)	Statistical P value	AUC	Sens.	Spec.	+LHR	-LHR
Phytanic acid	N vs P	68.2↑(N)	4.27E-08	0.974	95	0	n.c.	0.05
D-Frucotse	N vs C	11↑ (N)	5.66E-08	0.995	95	0	n.c.	0.05
Myo-Inositol	C vs P	15.5↑ (P)	1.41E-07	0.968	1	10	10	0
Sebacic acid	N vs C	90↑ (C)	3.24E-07	0.925	85	0	n.c.	0.15
Phytinol	N vs P	3.9↑(N)	3.67E-07	0.97	90	0	n.c.	0.1
D-Frucotse	C vs P	7.2↑ (P)	6.94E-07	0.953	90	15	6	0.12
Phytinol	N vs C	4.2↑ (N)	7.41E-07	0.958	95	10	9.5	0.56
Benzene acetic acid	N vs P	2.7↑(N)	9.78E-07	0.953	90	15	6	0.12
Galactose oxime	N vs C	5.1↑ (N)	1.69E-06	0.94	95	10	9.5	0.06
Oleic acid	C vs P	1.9↑ (C)	1.69E-06	0.943	95	15	6.33	0.06
Phytanic acid	N vs C	8.1↑ (N)	1.71E-06	0.925	95	15	6.33	0.06
Glycerol	N vs C	2.8↑ (C)	2.12E-06	0.938	85	5	17	0.16

Benzene acetic acid	N vs C	2.6↑ (N)	2.20E-06	0.938	90	10	9	0.11
β-Sitosterol	N vs C	2↑ (N)	2.52E-06	0.935	90	10	9	0.11
1-Hexadecanol	N vs C	2.6↑ (N)	4.25E-06	0.925	85	10	8.5	0.17

In order to determine the diagnostic accuracy with which metabolites identified are able to distinguish the different groups of elephants, diagnostic accuracy was determined using a number of parameters and the outcomes of this analysis are summarised in Table 4-1. The table shows that metabolite levels vary between a roughly two-fold change between comparator groups and an almost 100 fold change (sebacic acid between non-cycling and cycling elephants). All of the significantly different metabolites (following correction for multiple hypothesis testing) exhibit an area under the generated receptor operator characteristic curve (AUC) of greater than 0.9 representing a clear distinction between groups when these metabolites levels are analysed. The reported likelihood ratios indicate the probability of an elevation (+LHR) and decrease (-LHR) in the respective groups (derived from the sensitivity and 1-specificity values with greatest diagnostic benefit from the ROC curves) and all indicate good to excellent levels as outlined in a previous study categorising diagnostic benefit (Ray. P., et al, 2010). In order therefore to consider the power of the group of the 11 key metabolic biomarkers identified in Table 4-1 to allow for discrimination between the three separate groups of elephants, a heatmap was produced detailing the comparative levels of all 11 in each elephant within each of the three identified groups (Figure 4-5). This provides an overview as to how the animals were grouped by the collective levels of these potential biomarkers. As can be seen in this analysis, the two groups most commonly closely linked are the cycling and pregnant individuals, however the pregnant elephants exist as a separate subgroup mainly within this group. Five pregnant individuals exist on the non-cycling branch following this analysis, however 4 of these appear as a separate subgroup

within this branch and are distinguished from the other elephants within this area of the heat map. As outlined in Figure 3 previously, elevations often occur in the non-cycling individuals for the majority of the identified metabolites, with those not elevated in this group mainly exhibiting elevations in the cycling group rather than the pregnant individuals. Overall, Figure 4-5 demonstrates the ability of the panel of diagnostic metabolites to differentiate elephants from the three different groups well.

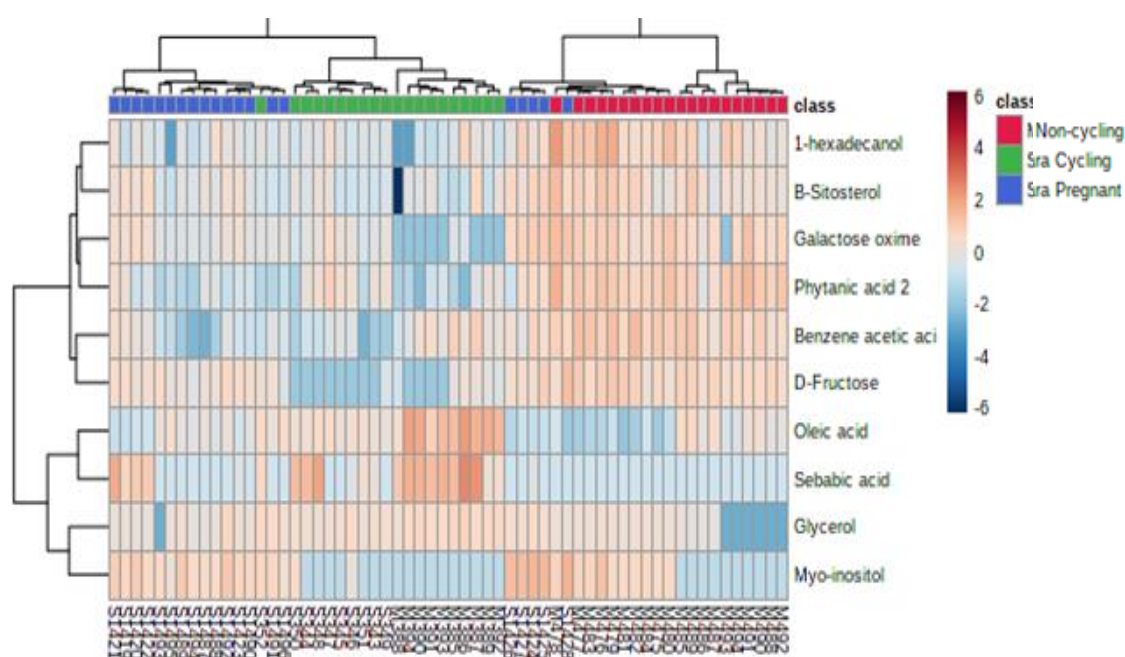


Figure 4-5- Heat map detailing the differentiation of the elephants within the three groups based upon normalised levels of the metabolites identified as having discriminatory power.

#### 4.4- Discussion

This study shows that metabolite profiling of faecal metabolites allows for the distinction of the three groups of elephants (cycling, non-cycling and pregnant) within this study and whilst as a panel the animals can be clustered mainly into three subgroups, certain metabolite levels have a better discriminatory power than others and the potential biological consequences of these metabolites are interesting to consider.

When considering the non-cycling group of elephants, three metabolites, 1-hexadecanol, Phytanic acid and Benzene acetic acid were consistently seen to be the most significantly upregulated metabolites when compared to both the cycling and pregnant groups of elephant. Additional to this, non-cycling elephants demonstrated the highest normalised levels of  $\beta$ -Sitosterol, D-Fructose and Galactose oxime, but not always at significant levels of statistical distinction. Non-cycling elephants always exhibited the lowest levels of Glycerol, Oleic acid and Sebacic acid, though again this was not always significant between all three groups. Unsurprisingly, many of these have an initial dietary source. Phytanic acid is an isoprenoid lipid occurring in the diet exclusively from green vegetation and has a number of biological roles including use as an anti-inflammatory, antiallergenic and as an anti-oxidant (de Moraes J, et al, 2014). Increases in the levels of Phytol within mice has been found to influence the usage and oxidation of other fatty acids in the body (Gloerich, J., et al, 2005) and also glucose metabolism (Heim, M., et al, 2002). Correct ability of animals to utilise phytanic acids is also believed to have preventative effects on metabolic dysfunction through its increase of genes involved in fatty acid oxidation, glucose uptake enhancement and hepatocytes and improved differentiation of brown adipocyte and expression of uncoupling protein-1 (Werner, L. B., et al, 2011). Any alterations in such fatty acid regulated processes could potentially lead to issues in fertility and as such this could explain the elevation of this metabolite in this group. Furthermore, clear associations between higher levels of phytanic acid and defects in steroid hormone synthesis have been reported (Wierzbicki, A.S., et al, 2002).



A further dietary derived metabolite,  $\beta$ -sitosterol was also found to be upregulated in non-cycling elephants when compared to cycling in this study. Numerous mammalian hormones can be derived from  $\beta$ -sitosterol obtained through the diet, including progesterone and the metabolite is weakly estrogenic and can therefore have various impacts on fertility in mammals, including infertility in male rats in which it inactivates sperm mobility and acting as a contraceptive in goats (Chattopadhyay, D., et al, 2005). Studies have also found that  $\beta$ -sitosterol can cause ovarian weight to be reduced with haemorrhage into cystic follicles in sheep (Burton, J. L., & Wells, M., 2002). Additionally misregulation of dietary phytosterols, such as  $\beta$ -sitosterol, is also known to lead to maturational, morphological, cyclic and fertility issues, such as phytoestrogen ingestion in sheep leading to temporary infertility in ewes (typically upto 6 months) or sometimes permanent infertility issues in some instances or prolonged exposure to the feed causing permanent alterations to the cervix (Adams N. R., 1995).  $\beta$ -sitosterol has been linked to cycling issues in various mammals through its ability to increase ovarian steroidogenesis, and therefore levels of estrogen, altering the hormonal balance required to maintain reproductive tissues and can therefore lead to issues with fertility (Shabanian S., et al, 2016).

When considering metabolites that differentiate cycling from non-cycling animals, a reduction in faecal fructose levels was determined. Fructose is one of the predominant sugars found in plants and therefore again could have an initial dietary source. It has a role in the body of measuring and maintaining sugar consumption levels and has been shown to positively correlate to the body conditioning score in female Asian elephants (Hannou, S. A., et al, 2018; Norkaew, T., et al, 2018). Elevated levels of fructose in association to non-pregnant status can however give indication of weight issues and high levels of stress, which in captive elephants, can in turn cause infertility issues (Norkaew, T. et al., 2018). Sebacic acid, a medium chain, saturated fatty acid, was found to be significantly elevated in the faeces of non-cycling elephants when compared to cycling elephants faecal samples. Deficiency in metabolism of fatty acids has long been understood to impact fertility in a number of ways, including the ability for the body to produce various hormones and maintain embryos during early-stage pregnancy (Ye, Q., et al, 2019). Further studies have also

shown that sebatic acid can contribute to a significant anti-inflammatory response through impact on TNF- $\alpha$ , and additionally through modulation of estrogen receptors (Moutsatsou, P., et al, 2010; Chen, Y. F., et al, 2016).

When considering the metabolites present in the pregnant elephant faecal samples, oleic acid exhibited the greatest variation between timepoints within the pregnancy group. The alterations in oleic acid during pregnancy (and also cycling) could reflect increasing fat content during pregnancy, supporting embryonic and mammary tissue growth, which intensifies more so in the later stages of pregnancy. Circulating oleic acid has been previously shown to increase significantly during pregnancy (Crocker, I., et al, 1999) and it has been reported to be critical to development, with involvement noted in germ cell growth and development (Fayezi, S., et al, 2018) implantation and the correct development of a foetus. Another metabolite, Myo-Inositol was found to be significantly elevated in pregnant elephant faeces compared to cycling elephants. Myo-inositol is a sugar alcohol that acts as a second messenger, regulating numerous hormones such as follicular stimulating hormone (FSH). Additionally, it holds a crucial role in regulating glucose levels through improving insulin sensitivity and as such has been used to treat polycystic ovary syndrome in human candidates as well as gestational diabetes (Unfer, V., et al 2017; Sobota-Grzeszyk, A., et al, 2019). It is also a major constituent of the milks of various mammalian species and so will be in greater demand during pregnancy (Unfer, V., et al 2017; Byun, S. M., & Jenness, R., 1982).

Though limited by study size being only two elephants could be included in the analysis, this study indicates the potential for untargeted metabolomic analysis of faecal samples to allow for metabolite panels that collectively and individually allow for the differentiation of animals (elephants in this current study), based upon their reproductive status. The faecal analysis allows for the study of what dietary elements are being utilised and rejected by the individual animal at given points in their reproductive cycle and hence can contribute to a greater understanding of the importance of dietary elements for the reproductive status of an individual. The sample also may reflect overall metabolism for the animal and provide insight into the bioavailable metabolites for an individual which, again may have implications for

reproductive and overall health status. Hence, it can be seen that this analysis allows for a non-invasive analysis of individual animals health status and may have benefits in successful care of animals and reproduction within the field of animal conservation.

#### 4.4.1- Future perspectives

This study managed to evidence several markers that could be associated to the fertility status of captive elephants. This study is limited by the low sample set being only two separate elephants were analysed and so expanding it to include more animals with better defined cycling points would vastly improve the understanding of what is happening during these stages. This study utilised wet samples as literature suggests this will give the best preservation of hormonal metabolites, however, this does allow for a bias due to elephant's hydration levels and so a further study making comparisons between wet and lyophilised samples would be beneficial to compare profiles obtained in both a qualitative and quantitative way. This study was conducted pre-covid and a lot of the markers of interest were attributed to stress related markers. Over the pandemic the zoo was shut down to the public and as such a follow up study to see how these markers are impacted with a lower footfall in the zoos could provide detail on the impact of visitors on stress.

#### 4.5- References

- Adams N. R. (1995). Detection of the effects of phytoestrogens on sheep and cattle. *J Animal Sci*, 73(5), 1509–1515.
- Allen W.R. (2006). Ovulation, pregnancy, placentation and husbandry in the African elephant (*Loxodonta africana*). *Philos Trans R Soc Lond B Biol Sci*. 361(1469), 821–834.
- Burton, J. L., & Wells, M. (2002). The effect of phytoestrogens on the female genital tract. *J of Clin Path*, 55(6), 401–407.
- Byun, S. M., & Jenness, R. (1982). Estimation of free myo-inositol in milks of various species and its source in milk of rats (*Rattus norvegicus*). *J of Dairy Sci*, 65(4), 531–536.
- Cesbron, N., Royer, A. L., Guitton, Y., Sydor, A., Le Bizec, B., & Dervilly-Pinel, G. (2017). Optimization of fecal sample preparation for untargeted LC-HRMS based metabolomics. *Metab*, 13(8), 1–6.
- Chattopadhyay, D., Dungdung, S. R., Das, K., Saha, S., Mandal, A. B., & Majumder, G. C. (2005). Sperm motility inhibiting activity of a phytosterol from *Alstonia macrophylla* Wall ex A. DC. leaf extract: a tribal medicine. *Ind J Exp Bio*, 43(11), 1104–1109.
- Chen, Y. F., Wang, K., Zhang, Y. Z., Zheng, Y. F., & Hu, F. L. (2016). *In vitro* anti-inflammatory effects of three fatty acids from royal jelly. *Meds of Inflam*, 3583684.
- Cheng, K., Brunius, C., Fristedt, R., & Landberg, R. (2020). An LC-QToF MS based method for untargeted metabolomics of human fecal samples. *Metabolomics: Official J Metab Soc*, 16(4), 46.
- Crocker, I., Lawson, N., Daniels, I., Baker, P., & Fletcher, J. (1999). Significance of fatty acids in pregnancy-induced immunosuppression. *Clin & Diag Lab Immuno*, 6(4), 587–593.

de Moraes J, de Oliveira RN, Costa JP, et al. (2014). Phytol, a diterpene alcohol from chlorophyll, as a drug against neglected tropical disease Schistosomiasis mansoni. *PLoS Negl Trop Dis*. 8(1), e2617.

Fayezi, S., Leroy, J., Ghaffari Novin, M., & Darabi, M. (2018). Oleic acid in the modulation of oocyte and preimplantation embryo development. *Zygote*, 26(1), 1–13.

Garnier J. N. , Green D. I. , Pickard A. R. , Shaw H. J. Holt W. V. (1998). Non-invasive diagnosis of pregnancy in wild black rhinoceros (*Diceros bicornis minor*) by faecal steroid analysis. *Repro, Fertility & Devel*. 10, 451-458

Ghosal R., Seshagiri. P.B., and Sukumar. R. (2021). Dung as a potential medium for inter-sexual chemical signaling in Asianelephants (*Elephas maximus*). *Behav Processes*. 91(1),15-21

Gloerich, J., van Vlies, N., Jansen, G. A., Denis, S., Ruiter, J. P., van Werkhoven, M. A., Duran, M., Vaz, F. M., Wanders, R. J., & Ferdinandusse, S. (2005). A phytol-enriched diet induces changes in fatty acid metabolism in mice both via PPARalpha-dependent and -independent pathways. *J of Lipid Res*, 46(4), 716–726

Hadinger,U., Haymerle, A., Knauer, F., Schwarzenberger, F., & Walzer, C., (2015) Faecal cortisol metabolites to assess stress in wildlife: evaluation of a field method in free-ranging chamois. *British ecological society*. 6 (11)1349-1357

Hannou, S. A., Haslam, D. E., McKeown, N. M., & Herman, M. A. (2018). Fructose metabolism and metabolic disease. *J of Clin Inv*, 128(2), 545–555.

Heim, M., Johnson, J., Boess, F., Bendik, I., Weber, P., Hunziker, W., & Fluhmann, B. (2002). Phytanic acid, a natural peroxisome proliferator-activated receptor (PPAR) agonist, regulates glucose metabolism in rat primary hepatocytes. *FASEB journal*, 16(7), 718–720.

Howell-Stephens, J., Bernier, D., Brown, J. S., Mulkerin, D., & Santymire, R. M. (2013). Using non-invasive methods to characterize gonadal hormonal patterns

of southern three-banded armadillos (*Tolypeutes matacus*) housed in North American zoos. *Animal Repro Sci.* 138(3-4), 314–323.

Kajaysri. J., and Nokkaew W. (2014). Assessment of pregnancy status of Asian elephants (*Elephas maximus*) by measurement of progestagen and glucocorticoid and their metabolite concentrations in serum and feces, using enzyme immunoassay (EIA). *J Vet Med Sci.* 76(3), 363-8

Keady, M. M., Prado, N., Lim, H. C., Brown, J., Paris, S., & Muletz-Wolz, C. R. (2021). Clinical health issues, reproductive hormones, and metabolic hormones associated with gut microbiome structure in African and Asian elephants. *Animal microbiome*, 3(1), 85. <https://doi.org/10.1186/s42523-021-00146-9>

Klinhom, S., Sriwichain, S., Kerdphoo, S., Khonmee, J., Chattipakorn, N., Chattipakorn, S. C., & Thitaram, C. (2023). Characteristics of gut microbiota in captive Asian elephants (*Elephas maximus*) from infant to elderly. *Scientific reports*, 13(1), 23027. <https://doi.org/10.1038/s41598-023-50429-1>

Kumar. V., Palugulla. R.V., Kokkiligadda. A., Shivaji. S., and Umapathy. G. (2014). Non-invasive assessment of reproductive status and stress in captive Asian elephants in three south Indian zoos. *Gen Comp Endocrinol.* 201, 37-44.

Lopez-Bascon, M. A., Calderon-Santiago, M., Arguello, H., Morera, L., Garrido, J. J., & Priego-Capote, F. (2019). Comprehensive analysis of pig feces metabolome by chromatographic techniques coupled to mass spectrometry in high resolution mode: Influence of sample preparation on the identification coverage. *Talanta*, 199, 303–309.

Metrione, L. C., & Harder, J. D. (2011). Fecal corticosterone concentrations and reproductive success in captive female southern white rhinoceros. *Gen & Comp Endo.* 171(3), 283–292.

Moutsatsou, P., Papoutsis, Z., Kassi, E., Heldring, N., Zhao, C., Tsiapara, A., Melliou, E., Chrousos, G. P., Chinou, I., Karshikoff, A., Nilsson, L., & Dahlman-Wright, K. (2010). Fatty acids derived from royal jelly are modulators of estrogen receptor functions. *PloS one*, 5(12), e15594.

Mushtaq, M. Y., Choi, Y. H., Verpoorte, R., & Wilson, E. G. (2014). Extraction for metabolomics: access to the metabolome. *Phytochemical analysis : PCA*, 25(4), 291–306.

Niemuller C.A., Shaw H. J., and Hodges J. K. (1998). Pregnancy determination in the Asian elephant (*Elephas maximus*): A change in the plasma progesterone to 17 $\alpha$  hydroxyprogesterone ratio. *Zoo Bio*. 16(5),415-426

Norkaew, T., Brown, J. L., Bansiddhi, P., Somgird, C., Thitaram, C., Punyapornwithaya, V., Punturee, K., Vongchan, P., Somboon, N., & Khonmee, J. (2018). Body condition and adrenal glucocorticoid activity affects metabolic marker and lipid profiles in captive female elephants in Thailand. *PloS one*, 13(10), e0204965.

Palme R., (2005) Measuring fecal steroids- guidelines for practical application. *Acad. Sci*. 1046(1)75-80

Ray. P., Manach Y., Riou B, and Houle T.T. (2010). Statistical Evaluation of a Biomarker. *Anesth*, 112 (4), 1023-1040.

Rees P.A., (2003). Asian elephants in zoos face global extinction: should zoos accept the inevitable? *Oryx*. 37(1), 20-22

Schwarzenberger F. (2007). The many uses of non-invasive faecal steroidmonitoring in zoo and wildlife species. *Int. Zoo Yb*. 41: 52

Shabani S., Bahmani. M., and Asadi-Samani . M. (2016). The medicinal plants effective on female hormones: A review of the native medicinal plants of Iran effective on estrogen, progesterone, and prolactin. *J Chem & Pharm Sci*. 9(3):1270-1276

Sobota-Grzeszyk, A., Kuźmicki, M., & Szamatowicz, J. (2019). Myoinositol in the Prevention of Gestational Diabetes Mellitus: Is It Sensible?. *J of Diab Res*, 3915253.

Unfer, V., Facchinetti, F., Orrù, B., Giordani, B., & Nestler, J. (2017). Myo-inositol effects in women with PCOS: a meta-analysis of randomized controlled trials. *Endocrine Comms*, 6(8), 647–658.

Veeraselvam, M., Selvaraj, P., Senthil Kumar, S., Senthilkumar, A., Senthilkumar TMA & C Sreekumar (2021) Elephants: this needs to be considered as a confounding issues that will affect your results..in the discussion- not knowing time point of cycle. *The pharma innovation journal*. SP-10(10), 892-895

Wasser, S. K., Monfort, S. L., & Wildt, D. E. (1991). Rapid extraction of faecal steroids for measuring reproductive cyclicity and early pregnancy in free-ranging yellow baboons (*Papio cynocephalus cynocephalus*). *J. Repro & Fertility*. 92(2), 415–423

Werner, L. B., Hellgren, L. I., Raff, M., Jensen, S. K., Petersen, R. A., Drachmann, T., & Tholstrup, T. (2011). Effect of dairy fat on plasma phytanic acid in healthy volunteers - a randomized controlled study. *Lipids in Health & Disease*, 10, 95.

Wierzbicki AS, Lloyd MD, Schofield CJ, Feher MD, Gibberd FB. (2002). Refsum's disease: a peroxisomal disorder affecting phytanic acid alpha-oxidation. *J Neurochem*. 80(5), 727-735.

Williams, C. L., Ybarra, A. R., Meredith, A. N., Durrant, B. S., & Tubbs, C. W. (2019). Gut Microbiota and Phytoestrogen-Associated Infertility in Southern White Rhinoceros. *mBio*, 10(2), e00311-19

Ye, Q., Cai, S., Wang, S., Zeng, X., Ye, C., Chen, M., Zeng, X., & Qiao, S. (2019). Maternal short and medium chain fatty acids supply during early pregnancy improves embryo survival through enhancing progesterone synthesis in rats. *The J of Nut. Biochem*, 69, 98–107.



## Chapter 5 - Seeking plasmalogen biomarkers from *Drosophila* flies for utilisation as a model for Alzheimer's disease

### 5.1 Aims

Lipids constitute a significant proportion of the brain's composition, with ethanolamine glycerophospholipids, including plasmalogens, being the second most prevalent. Plasmalogens play critical roles in membrane fusion, fission, protein integration, fluidity, and ion permeability, and disruptions in their levels have been observed in the brains of Alzheimer's patients. A large issue with Alzheimer's research is capturing diagnostic markers at an early point, often the disease is well established by the point of diagnosis in a large enough sample number for evaluation. This study explores the potential of *Drosophila melanogaster* (fruit flies) as a model to address these challenges, given their ability to convey the earliest stages of Alzheimer's onset through exhibiting traceable behaviours and the presence of orthologues genes with humans associated the condition. Through lipidomic profiling of drosophila brain samples, this research provides the first evidence that *Drosophila* produce plasmalogens. Furthermore, disruptions in lipid production caused by mutations in plasmalogen synthesis genes associated with the enzymes GNPAT, FAR1, AGPS and PEDS1 were mirrored in the fly model, reflecting similar effects observed in humans. These findings demonstrate that *Drosophila* can serve as a viable model for studying Alzheimer's disease, offering a promising avenue for early-stage research and biomarker discovery.

### 5.2- Introduction

#### 5.1.1- Lipids and neurological disruptions

Lipids provide a fundamental structural, regulatory, signalling, and metabolic role within all tissues of the body, one such tissue of major lipid prevalence is the brain. Approximately 50% of the brains dry weight is made up of lipids, here it has a more specialist role, with high prevalence noted in various neurological structures including in the neurons, myelin and membrane (Hornemann T. 2021). It is therefore no

surprise to see associations between lipid mis-regulation and various neurological disorders. There are eight different categories of lipids according to LIPID MAPS, with varying subclassifications within each; Fatty acyl (FA), Glycerolipids (GL), Glycerophospholipids (GP), Sphingolipids (SP), Sterol Lipids (ST), Prenol Lipids (PR), Saccharolipids (SL) And Polyketides (PK) (Figure 5-1)

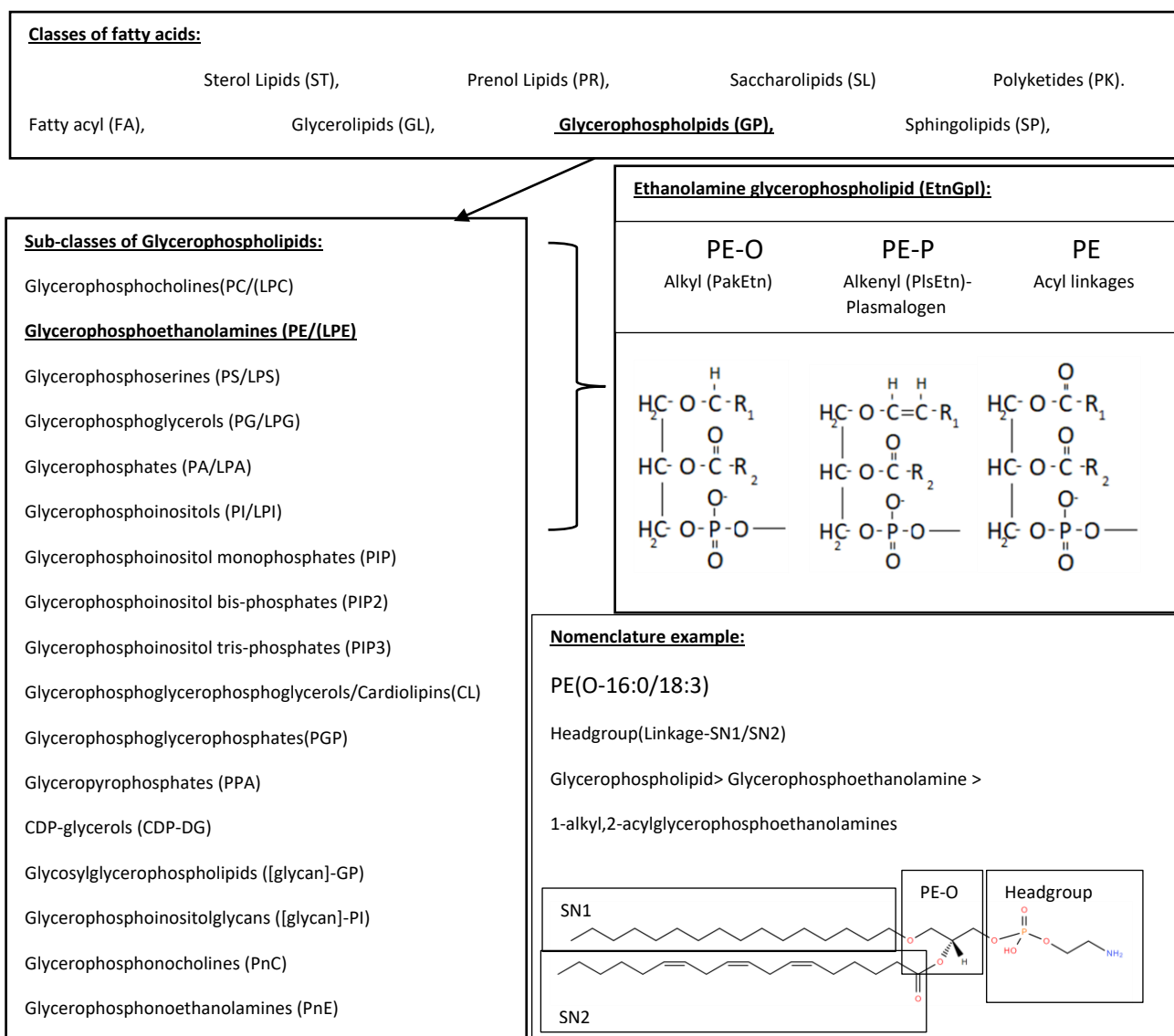


Figure 5-1- Lipid nomenclature with focus on Glycerophospholipid group glycerophosphoethanolamine (PE)

GP encompass any derivative of glycerophosphoric acid including PC and PE (including plasmalogens), which biologically contribute to the vast number of structural/membrane lipids as well as signalling molecules. The sub-classes of GP's of interest, namely PC and PE, exists in three forms dependent on the sn-1 position linkages attached to the glycerol residue. These consist of alkyl (PE), alkenyl (PakEtn) and plasmalogen (PlsEtn) (Figure 5-1).

Glycerophosphocholines (PC) are the most abundant glycerophospholipids in eukaryotic cells being a major component of the cellular membrane as well as baring roles in cell division, growth and signalling and holds a particular prevalence in the brain. Additional to this PC is involved in major biosynthesis pathways enabling the production of other lipids required for neurological function and structure. Specific PC's reduction has known association with Alzheimer's disease (Whiley, L., et al., 2014)

Phosphatidylethanolamine (PE) is the second most abundant glycerophospholipid in eukaryotic cells (15% prevalence), after the previously discussed PC, and is particularly enriched in the inner leaflet of the plasma membrane as a non-bilayer forming phospholipid. The conical shaped structure of PE is formed due to its small polar head group and longer fatty acid chain moiety leading to PE's existence in hexagonal phases. This structural adaptation enables it to be utilised in membrane curvature and key membrane functioning including membrane fusion, fission, protein integration (via glycosylphosphatidylinositol anchors) and allowing for conformational changes in protein structures (Calzada, E., et al., 2016; Phan, N. T., et al., 2015; Vance J.E. 2015) as well as impacting the neural membranes fluidity and ion permeability (Farooqui, A.A., et al., 2009). The quantities of PE vary throughout different biological systems, with it representing approximately 45% of the total phospholipid content of mammalian brains. PE is a precursor for several other classes of phospholipids, mainly PC, PS and PI lipids. This conversion occurs in the mitochondria and as such there is typically a high abundance of PE present in this organelle. In the endoplasmic reticulum the mitochondrial associated membrane (MAM) is the site of the enzymes involved in the lipid synthesis including PS/PI synthase and PE methyl transferase. PE also have a role in mitochondria biogenesis,

which may also explain the close association. It is also a substrate for post-translational modifications and has a role in oxidative phosphorylation and autophagy.

Of particular interest in this study is the presence of plasmalogens, a subclass of both lipidomic groups, in neurological disease. As previously discussed, these are a subgroup of specific glycerophospholipids identified through the presence of vinyl ether linkage in the sn-1 position and an ester linkage in the sn-2 position, Figure 5-1. Plasmalogens have presence in most cells of the body but have particularly dominant presence in the brain being the main glycerophospholipid. Plasmalogen's concentrations vary in different elements of the brain, accounting for approximately 60% of the total ethanolamine phospholipids in grey matter and 80% in the white matter as well as being found to be prevalent in the sarcolemma, myelin sheath (85%), neurones (50%) and synaptic vesicles (Gu, J. et al., 2022; Kao, Y. C., et al., 2020). Ethanolamine plasmalogens is the most abundant form of glycerophospholipid in the brain, followed by PE and then PC. PE plasmalogens are more abundant than the PC plasmalogens in all cells except the brain (Kao, Y. C., et al., 2020). Alterations in their levels are known to be causative of various health consequences, of note, a depletion in the overall presence of plasmalogens is observed in Alzheimer's disease, Parkinson's disease and Multiple sclerosis (MS) disorders with a direct correlation seen between disease severity and lipid fluctuations (Calzada, E., et al., 2016). Plasmalogens have additional association with other inherited conditions such as Rhizomelic chondrodysplasia punctata, Zellweger spectrum disorders (Dean, J. M., & Lodhi, I. J., 2018). Plasmalogen dysregulation in human brains is a known trait of Alzheimer's sufferers. Due to the large presence of myelin in the human brain, accompanied by heavy plasmalogen content, this accounts for the majority of the change being noted. Neuronal fluctuations need further investigation to see how they are impacted in diseased states as well as what implications these may hold.

There are several key roles of plasmalogens in the brain that enable normal neurological functioning through various routes. One of the most important roles of plasmalogens is structural, with their cell membrane functioning. The chains of

plasmalogens, due to their unique linkages, allow for a tighter packing of lipids in the membranes. As a result, the presence of plasmalogens decreases the fluidity of the membranes. This is particularly important in stable myelin sheath structures as well as in membrane trafficking, mediating secondary messengers for signalling molecules. Plasmalogens additionally have roles as secondary messenger precursors, neurotransmission, cell cycling, energy sources and PUFA presence as well as acting as antioxidants and cell signalling molecules (Dean, J. M., & Lodhi, I. J. 2018 & Phan, N. T. N., et al.,2017).

## **Lipid Biosynthesis**

### **PE**

The main route for PE biosynthesis is via the CDP-ethanolamine and phosphatidylserine decarboxylase pathway, both of which are the major sources of PE production in mammalian systems. During this pathway ethanolamine is first converted to phosphoethanolamine via ethanolamine kinase enzyme. The second step involves the activation of to CDP- ethanolamine by CTP:phosphoethanolamine cytidyltransferase, which is the rate limiting step, before the transfer of the phosphoethanolamine group to Diacylglycerol (DAG) by ethanolamine transferase (Bleijerveld, O. B., et al., 2006). As an alternative pathway, PE can also be formed through the acylation of Lyso-PE via the enzyme 1-acylglycerophosphocholine-O-acyltransferase, which additionally generates coenzyme A as a by-product.

### **Plasmalogen**

Ether lipids, such as plasmalogens, undergo biosynthesis initially in the peroxisome via the acyl dihydroxyacetone DHAP pathway, Figure 5-2a. The first step converts glycerone-phosphate (DHAP), a product and intermediate of glycolysis, into Acyl-DHAP. This is done through the acylation of DHAP with a long chain Acyl-CoA to the sn-1 position via the enzyme glycerone phosphate O-acyltransferase (**GNAPT**). Alkylglycerone phosphate synthase (**AGPS**) then exchanges the added acyl chain for an alkyl chain, which is provided by fatty acyl-CoA reductase (**FAR1/2**). The fatty acid

element is derived from lipogenesis/ dietary sources and are first activated by acyl-CoA synthetase (ACS), which is also linked to the peroxisome membrane (Dean, J. M., & Lodhi, I. J. 2018). Acyl/alkyl-DHAP reductase, then reduces alkyl-DHAP into the ether lipid precursor 1-O-alkyl glycerol-3-phosphate (AGP). The formation of plasmalogens beyond this point is catalysed by plasmalogen ethanolamine desaturase (**PEDS1**) through generating the vinyl ether double bond characteristic of plasmalogens (Koch, J., et al., 2022).

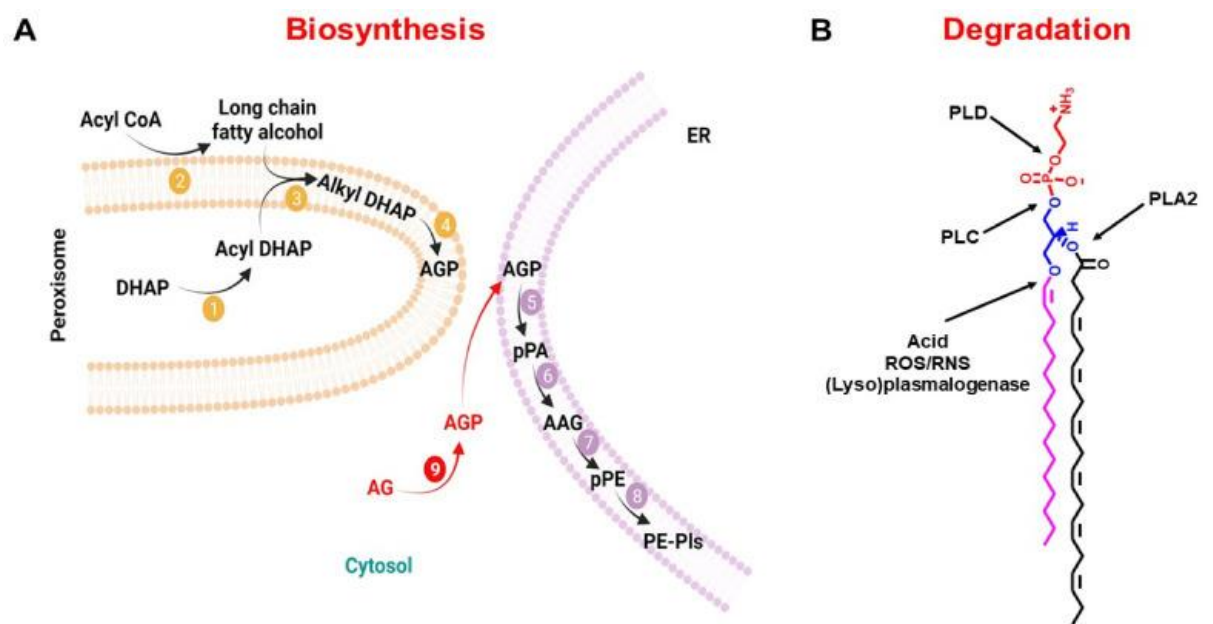


Figure 5-2- Biosynthesis and degradation pathways of plasmalogens within cells. The biosynthesis is catalysed by the enzymes **1) GNPAT 2) FAR1 3) AGPS 4) AADHAP** in the peroxisome and **5) AAG3P-AT 6) PAP-1 7) EPT and 8) PEDS1** in the endoplasmic reticulum. **9) alkyl glycerol kinase** phosphorylates **AG** which then crosses the membrane entering the endoplasmic reticulum (Bozelli J.C.JR and Eband R.M., 2021).

There is a common enzyme that links plasmalogen synthesis and PE synthesis, ethanolamine transferase indicated by 9 in Figure 5-2, which enables a systemic cross-talk controlling the balance of lipidomic presence, whereby the loss of the affected lipid for example plasmalogen will lead to the increase in the counterpart lipid as a part of homeostatic regulation.

Similarly to PE biosynthesis, there is also a secondary biosynthesis pathway that utilises the lyso-product of degradation and as such lysoplasmalogen (LPE-P) can also hold an essential role in plasmalogen synthesis acting as a precursor whilst also being

recognised as metabolism products or transient products during plasmalogen modification (Dorninger, F., et al., 2020).

### **Regulation**

The levels of plasmalogen presence are largely governed by the rates of biosynthesis and degradation that occur within the specific tissue structures. The biosynthesis pathway is controlled by a number of enzymatic reactions as shown in Figure 5-2 and occurs in the peroxisomes, but one of the main points of enzymatic regulation is through the enzyme FAR1. The method that FAR1 uses to regulate plasmalogen synthesis is through feedback inhibition, though this is not yet understood very well. There is belief that the process links to protein degradation patterns/products when plasmalogen levels are elevated (Honsho, et al., 2010).

Neurodegeneration is the resultant effect of loss of neuron structure/ function leading to neural degeneration. This can occur for various reasons and have various knock-on effects as a result of the changes seen. A lot of new research shows evidence of lipidomic profile changes holding key prevalence in various neurodegenerative disorders and therefore warrant further investigative analysis to better understand their role and seek out potential therapeutic targets directed at the lipidome (Chakraborty A., et al., 2020; Calzada, E., et al., 2016; Liu. Y., et al., 2021; Kao, Y. C., et al., 2020).

#### **5.1.2- Alzheimer's disease**

Dementia encompasses a broad range of debilitating neurological conditions that affect the way in which a person's neurones function leading to various neurological symptomatic presentations such as memory loss, motor function, cognitive and behavioural changes. Alzheimer's accounts for almost 60% of all dementia cases diagnosed in the UK (Dementia UK, 2023). The exact cause of Alzheimer's disease is poorly understood at present. The neuronal onset is associated with an accumulation of amyloid plaques due to an impaired clearance of  $\beta$ -amyloid peptides and build-up of neurofibrillary tangles. This in-turn gives rise to synaptic loss and



neurodegeneration but is unlikely the cause of all of the disease presentations observed in sufferers (Kao, Y. C., Et al., 2020).

Decreased levels of plasmalogens are commonly found in AD patients with a correlation to cognition deficit and disease severity, this could be due in part to their involvement with synaptic vesicle cycle and neurotransmissions due to principle roles in synaptic membranes and vesicle structures (Honermann T., et al., 2021 & Gu J., et al., 2022 & Moser, A. B., et al., 2011).

A difficulty in understanding the pathology of Alzheimers is a lack of clarity surrounding the exact functioning of the different classes of lipids seen to fluctuate in sufferers' brains including plasmalogens and plasmalyns. More clarity of role of all PE in neurodegenerative disorders, including which have the most significant impact on physical affiliations can drastically improve research and therapeutical advancement in this field.

This makes it critical for researchers to develop better understanding of the lipidomic complexities and aberrations that occur in neurological systems.

#### 5.1.3- *Drosophila*

Due to the complexities of the human brain, the first step to achieving a better understanding of aberrations in specific pathways is to utilise a simpler model. *Drosophila* offer such a system that is frequently utilised in such studies due not only to its many orthologues with humans which include approximately three quarters of human disease genes (Reiter, L. T., et al., 2001), but the ease with which these genes can be manipulated for study. Additional to this, *drosophila* offer a model system that can convey traceable behavioural markers mirroring human complexities such as aggression, flight patterns, learning and memory (Shin, M., & Venton, B. J., 2018). These have been found to directly mimic those found in humans for Parkinson's disease previously during a study that captured proteomic changes that occurred in diseased candidates (Xun, Z., et al., 2007). *Drosophila* also have a full central nervous system, composed of approximately 200,000 neurones, but neurologically are less

advance and so lack the presence of myelin. A key advantage to this model organisms is that, in offering a myelin-free model, drosophila can aid in narrowing down plasmalogen function in neurones without their confounding role in myelin, which is so lipid rich it may mask other more subtle changes occurring that may be missed in myelin containing model systems. Plasmalogen presence is known to first occur in gray matter before progressing to alter the composition of myelin. Drosophila may offer a unique way to better understand these prior alterations through being a simplified model. Furthermore, drosophila have a much shorter life-span and so it is faster to trace longer term system impacts and lifetime changes and the brain of drosophila have a number of conserved functional units and biochemical signatures similar to human brain models including the composition of sensory optic and antennal lobes, a mushroom body involved in memory and a central complex that operates motor output (Moloney, A., et al.,2010).

There are certain differences that need to be noted when using Drosophila as a model for neurological lipid studies. Firstly, there will be differences in the length of fatty acids observed in the brain. Drosophila lipids do not exceed eighteen carbons in length and do not conserve the ability to utilise longer chain fatty acids from their diets in the same way mammalian systems do (Shen, L. R., et al., 2010 and Rietveld, A., et al., 1999). Additional to this, there are slightly different proportions of lipidomic subclasses present in the brain of drosophila when compared to humans. 50% of lipids present in drosophila membranes are composed of PE, 25% PC and the remainder other lipids, as opposed to mammalian system whereby 50% of the total lipid content is PC (Hammad, L. A., et al., 2011). These impacts will all need to be considered when experimentally considering drosophila suitability as a model. The biggest challenge faced in this study lies in the fact that to date plasmalogen presence has not yet been confirmed in drosophila though it has been well documented in animals and anaerobic bacteria species.

Despite all of the uses of drosophila for disease study including neuronal based diseases there is still very little detail available regarding the neurological lipidome that would aid future study developments that utilise this model.

#### 5.1.4- Aims

The main aims and objectives in this chapter are to optimise a method for successfully extracting and detecting lipids present in drosophila heads in order to:

1. Determine if plasmalogen lipid species are present in drosophila head extracts
2. Evaluate which lipid species can be successfully detected and quantified in these heads
3. Determine whether different mutations in plasminogen biosynthesis pathway genes can be accurately reflected in the resulting plasminogen production  
Further determine the impact of such mutations on alterations in PE-P, PE, LPE and PE-O profiles in order to gain a better understanding of the metabolic consequences of the disruption in plasmalogen synthesis

## 5.2- Methodology

### 5.2.1- Samples

Different mutants have been utilised in the study to analyse alterations in lipidomic profiles in drosophila. Mutant and healthy drosophila fly heads were supplied from collaborator Gaynor Smith from Cardiff university generating total knockouts of the following enzymes; FAR-1, GNPAT, AGPS and PEDS-1, the biosynthetic role associated with these enzymes have been shown previously in Figure 5-2.

### 5.2.2- Extraction

Five drosophila heads were extracted using a modified Bligh Dyer lipidomic extraction whereby 5 heads were manually crushed using a micro pestle in a 1.5mL microtube before being transferred into a glass vial using four lots of 50µL of 155mM ammonium acetate buffer, ensuring all material is transferred across. The sample was then spiked with 300pmoles of PE 15:0/18:1 internal standard (purchased from Avanti lipids). This was kept on ice whilst the addition of 500µL of MeOH was performed and then vortexed for 5 minutes before adding 1mL of  $\text{CHCl}_3$  and further vortex mixing in bursts for 30 minutes with incubations kept on ice. Phase separation was allowed through the further addition of 200µL of ammonium acetate buffer and centrifugation at 3000G for 15 minutes at 7°C. The chloroform layer was collected before the remaining sample was re-extracted with a further 1mL of  $\text{CHCl}_3$  and the steps following this addition above being repeated. The two collections were pooled and dried down under  $\text{N}_2$  whilst kept on ice before being resuspended in 100µL of 2:1 ( $\text{CHCl}_3$ :MeOH). The extract was stored at -20°C until analysis.

For the analysis, samples were briefly centrifuged to ensure no debris remained and a stable spray could be achieved. 30µL of extract and 60µL of 0.2% DEA (in 5:1 MeOH: $\text{CHCl}_3$ ) were loaded onto 96-well plate for direct infusion via orbitrap analysis. A Thermo scientific LTQ Orbitrap XL with a Triversa NanoMate(Advion) was used for injections with a plate temperature set to 6°C ensuring lipid preservation without condensation. Eight microlitres of sample was injected as spray for analysis. The API Source was set to a 0.01kV Source voltage, with a vaporizer temperature of -60.74°C,

capillary voltage of -23.89V and capillary temperature of 200°C. The Ion optics were set to achieve a lens voltage of 5.16V, a gate lens of 34.31, Front lens of 5.52V and back lens of 0.17V. The heated ESI source had a heater temperature of 150°C, I Spray voltage of 4kV, capillary temperature of 275°C, capillary voltage of -9V and a tube lens of 150V. The MS1 data was obtained using FTMS mode over a full 10 minute run whereas identification data, both MS2 and MS3, was obtained instead using ITMS mode.

Standard curves were generated to check the linearity of different PE species that needed to be quantified during the analysis. Lipid standards purchased from Avanti lipids were used to generate triplicate desired stock concentrations between 0.1 and 10µM solutions before the addition of ISTD, 300pmoles of PE 15:0/18:1. These were then extracted and analysed using the same methodology as drosophila samples, see section 5.2.2- Extraction.

### 5.2.3- Data analysis

Following analysis, spectra were extracted using Xcalibur Freestyle software and then statistically analysed using Microsoft Excel whereby peaks were first filtered out and excluded if the signal/noise ratio fell below 2. The closest peaks' relative intensities to the PE and PE-O of interest were then obtained and these underwent isotopic corrections before being normalised according to the internal standard. During this process if any data points achieved a signal: noise <2 they were denoted as having a zero intensity during the comparison. This caused some lipids that were identified such as PE-O(18:1/18:0) to be excluded from further analysis. ANOVA was used for the statistical analysis of this data set, which was conducted on SPSS with Tukey's method used for post-hoc analysis.

## 5.3- Results

### 5.3.1- Standard curves

The standard curves showed a very good linearity, with low level concentrations still falling on the standard curve, and reproducibility, illustrated by  $R^2$  values of 0.99 being achieved in both instances Figure 5-3.

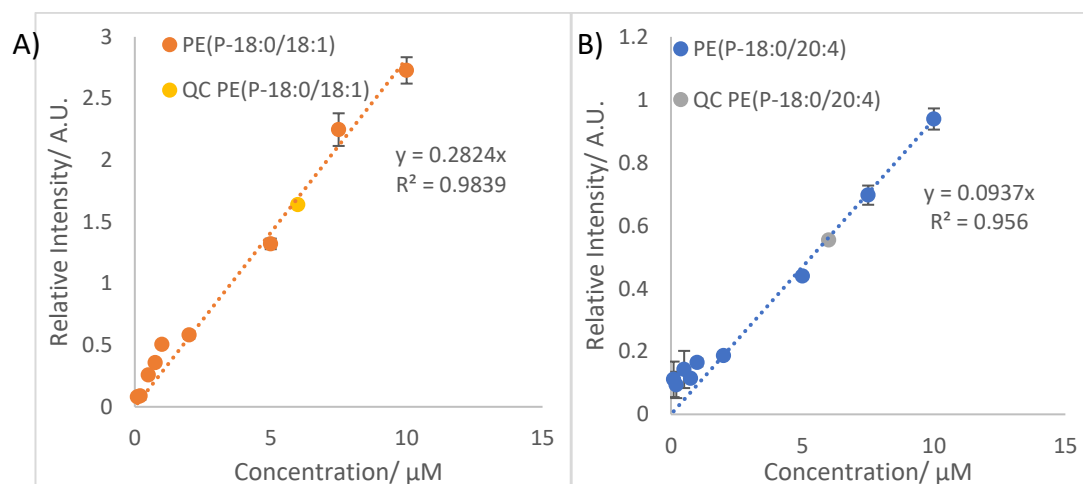


Figure 5-3- A) Plasmalogen (PE-P) and B) phosphatidylethanolamine (PE) standard curves showing linearity and good reproducibility for use in quantification of lipid classes. Standards have been normalised to the ISTD and have been extracted in the same way as samples were prepared. The PE-P standards were checked with the separate preparation of a quality control samples (QC) at a concentration of  $6\mu$ M to ensure that this could be correctly quantified from the graphs.

The next step in the standard curve evaluation was to assess the suitability of this methodology for the quantification of lipids present in the sample set. This involved a further critique of the reproducibility of the extraction set using this methodology. This was done through plotting the replicate samples against each other, Figure 5-4, with straightness of the line and  $R^2$  values closest to 1 noted as depicting a good reproducibility. These evaluations showed a good level of reproducibility was achievable using the extraction methodology being implemented in this study and so the sample set was evaluated using the same methodology.

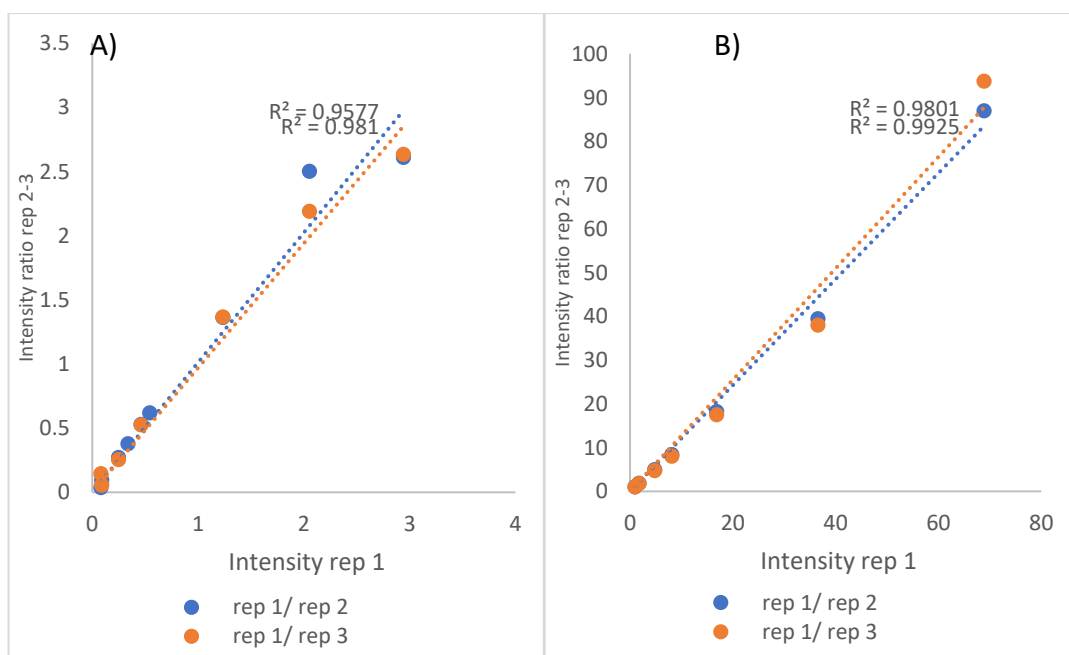


Figure 5-4- Plot showing a reproducibility assessment for the quantification of A) plasmalogen and B) phosphatidylethanolamine using the standard curves generated in Figure 5-7. This was conducted through taking the values of the first replicate set and dividing by the second or third replicate set and plotting the values generated were then divided from the values of replicate 1 sample set. The strong values of  $R^2$  generated the good level of reproducibility achievable using this range and methodology.

Upon acquisition of sample data (n=3 replicates for each mutant sample set) the range of the standard curve was evaluated against the PE lipid intensities under analysis to evaluate how they compared to the dynamic range of the curve. Figure 5-5 shows the majority of the samples PE classes fall within the working range of the standard curves generated with only a small number of lipids dropping below this range, namely PE(O-16:0/18:3), PE(O-16:0/18:2), PE(18:0/18:1), PE(16:0/20:0) and PE(19:0/20:5). No lipids exceeded this calibration range.

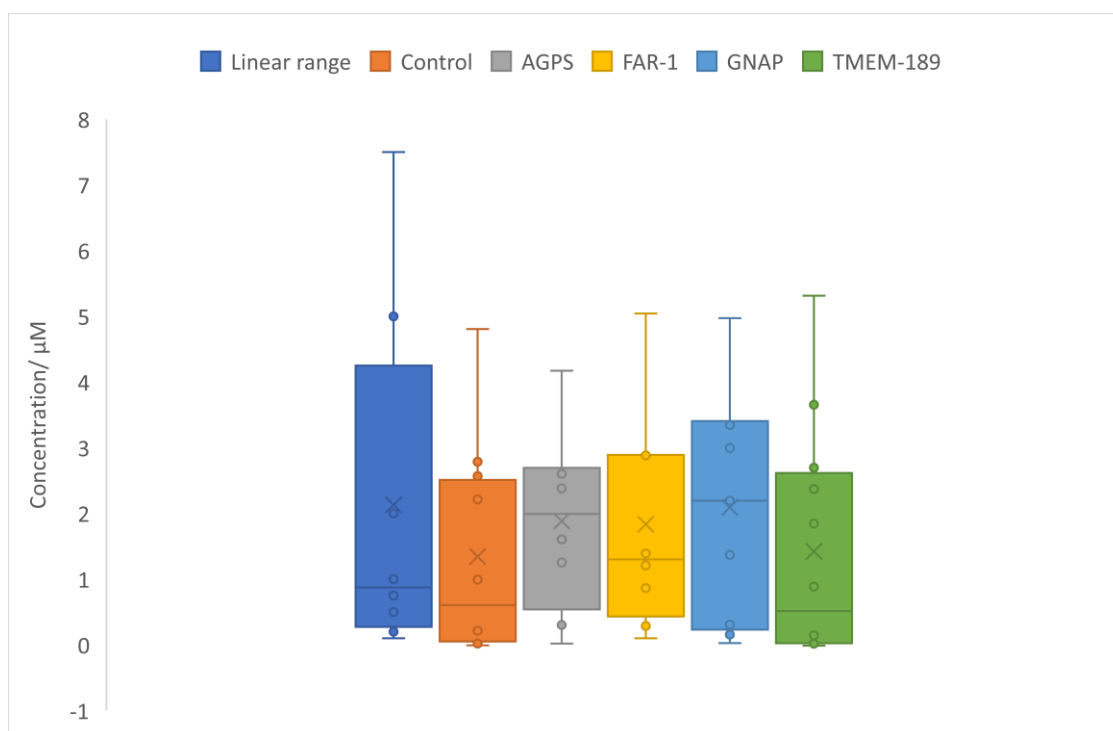


Figure 5-5- Box and whisker plot demonstrating the linear range of each mutant extract vs the dynamic range of the standard PE(16:1/16:1) with the majority of the detected phosphatidylethanolamine lipids falling within the working range of the standard curve but some dropping below, namely PE(O-16:0/18:3), PE(O-16:0/18:2), PE(18:0/18:1), PE(16:0/20:0) and PE(19:0/20:5)

### 5.3.2- Biological Extracts

After determining the methodology and corresponding dynamic ranges that were to be utilised for sample analysis A series of fly heads were extracted to confirm which lipidomic species could be qualitatively identified during the analysis with confirmations made through analysing fragmentation patterns generated.



A number of phosphatidylethanolamine (PE) and Plasmalogen (PE-P) and Plasmalogen (PE-P) and Plasmalogen (PE-P) as well as LPE and LPE-O/P were successfully identified using MS2 and MS3 fragmentation patterns. Additionally, a number of PI and PS groups were also identified and studied by semi-quantitative analysis as shown in Table 5-1.

Figure 5-6 below is an example of the identifying fragmentation spectral patterns acquired for a PE class of lipid, PE(34:1) identified as PE(16:0/18:1),  $C_{39}H_{76}NO_8P$ . This is a common lipid found in previous studies to be evenly distributed across the fly

brain and was similarly found to be one of the more dominant lipids in the extracts analysed during this study (Phan, N. T. N., et al., 2017).

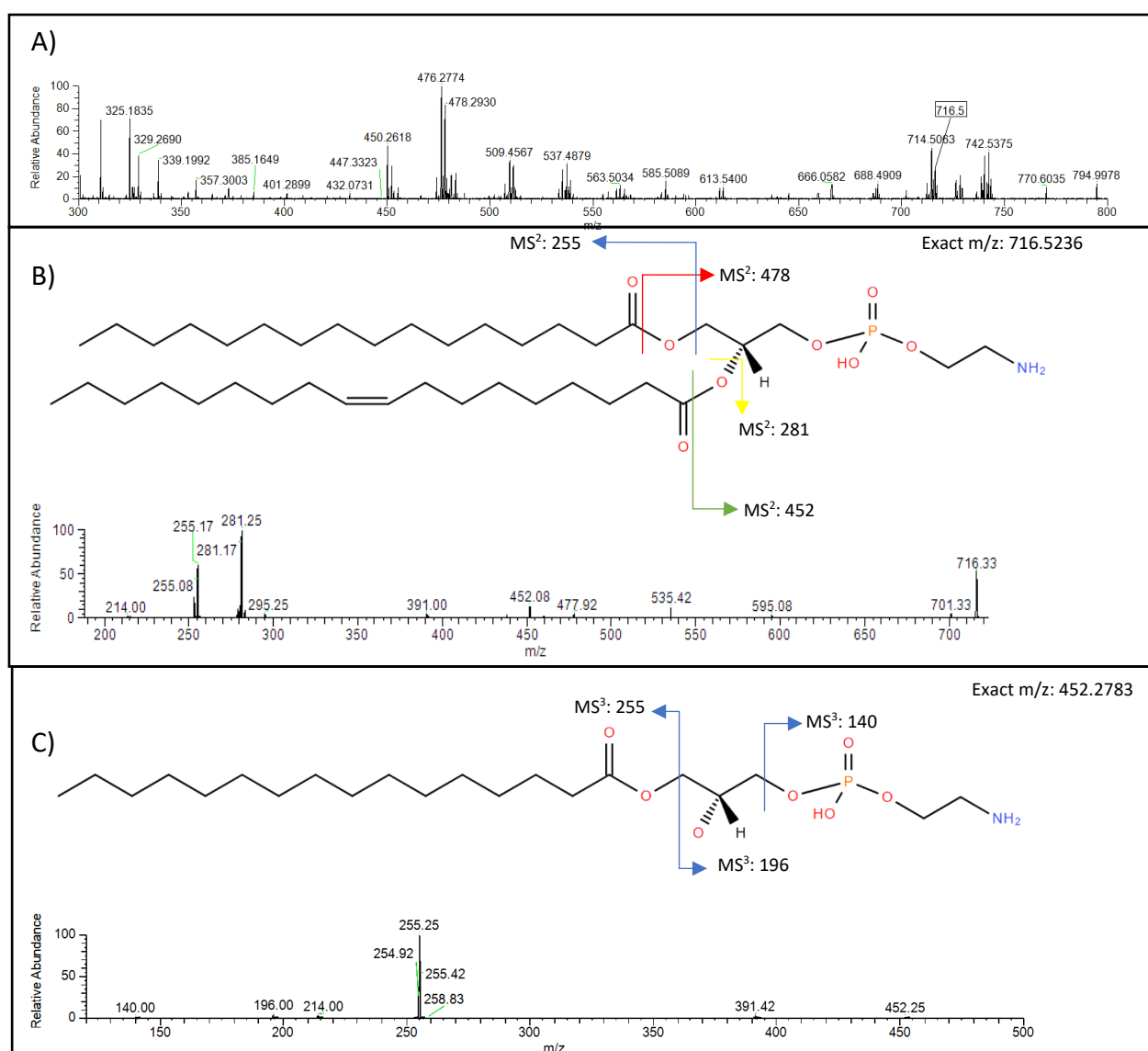


Figure 5-6- Fragmentation patterns obtained for ion 716.5 illustrating the key fragmentation patterns associated with lipid PE(16:0/18:1). A) Illustrates the full MS spectra obtained for a healthy drosophila head extract with the peak of interest highlighted. B) MS2 structural fragmentation seen for PE(16:0/18:1); red 478.33 resultant of the loss of the sn-1 C16:0 acyl chain as ketene. 452.17 is shown in green which is resultant of the loss of C18:1 sn2 as ketene. 434.25 was the loss of sn2 RCOOH group shown. 281.25 is reflecting of C18:1 shown in yellow, 255.25 is C16:0 shown in blue and finally 253.2 is reflecting of C16:1 similarly shown in blue but with one less H present. MS2 spectra obtained from peak 716.5 generated with corresponding peaks to those expected in B. C)

MS-2 fragmentation pattern anticipated for  $m/z$  716.5, 452.5. E) MS-3 fragmentation seen for peak 716.5  $\rightarrow$  452.5.

The MS2 fragmentation showed spectral peaks at 478.33, 452.17, 434.25, 281.25, 281.3, 255.25, 255.2 and 253.2. These fragmentation peaks give key identifiable markers for the lipid PE(16:0/18:1) as shown in Figure 5-6. Upon conducting ms3 on fragment ion 452.17 we see additional confirmation of the presence of PE through the fragmentation pattern of 255.25 which is the previously seen C16:0. Further peaks at 140.0 and 196 correspond to the PE headgroup  $C_2H_{29}O_2^-$  which is ethanolamine phosphate ion.

PE and PE-O can be identified through a different fragmentation profile signatures to the previously discussed. A peak at  $m/z$  726.5 was identified as a major lipid PE(O-18:0/18:3) with fragmentation pattern shown in Figure 5-7.

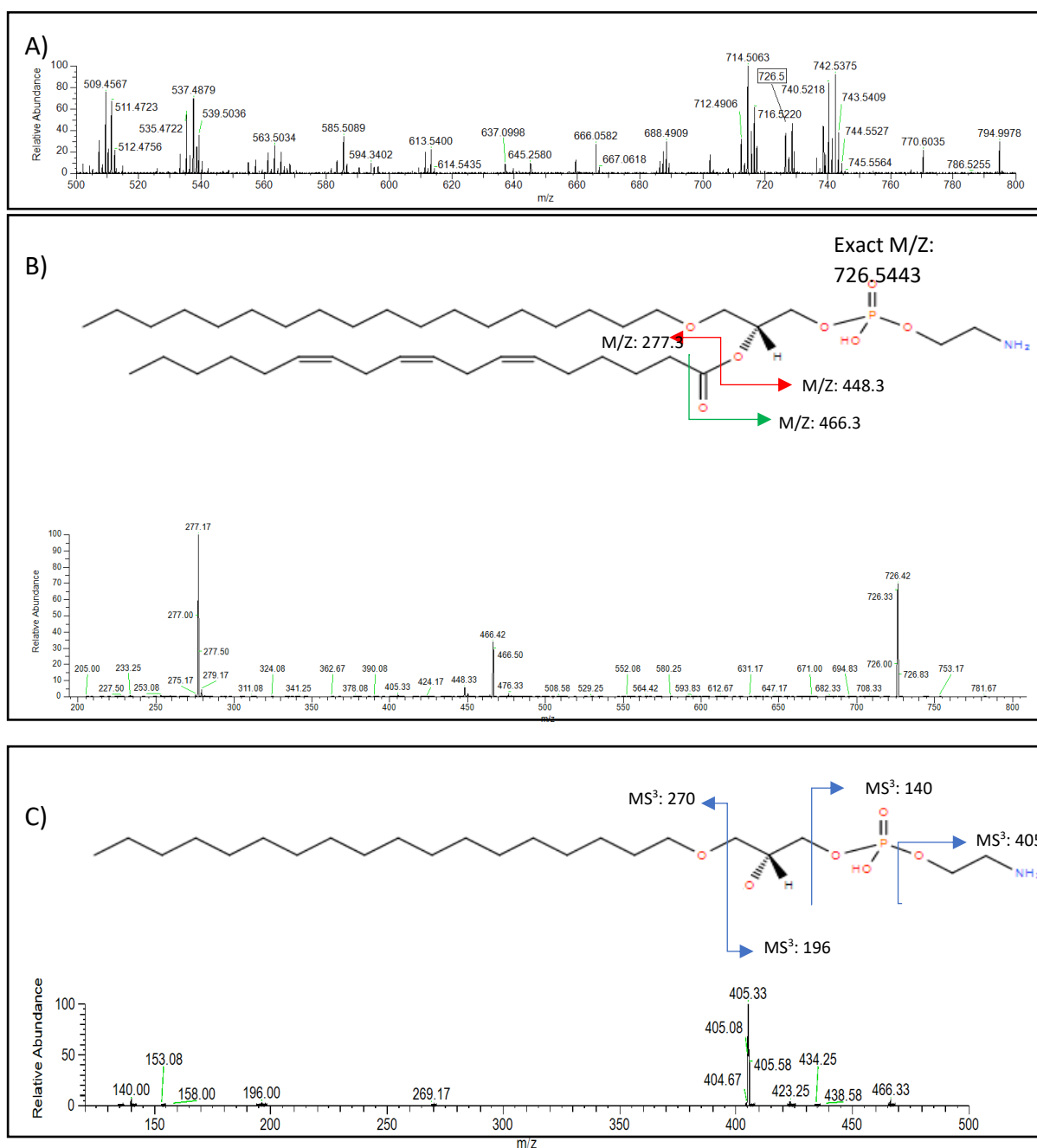


Figure 5-7- Fragmentation pattern of ion 726.5 A) Shows the full spectra with  $m/z$  peak, 726.5, of interest highlighted. B) Shows MS2 fragmentation pattern and structural details seen 466.3 seen in green with the loss of sn2 acyl chain, shown in red is  $m/z$  448.3 and 277 due to loss of sn2 RCOOH and sn2 RCOO ion respectively. C) Shows the MS3 fragmentation structure and spectra obtained from  $m/z$  726.5>466.5 generating identifying fragmentation patterns seen at and 140 loss of ethanolamine phosphate, 196 and 270 and 153 which shows the fragmentation product Glycerol-3-phosphate ion with loss of water.

The prominent lipid present here are shown in Figure 5-7B with identifying fragmentation patterns seen associated with the loss of SN2 acyl chain as a ketene at

peak 466.5, 277 showing the fragment SN2 group and 448.3 is attributed to the loss of sn2 RCOOH group. The MS3 fragmentation of product ion 466.5 gave further key fragment markers at 405.3, 269, 196 and 140 as seen in Figure 5-7C.

Alongside this more abundant lipid species, plasmalogen lipid signals were sought with a minor plasmalogen peak present also noted at  $m/z$  726.5. This was identified as being PE(P-18:0/18:2). The fragmentation pattern for this can be seen in Figure 5-8. This shows distinctly different fragmentation from the prior 726.5 PE(O-18:0/18:3). The key and most prominent fragmentation difference arises at 464.5 rather than 466.5 in addition to 476. The MS3 fragmentation then depicts the different plasmalogen structure with the 267 fragmentation rather than 270 due to the presence of an additional c-c double bond.

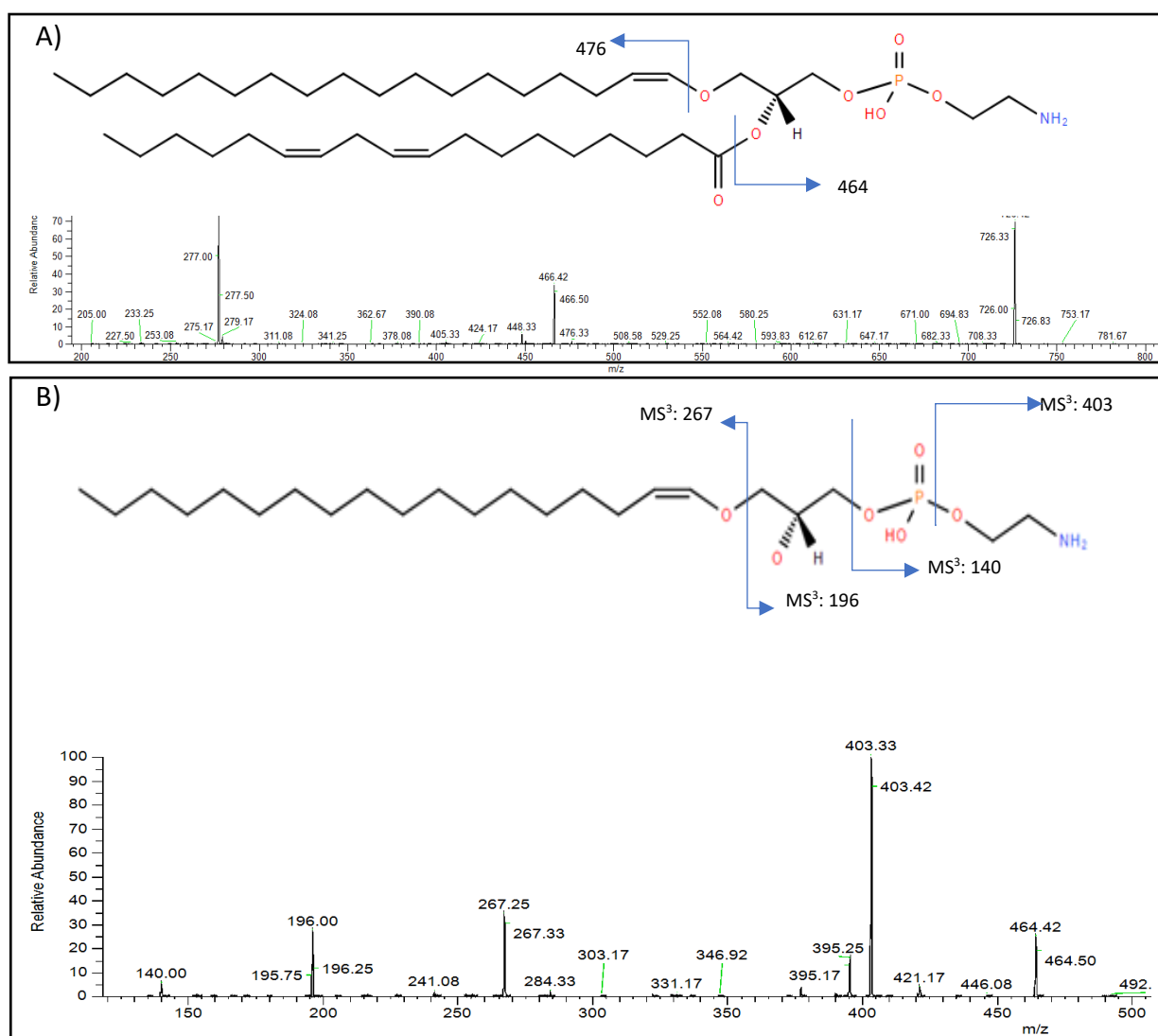


Figure 5-8- A) Shows the fragmentation pattern obtained from the peak at 726.5 with the PE-P(18:1/18:1) plasmalogen fragmentation pattern shown structurally with peaks seen at 476 and 464. This is further evidenced in the B) MS-3 fragmentation through its presence of ions at 403.3 (further fragmentation from the phosphate group), 267 (loss of the SN chain at the oxygen ion) 196 and 140 (indicative of fragmentation at sn-1 1-o-alk-1'-enyl residue as an alcohol). 464 & 446 loss of 18:1 group as acid and ketene respectively.

Table 5-1 below shows all of lipids utilised in this study that were identified using orbitrap analysis. The classification, identity and fragmentation pattern are listed for each ion identified. In instances when more than one product ion was present, such as the previously discussed 726.5, the most dominant peak identity was utilised.

Table 5-1- Lipids Identified and used during the study with *m/z*, subclass, assignment and fragmentation pattern seen listed

Accurate <i>m/z</i> [M-H]	Sub- class	Assignment	Fragment ions ms2	Fragment ions ms3
476.285 542	LPE	PE(18:2/0:0)	458.3/ 432.3/ 415.3/ 388.2/ 348.3/ 325.3/ 292/ 279.5/ 253.3/ 233/214.2/196/153/140	261.3/ 243.2/ 233/ 205.3/ 177/ 165/ 151/ 133/ 97/ 83
478.301 192	LPE	PE(18:1/0:0)	478.3/ 417.3/ 406.3/ 388.4/ 380.4/ 372.3/ 327.4/ 281.3/ 267.2/ 214/ 196/ 174.8/ 153/ 140	281.3/ 263.3/ 245.3/ 235/ 219/ 197/ 193/ 183/ 169/ 149.3/ 141.2/ 127/ 111/ 97
465.293 9	LPE	PE(17:0)	465.4/ 447.3/ 421.4/ 403.3/ 385.4/ 297.3/ 281.3/ 268.3/ 255.3/ 239.2/ 215.2/ 195/ 187/ 177/ 167/ 159/ 157.3/ 144.8/ 142	
450.262 6	LPE	PE(16:1)	450.4/ 388.4/ 358.3/ 299.3/ 297.2/ 279.4/ 253.4/ 227.3/ 214/ 196/ 185/ 171/ 162/ 153/ 140/ 130	
452.278 3	LPE	PE(16:0)	452.5/ 434.3/ 408.3/ 390.4/ 378.3/ 349.4/ 325.3/ 311.3/ 282.2/ 255.7/ 242.3/ 226.3/ 214/ 196/ 181/ 171/ 153/ 140/ 128	255.3/ 237.3/ 225.3/ 97/ 83
453.4	LPE	PE(16:0)	453.3/ 435.5/ 409.4/ 391.4/ 379.3/ 348.2/ 325.3/ 297/ 283/ 256/ 237/ 209/ 196/ 187/ 167/ 153/ 145/ 139/ 130	
435	LPE- P LPA	PE(P-16:0) LPA(18:1)	435.4/ 417/ 391/ 373/ 355/ 349/ 321/ 305/ 283/ 271/ 255/ 226/ 217/ 208/ 187/ 171/ 157/ 153/ 145/ 135/ 124	
439.4 440.278 3	LPE	LPE(O-14:0)	439.4/ 421/ 395/ 381/ 377/ 359/ 345/ 311/ 269/ 242/ 215/ 195/ 187/ 173/ 153/ 173/ 153/ 145/ 135/ 125	
449.5 450.269 9	LPE	PE(16:1)	449.5/ 431.3/ 405.3/ 389.5/ 387/ 369/ 357/ 347/ 333/ 317/ 281/ 253/ 237/ 209/ 199/ 187/ 171/ 162/ 153/ 145/ 128	

461.3 462.269 9	LPE	PE(17:2)	461.4/ 443.3/ 425/ 417/ 399/ 371/ 367/ 347/ 333/ 279/ 265/ 251/ 195/ 185/ 167/ 153/ 145/ 137/ 129	
476.571 7	LPE	PE(18:2)	476.5/ 415/ 402/ 360/ 325/ 318/ 302/ 279 / 261/ 253/ 242/ 222/ 214/ 205/ 196/ 189/ 171/ 160/ 153/ 140	
698.520 3	PE-O	PE(O- 16:0/18:3)	680.42/654.25/528.42/438.33/277.25	
700.535 9	PE-O	PE(O- 16:0/18:2)	678.17/656.58/593.5/568.08/512.5/478.08 /450.33/438.33/420.25/353.08/279.25/253 /240.83	438.42/377.08
702.515 2	ISTD PE	PE(15:0/18:1)	630.5/478.42/466.3/438.25/420.25/340.05 /281.25/253.17/241.25	241.08
716.530 9	PE	PE(16:0/18:1)	700.25/678.42/655./636.25/596.2580.5/47 8.33/452.17/434.25/418.42/384.33/372.17 /309.42/281.33/281.25/255.25	452.25/391.42/255. 25/214.0/196.0/14 0.0
718.546 5	PE	PE(18:0/16:0)	700.17/682.33/660.42/630.33/602.25/563. 25/499.08/480.42/453.25/434.17/390.33/3 11.17/297.25/282.33/277.17/255.25/241.3 3	480.58/283.5/25.25
726.551 6  726.551 6	PE-O  PE-P	PE(O- 18:0/18:3)  PE(P- 18:1/18:1) PE(P- 18:0/18:2)	726.42/466.33/448.33/405.33/277.25/233. 25  726.5/462/401/265/196/140 (minor) 726.5/464/403/267/196/139 (minor)	405.33/269.25/196 /140
728.567 2	PE-O  PE-P	PE(O- 18:0/18:2)  PE(P- 18:0/18:1)	466.33/448.33/279.23/210.97/  464/403/267/196 (minor)	405.33/269.25/255. 33/227.0/196/155. 58/140
734.483 9	PE	PE(18:3/18:3)	734.5/690.6/474.3/277.3	233.1/214/196/140



736.499 6	PE	PE(18:2/18:3)	736.7/718/651.2/474.1/458/279.3/277.3/2 77	
738.499 6	PE	PE(18:2/18:2)	738.4/738.2/476.3/279.3/279.1/277.25	
742.546 5	PE	PE(18:1/18:1)	742.5/480.42/476.3/460.25/439.83/417/39 8/373/322/307.5/286/281/279.3/277.3/25 5/214.25	480.25/450.08/419. 17/283.25/282/255 .25/196
744.562 2	PE	PE(18:0/18:1)	744.5/712.2/696.5/630.2/606.9/508.4/480. 3/463.2/435.3/411.3/311.3/283.3/281.3/27 9.3/261.5	
746.577 8	PE	PE(16:0/20:0)	688.5/508.3/490.3/465.3/440.2/392.3/311. 3/279.3/255.3/253.3	
750.544 3	PE-P	PE(P- 18:0/20:4)	750.5/464/403/267/196	
770.570 5	PE	PE(16:1/22:1)	763.1/738/726.3/696/648/542.9/516/478/ 437.3/418.2/389.3/297.3/281.3/253.3	
778.546 5	PE	PE(19:0/20:5)	777.8/746.2/718.5/691.3/623.7/562.6/522/ 518.3/500.3/498/476.3/442.3/431/413.2/3 78.3/374.2/320.3/297.9/277.2/253.5/236	
782.505 0	782. 5	PS(18:2/18:2)	782.5/782.3/695.3/654.6/616.5/597.3/570. 8/542.6/527.3/507.9/476.5/415.3/384.2/35 5.2/309.3/279.3/238	
784.520 7	PS	PS(18:1/18:2)	784.5/740.4/726.5/697.5/667.6/633/615.5/ 586/567.5/540/506/478/435.5/417.2/356/ 335/311.2/277.3	
786.563 4	PS	PS(18:1/18:1)	786.7/ 728.3/ 699.5/ 689.4/ 654.2/ 581.5/ 531.3/ 519.2/ 466.6/ 417.3/ 393/ 337.5/ 311.3/ 279.5/ 241	
859.541 5	PI	PI(18:2/18:1)	859.5/697.5/597.5/595.5/579.3/435.3/417. 3/297.3/281.3/241.2	
805.494 5	PI	PI(16:1/16:1)	805.5/643.5/551.5/389.3/297.2	
807.510 2	PI	PI(16:1/16:0)	807.5/571.3/553.3/551.3/391.3/389.3/315/ 297/255.3/253.3/241	

831.510 2	PI	PI(18:2/16:1)	831.5/ 668.5/ 595.3/ 577.3/ 551.3/ 415.3/ 389.2/ 315/ 297/ 253/241	
833.525 8	PI	PI(16:0/18:2)	833.5/ 681.8/659.3/597/ 579.4/ 553.3/ 433.3/ 315.3/ 391.3/ 315/ 255.3/ 223	
835.541 5	PI	PI(16:0/18:1)	835.6/ 803/ 778/ 736/ 682/ 649/ 597/ 579/ 553/ 505/ 485/ 417/ 391/ 297/ 281/ 255	
857.525 8	PI	PI(18:1/18:3)	857.5/ 813.3/ 777.4/ 740.8/ 695.3/ 579.3/ 552/ 505/ 473.5/ 417.4/ 415.3/ 387.4/ 297.2/ 281.3/ 241	
861.557 1	PI	PI(18:2/18:0)	861.5/ 843.2/ 743.3/ 669.3/ 599.3/ 581.3/ 556.2/ 419.3/ 391.8/ 334.3/ 283.3/ 241	

The majority of PE lipids of interest in the study were seen as individual lipidomic profiles with clear fragmentation patterns seen, which enabled the absolute quantification with use of standards. For those that had more than one lipid present, the dominant lipid was taken as being the class for quantification.

#### 5.3.1- Plasmalogens

During this study no plasmalogen lipids were able to be specifically quantified due to dominant occluding lipids or the presence of other species. Plasmalogen PE-(P-18:0/20:4) seen at ion 750.5 is an example of this, which showed additional fragmentation corresponding to an unknown species, which was an isotope of 749.5. This unknown species gave a fragmentation pattern of ions 534.5, 479.3, 282.5 and 253.25 as shown in Figure 5-9.

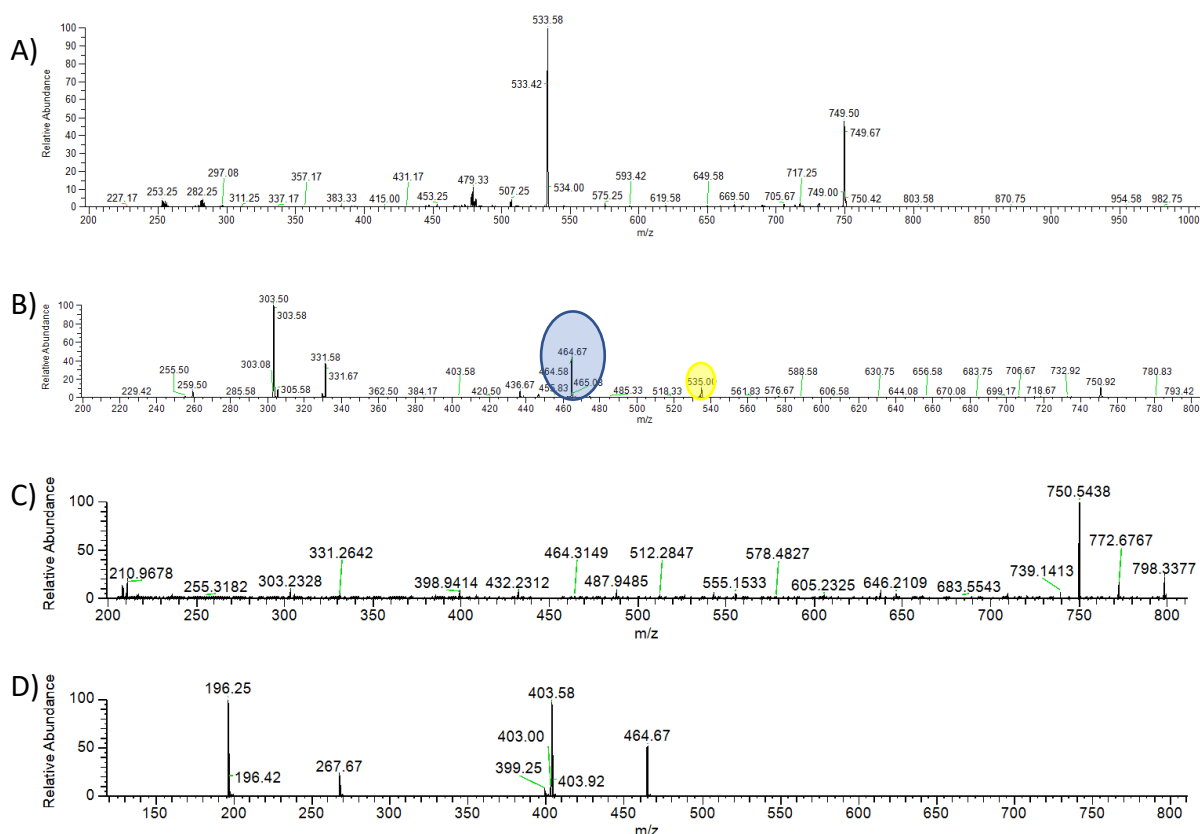


Figure 5-9- Seeking proof of plasmalogen presence using peak 750.5. A) Shows the peak at 749.5, which generates fragmentation patterns of 534, 479, 282 and 253. This peak has an isotope that is also present at 750.5. B) Shows the full MS-2 fragmentation pattern generated at peak 750.5. The blue circle highlights a principle plasmalogen related fragment ion 464, the yellow circle highlights the isotopic peak generated from the previously discussed 749.5 with principal fragment ion 535 generated. C) Shows another example whereby the ratio of the two fragment ions has changed with the 749.5 isotope now being the more prominent. D) Illustrates the correct ms-3 fragmentation of the 750.5/464.5 plasmalogen lipid present confirming the presence of plasmalogen PE-P 18:0/18:1. This fragmentation pattern is completely unlike that of the isotopes.

The presence of this unknown species fluctuated between identical replicate samples when compared to the plasmalogen fragmentation key peak of 464 and as such could not be considered a constant to enable relative abundance comparisons to be made either.

Despite the lack of ability to quantify the presence of PE(P-18:0/18:1) for the reasons discussed, there was clear evidence of the plasmalogen reducing to the point of complete absence in mutant strains.

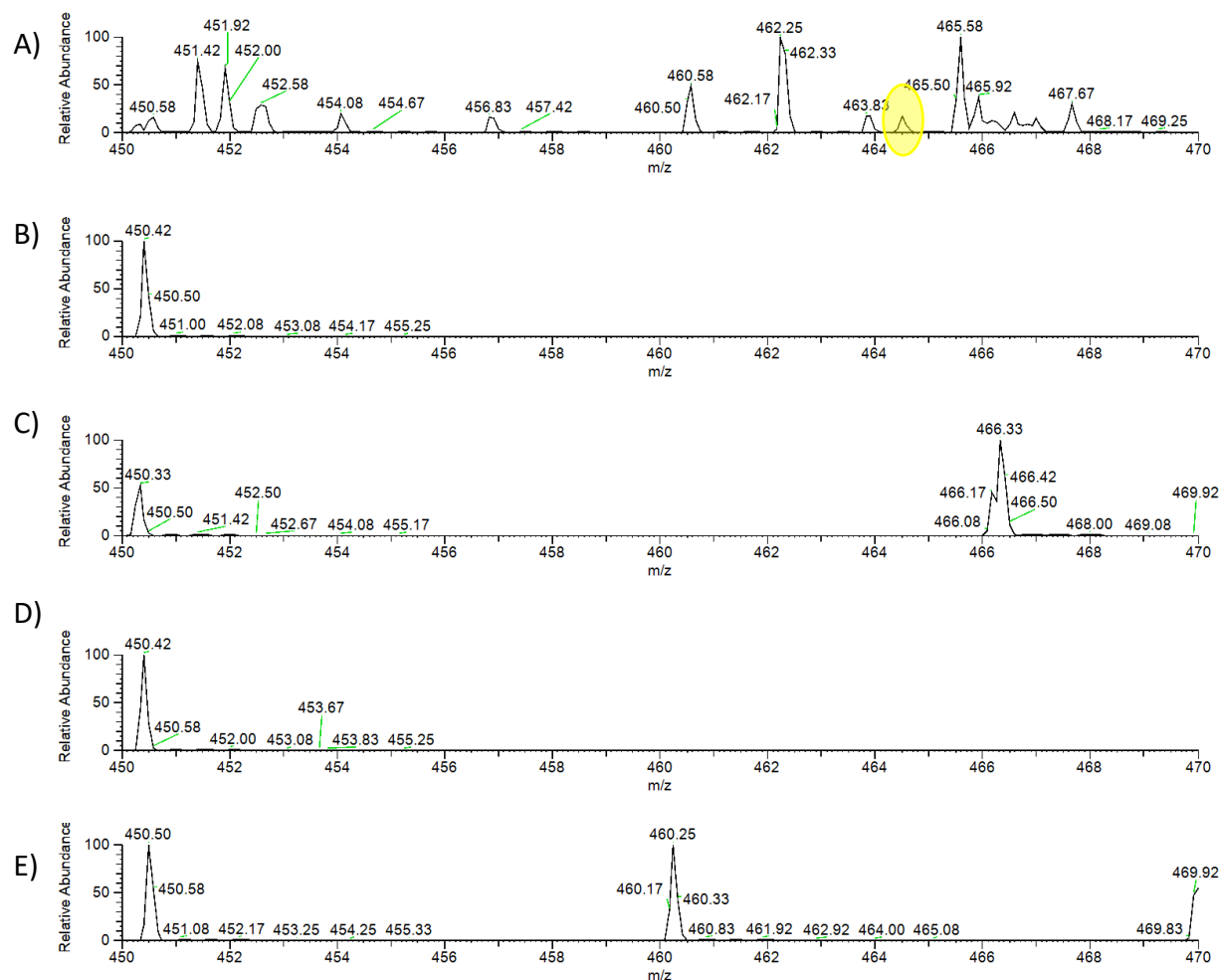


Figure 5-10- MS-2 spectra generated for ion peak 750.5 with focus on the region between 450 and 470 for samples A) Normal drosophila brain extract showing the presence of plasmalogen through the presence of fragment ion 464.5 B) Mutant extract AGPS C) GNAP D) FAR and E) PEDS1 with all mutant strains (B-E) showing complete loss of plasmalogen with no detectable 464.5 fragment ion.

### 5.3.3- Mutant flies

After establishing the lipidomic analysis scope to be undertaken from biological extracts, the methodology was applied to mutant fly extracts, whereby the lipid profiles were analysed following various enzymatic knock-downs impacting the plasmalogen biosynthesis pathway.

Data was acquired to look for lipidomic comparisons between control and mutant fly extracts FAR1, AGPS, TMEM-189 and GNAP's in order to compare the lipidomic profiles. This was done with a focus on EtnGpl's PE and plasmalogen impacts caused by the mutations whilst also considering PS and PI presence. Figure 5-11 shows example spectra obtained for each of the different mutant strains in one of the regions of interest, selected for its high presence of the PE groups between  $m/z$  700 and 800.

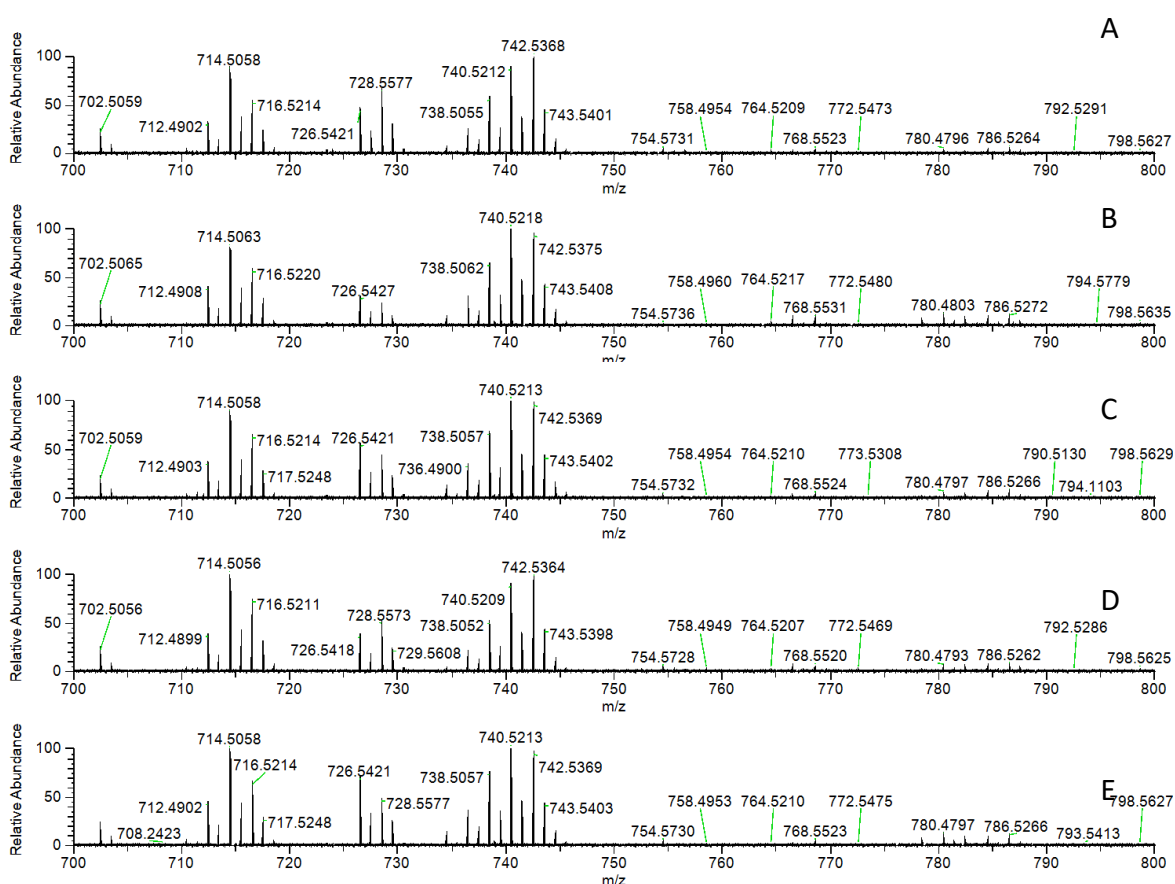


Figure 5-11- example lipidomic chromatogram obtained from A) healthy control, B) FAR1, C) AGPS, D) TMEM189 and E) GNAP mutant drosophila heads over the region of  $m/z$  700-800 focusing on the region which is indicative of the presence of PE, PE-O and PE-P EtnGPL.

When evaluating all of the GpEtn lipidomic species present in the fly heads using a heatmap analysis, a good variation of lipidome presence was seen which can be used to analyse how the different mutations relate to each other and emphasise any key differences observed between the groups as seen in Figure 5-25.

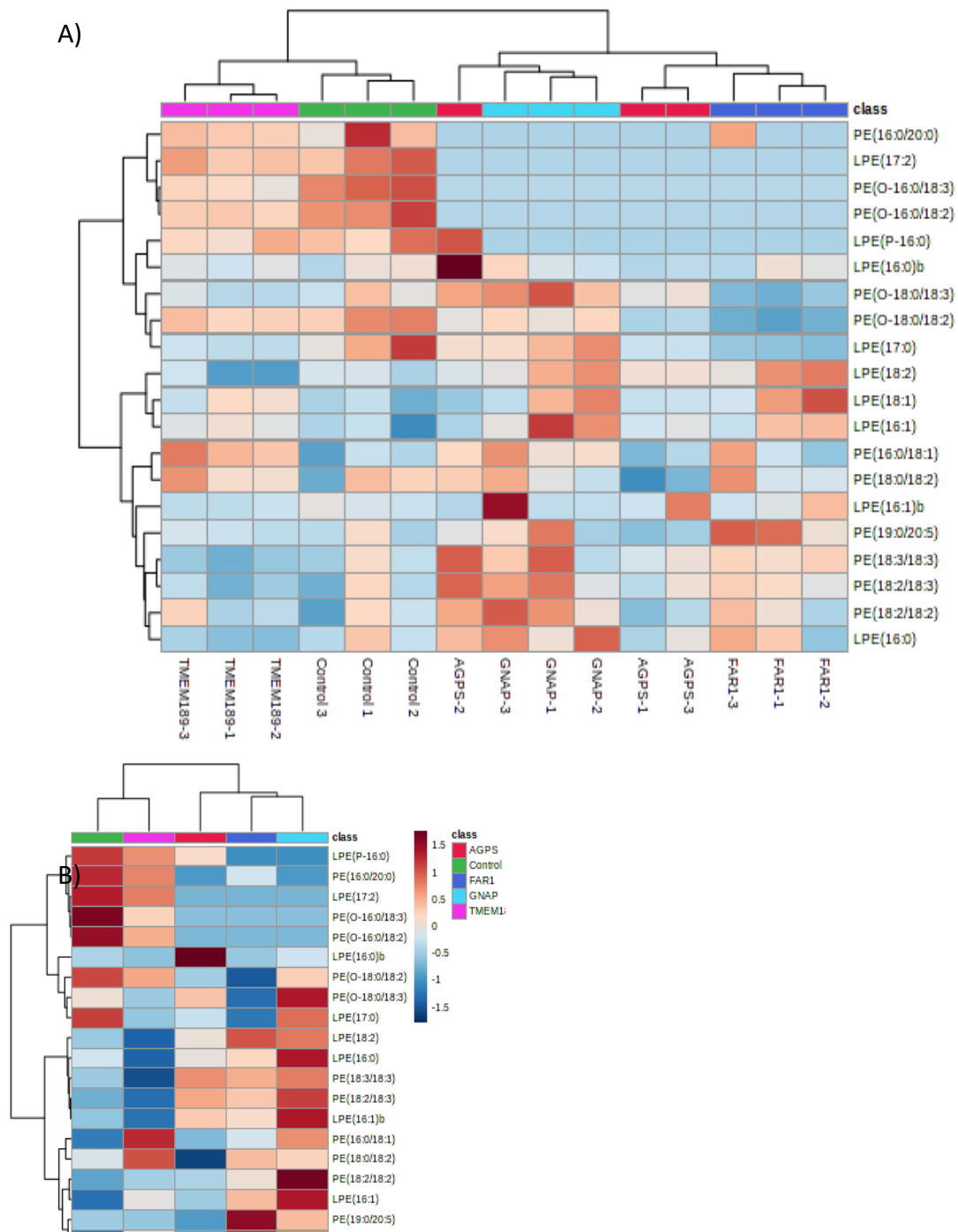


Figure 5-12- heatmap showing the identified lipids relative abundance changes between each of the different mutant strains samples and overall how these then relate to each other in terms of pattern similarity. A) demonstrates the individual samples alterations in lipidomic profiles whereas B) shows the groups average changes. All of the groups, with exception of one replicate in AGPS mutation are seen to hold very similar profile changes, noted by there close proximity to each other. The TMEM-189 (pink) and the control group (green) had the most similar profiles to each other coming from the same primary branch, whereas the other three mutant strains differed more from the control group with FAR-1 (dark blue) and AGPS (red) bearing more similarity than GNAP (turquoise).

The trend shows that TMEM-189 is the most similar to the control group. With the other three mutational groups branching off separately together, this can be expected with this mutation being the latest to occur in the plasmalogen synthesis pathway, only impacting the final step and not the prior use of PE-O, unlike the other markers that occurred much earlier on. Of those FAR-1 and GNAP are the next most similar to each other with AGPS being the most unique amongst the class. This is shown by visually greater variance in lipidomic concentrations being seen across all the subclasses under evaluation. This similarity trait is slightly altered when considering only the most varied changes via a PCA plot, , in this instance the control group has the least separation from AGPS and GNAP followed by FAR-1 and TMEM-189. There is only a small level of variation seen between the experimental classes, this is to be expected due to them being mutations that arise at different points of the same biological pathway.



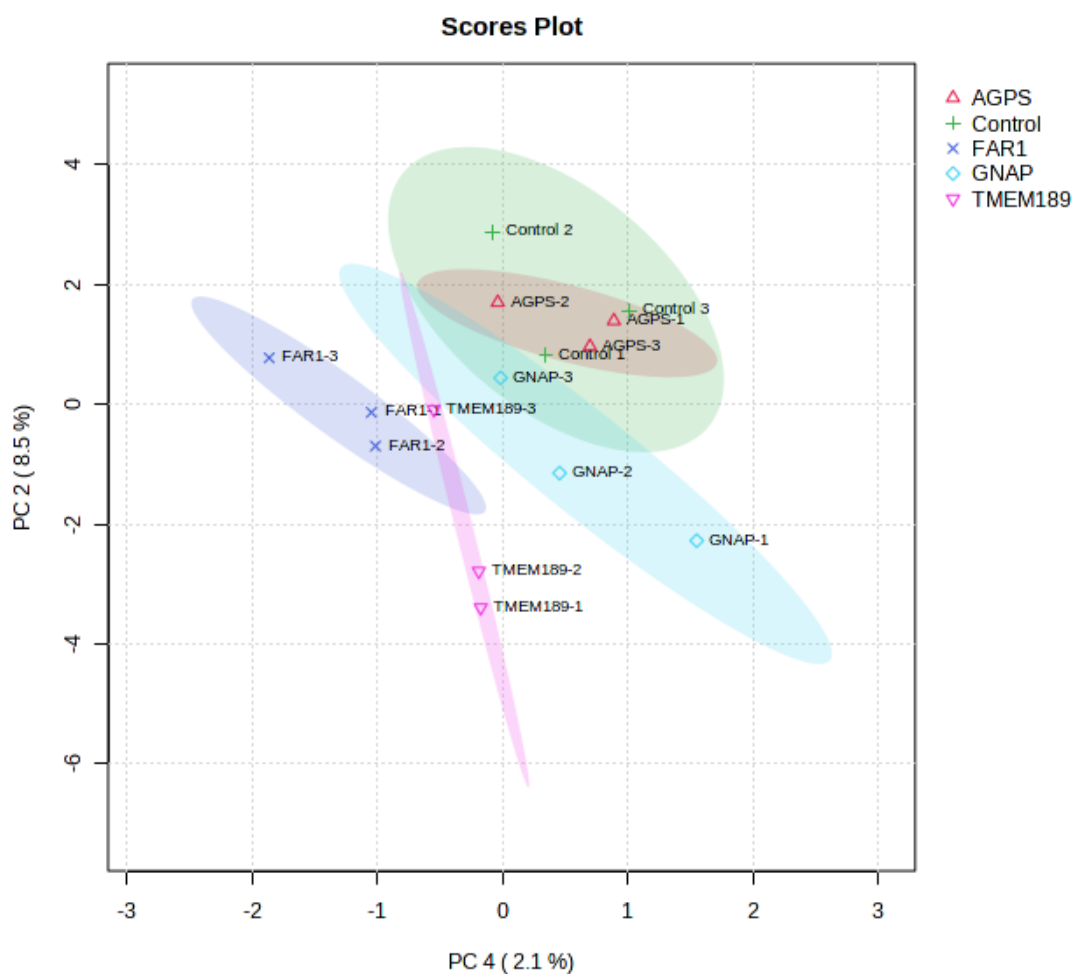


Figure 5-13- PCA analysis of all groups considering all three lipidomic subclasses identified in the study. FAR-1 showed the greatest variation to the control group, followed by TMEM-189, GNAP and AGPS.

When focusing on the variation of just the PE and PakEtn species alone, a better group dissociation is seen. AGPS has a general overall lower presence of all the groups, whereas GNAP has the opposite though this is largely due to the higher proportion of PE present. Similar to when all of the groups were included in the analysis, of all of the mutant strains included, TMEM-189 holds the greatest similarity in lipidomic profile changes compared to the control group. GNAP and AGPS are then the most similar of the remaining mutants with FAR-1 being the most dissimilar group, Figure 5-26.

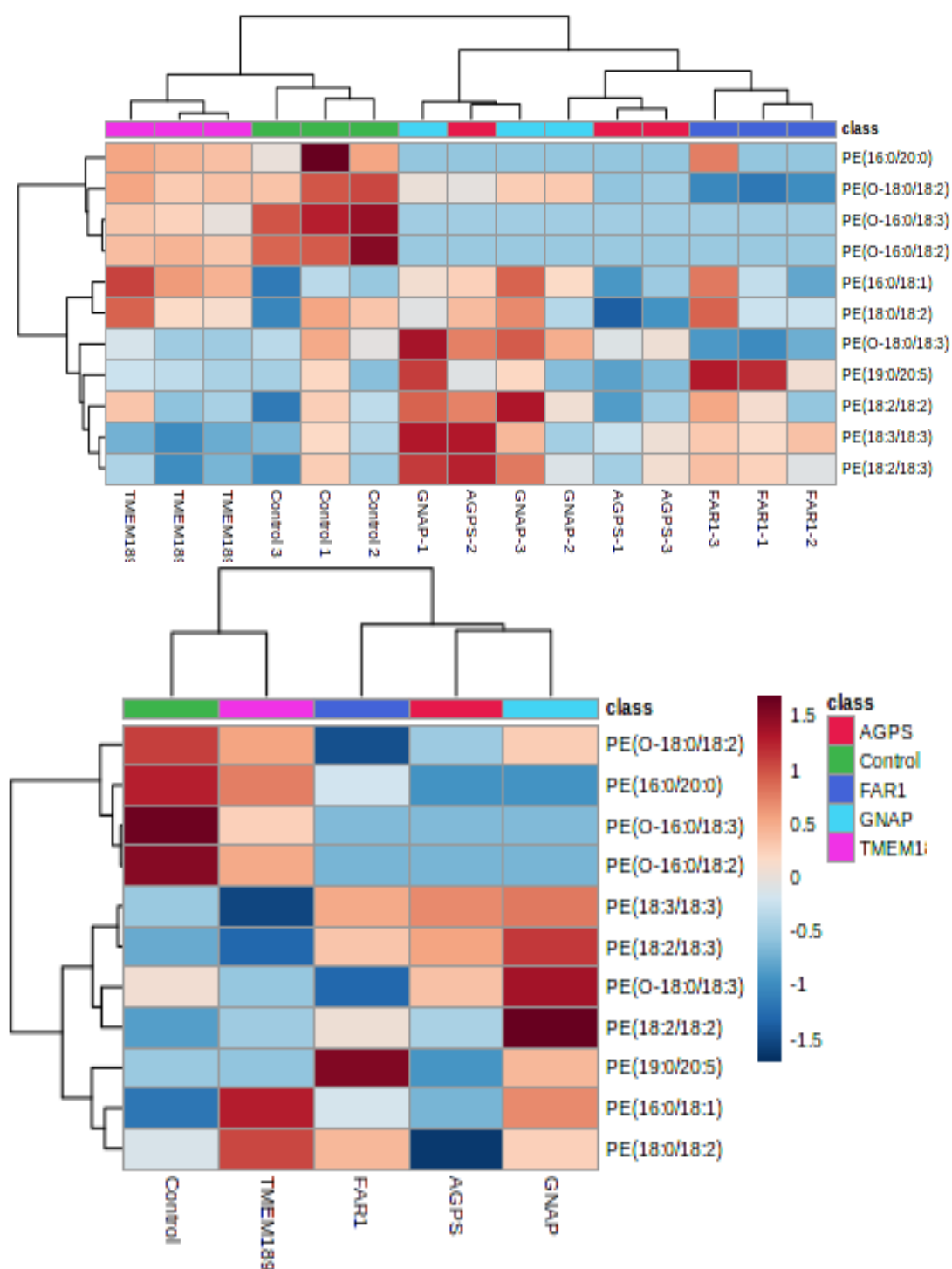


Figure 5-14 A) Shows the variation of phosphatidylethanolamine (PE) and alkyl-phosphatidylethanolamines (PE-O) species in individual mutant samples B) shows the mutant group average associations. Looking at the variations in PE and PE-O TMEM-189 holds the most similarity to the control group, illustrated through these both coming off the same major branch. There is a good differentiation of these groups through PE/PE-O profiles as all samples sub-branched in their group clusters. GNAP and AGPS showed the greatest similarity in PE/PE-O profiles of the remaining mutant strands

When analysing all of the different lipid classes identified in healthy drosophila heads during this study we found that the top most abundant lipids in sequential order were LPE's (18:2)/(18:1)/(16:1)/(16:0); PE(18:0/18:2)/(18:2/18:2)/(16:0/18:1)/(18:2/18:3) and PE's (O-18:0/18:3)/(O-18:0/18:2). Overall, the LPE class were at the highest abundance in healthy drosophila heads followed by, PE, PE-O, PI and finally PS, as shown in Figure 5-15, this reflects findings in drosophila brain by Carvalho, M. et al., 2012, though in that study more PS was found to be present in the brain than both PI and PE-O. This high abundance of PE and PE-O bares great similarity to the human brain (Choi, J., et al., 2018). In mouse models PE(36:2) and PE(38:2) most dominating in control models, which agrees with the findings presented in Figure 5-11, in part though due to the lack of lipids greater than 18C in length the 38:2 lipid was not found to be present in our sample set (Rodemer, C., et al., 2003

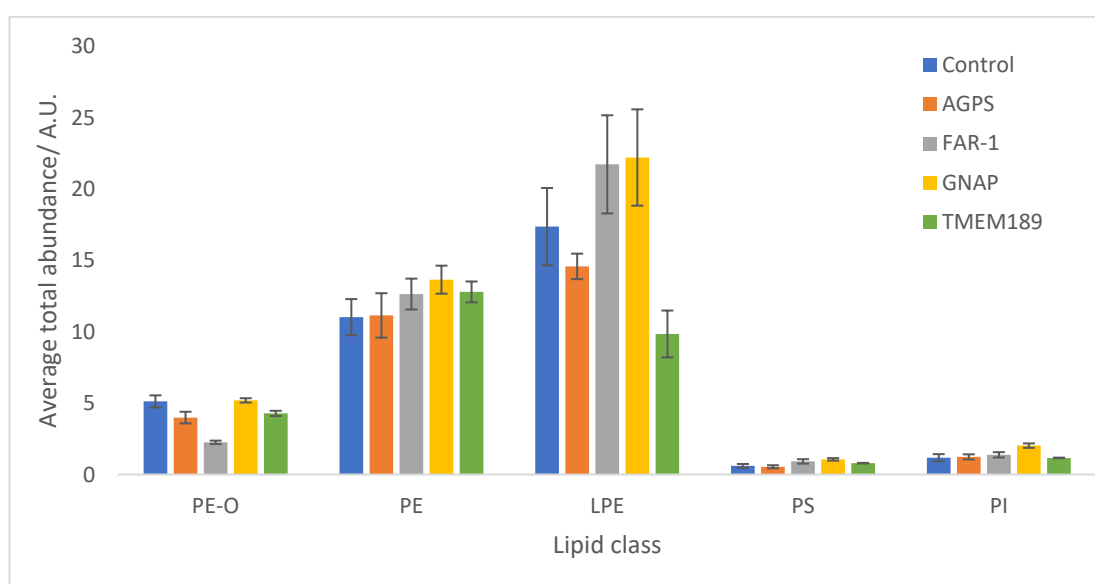


Figure 5-15-Average total abundance of each different class of lipids identified with standard error shown for n=3 samples/ mutant group.

The data also showed a great abundance of linoleic 18:2, 18:1 oleic acid, stearic acid 18:0, palmitic acid 16:0,  $\alpha$ -linoleic acid 18:3, which agrees with other research on the drosophila organism as a whole (Hammad, L.A., et al., 2011 & Overgaard, J., et al., 2008), the presence of each lipid is shown graphically in Figure 5-16.

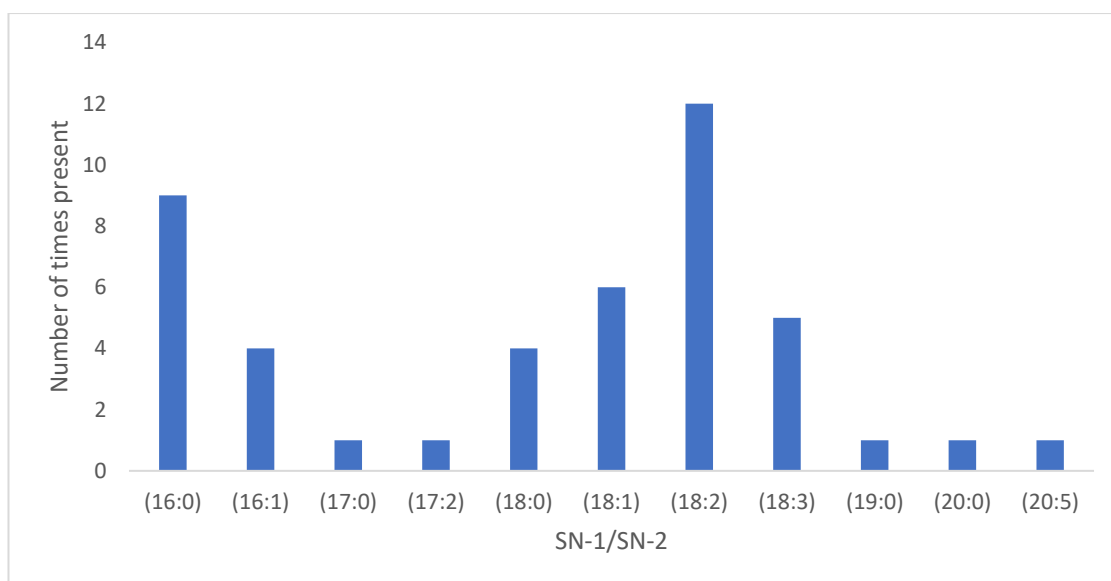


Figure 5-16- Graphical representation of the number of times each different length of carbon chain with specific level of saturation was seen in samples

The data additionally found that the majority of phospholipids had a total of 36 fatty acid carbon species present (excluding LPE as these are composed of only one fatty acid chain, which has been previously discussed).

The most common multiple saturations were more common than mono/unsaturated fatty acids found to be present in the fly heads, with a total of 3 double bonds present in the lipid as shown in Figure 5-17. The increase in PUFA species is significant as it is known to have an impact on membrane fluidity increasing and the neurological consequences of this changing the in brain.

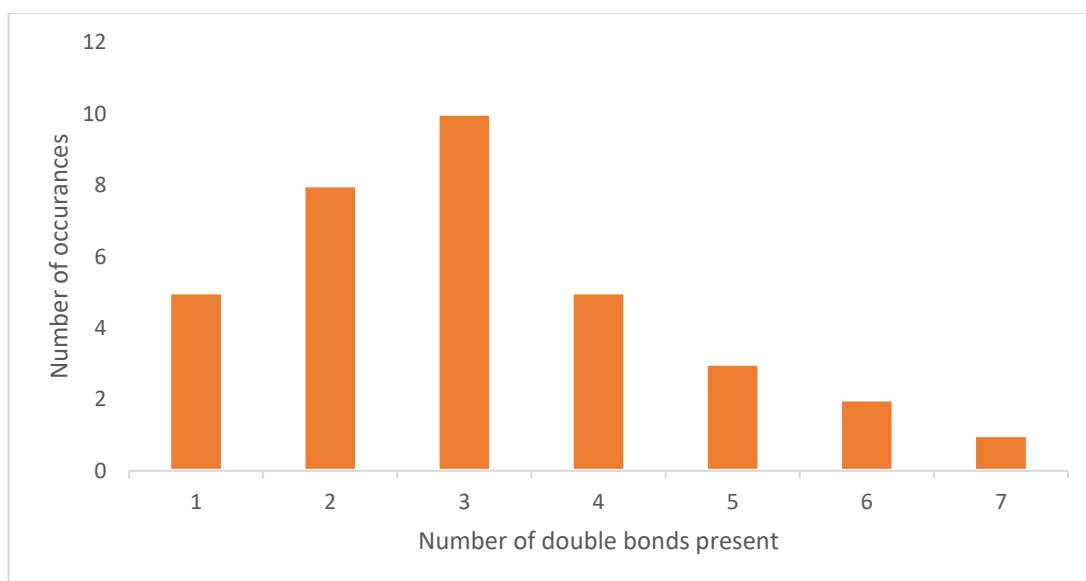


Figure 5-17- graphical representation to show the most common degree of saturation identified in all lipid species under study, with 3 double bonds being the most prevalent.

This group relationship can be clearly shown using PCA plots as seen in Figure 5-27. These further evidence FAR-1's most influencing affects on PE/PE-O presence, more so than the other mutations under study. This is shown by the greatest distance and completely distinct group clusters being achieved in the PCA plot analysis between the entire FAR-1 group and that of the controls. TMEM and AGPS also achieve very good separations from the control group, with both groups again showing completely distinct groups and confidence intervals from that of the control groups. GNAP has a

slight overlap with the control group under PCA analysis, but it is still well defined from the control group at this point.

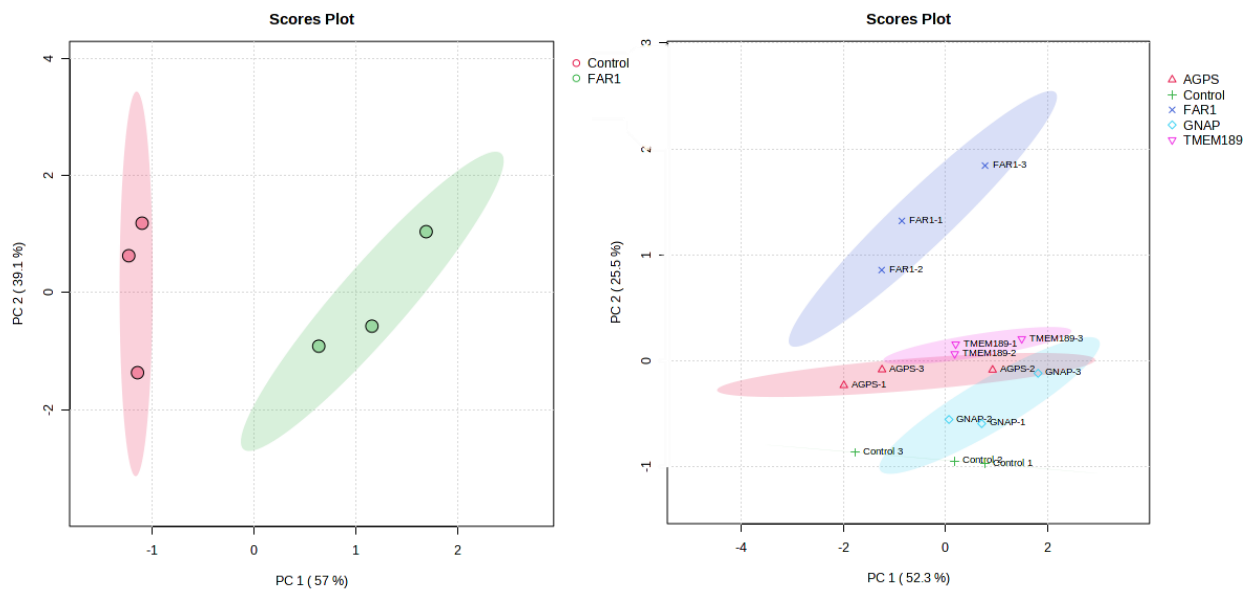


Figure 5-18- PCA analysis showing the differentiation of the mutant/ control strains from each other. A) shows the greatest distance achieved between Control and FAR-1 which was 57% accountable of PC1 and 39% PC2. B) Shows all of the groups separation using PCA analysis with PC1 accounting for 52% of the variation and PC2 25.5%. The confidence intervals around the groups is plotted with the ovals.

The earliest mutations under study (FAR1, GNPAT and AGPS), occurring in the peroxisomes, have key functioning in the regulation of the relative amount of diacyl phospholipid and ether-linked phospholipid synthesized (Enomoto, H., et al., 2020).

In almost all mutant cases there is a decrease in the level of PakEtn detected in comparison to the control sample. This is found to be significant in all FAR1 species analysed excluding PE(O-16:0/18:3), GNAP mutants only showed significant change in PE(O-16:0/18:2) and AGPS PE(O-18:0/18:2). For PE(O-16:0/18:3) & PE(O-16:0/18:2) the levels of detection of the plasmalogens in the mutant strains was recorded as zero due to no levels greater than the background noise being detected. PE(O-18:0/18:3) was seen in higher levels in the AGPS and GNAP mutant strain though this change was not found to be significant.

When looking at the changes in presence of PE excluding PE-O the mutant strains generally showed slightly higher levels than the control group. The only instances that showed significant increase compared to the control group was for PE(16:0/18:1) for both GNAP and TMEM189 and additionally PE(16:0/20:0) for AGPS and GNAP.

### 5.3.3- Plasmenylethanolamines

Tracking PE fluctuations can give an indication at the production of plasmalogen in a system through its indication of a use of an alternative pathway during lipid production. PE showed a general overall increase in all mutant strains analysed, though this fluctuation was not found to be significant as seen in Figure 5-19. Though this was not found to be the case in mutant strain TMEM-189, which showed a decrease in strains. PE(18:2/18:3), PE(18:3/18:3) and PE(19:0/20:5). PE(16:0/20:0) showed a decrease in all mutant strains when compared to the wild type, significantly so in FAR-1, GNAP and AGPS with a complete absence noted in each case. Similarly AGPS showed a decrease in the abundance of PE(18:0/18:2) and PE(19:0/20:5) but this change was not found to be significantly so. PE(16:0/18:1) was found to be significantly up-regulated in GNAP and TMEM189 mutant strains.

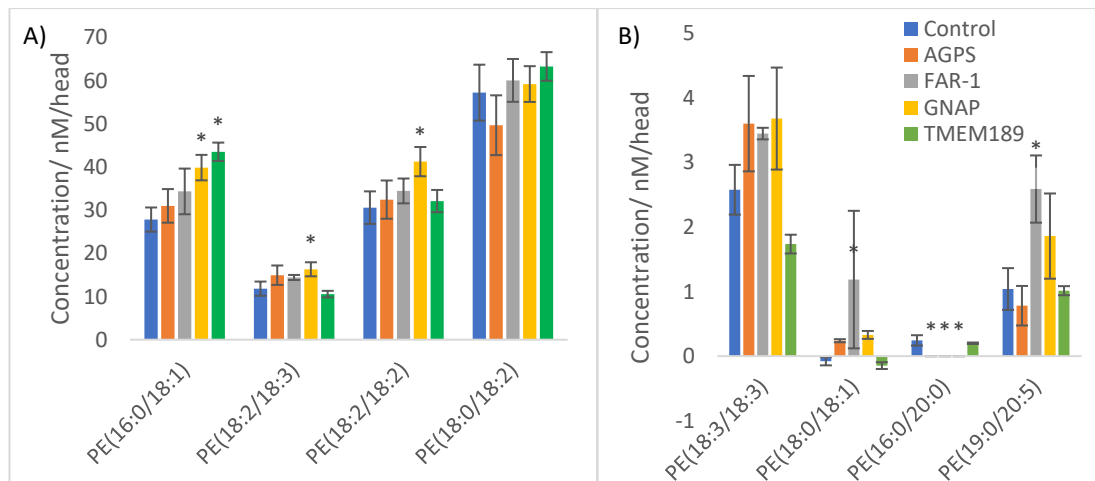


Figure 5-19- Fluctuations shown between different plasmenylethanolamines 's with an overall trend of upregulation seen in mutant strains. T-test variance for significant changes were made between each mutant group and the control groups values with bars noted with \* $>0.05$  \*\* $>0.01$  \*\*\*  $> 0.001$

#### 5.3.4- Plasmalogen GP

When considering the presence of plasmalogen GP a number of significant alterations to their presence were noted in all mutant groups. PE(O-16:0/18:3) and PE(O-16:0/18:2) was seen to be significantly decreased in all mutant strains under analysis with complete absence reported in AGPS, FAR-1 and GNAP. The remaining plasmalogens under comparison similarly almost all showed a decrease in their relative presence in comparison to control models with exception of PE(O-18:0/18:3) which instead showed a slight increase in the GNAP and AGPS mutant groups, though this was not found to be a significant change, Figure 5-20.

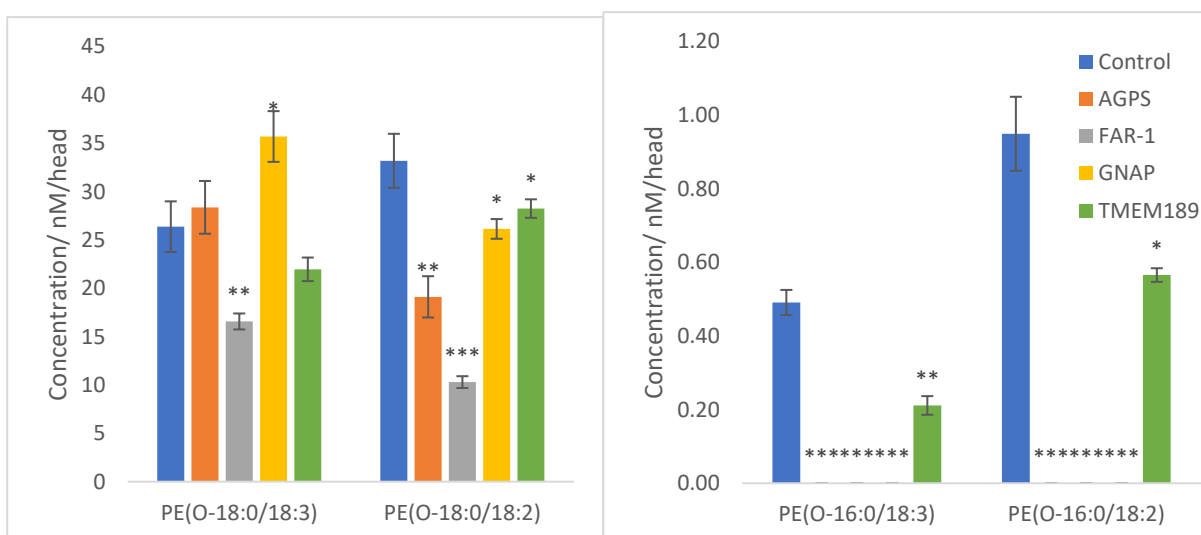


Figure 5-20- The relative alterations in plasmalogen abundance for PE(O-16:0/18:3), PE(O-16:0/18:2), PE(O-18:0/18:3) and PE(O-18:0/18:2) plasmalogen-glycerophospholipids. Significant variance between each of the mutant groups and the control group is denoted with p values as per \* $>0.05$  \*\* $>0.01$  \*\*\* $>0.001$

#### 5.3.5- LPE

In a similar fashion to PE production observations, LPE was monitored to gauge the production of plasmalogens being a further downstream product of PE. LPE lipids showed an overall increase in the mutant strains though these were never seen to vary significantly (FAR-1 and GNAP achieve p values of 0.0504 and 0.05339 respectively). Two LPE lipids were seen to significantly decrease in the mutant strains;



LPE(17:0) decrease by a significant amount in FAR-1 mutants and LPE(17:2) was seen to significantly decrease to the point of total absence in all mutant strains with the exception of TMEM-189.

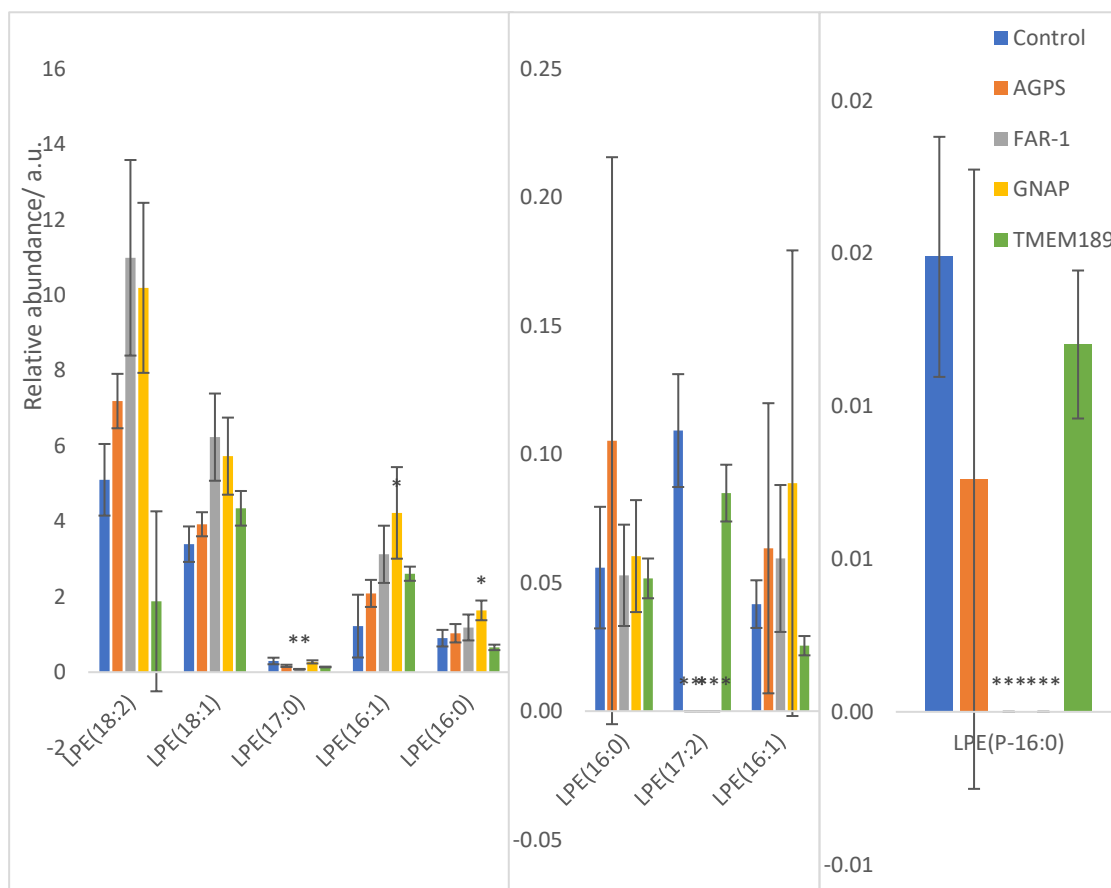


Figure 5-21- LPE(P-16:0) was shown to decrease in all mutant strains with a significant decrease due to total absence reported in FAR-1 and GNAP. Significant variance between each of the mutant groups and the control group is denoted with p values as per \* $>0.05$  \*\* $>0.01$  \*\*\*  $>0.001$ .

### 5.3.6- PS and PI presence

PS and PI lipids showed far less variance between the control and mutant groups than any of the other groups under study. This was likely due to the fact that none of the mutations applied to the drosophila directly affected the biosynthesis pathways for either subclass of lipid and as such these lipids can be considered to represent almost a standard to ensure other pathways have not been impacted by the mutants.

The only strain that showed a lot of variance in level presence when compared to the control group was GNAP. This group had a significant variance in both total PI and PS

presence as shown in Figure 5-22. No other group showed significant change in this manner.

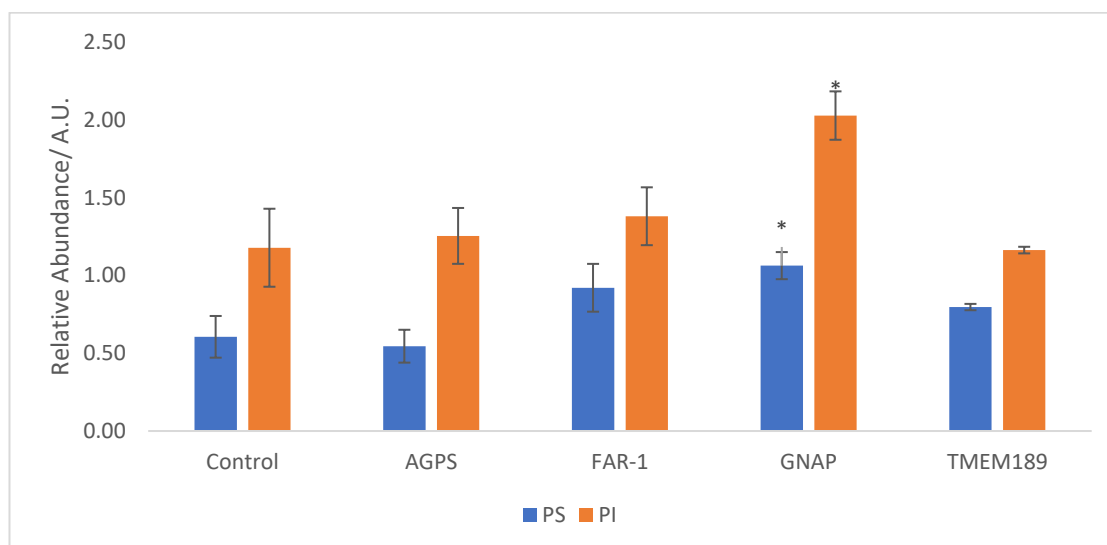


Figure 5-22- Total relative abundance of the sum of all Phosphatidylserine (PS) and Phosphatidylinositol (PI) classes of lipids observed in samples with only GNAP showing a significant variance with p.value of 0.04 for both PS and PI presence when compared to the control group. Significant variance between each of the mutant groups and the control group is denoted with p values as per \*>0.05 \*\*>0.01\*\*\* > 0.001

Additionally GNAP showed significant variance in the total PI and PS presence when compared to control groups; PS(18:2/18:2), PS(18:1/18:2), PS(18:1/18:1), PI(16:1/16:1), PI(16:1/16:0), PI(18:2/16:1), PI(16:0/18:2), PI(16:0/18:1), PI(18:1/18:3). Only one other lipid class exhibited a change with the control groups that was considered significant which was for mutants FAR-1, GNAP and AGPS for their increase in PI(16:1/16:1) and an additional change in mutant AGPS with its decrease in PI(16:0/18:1). The individual PS and PI presence can be observed in Figure 5-23 & Figure 5-24.

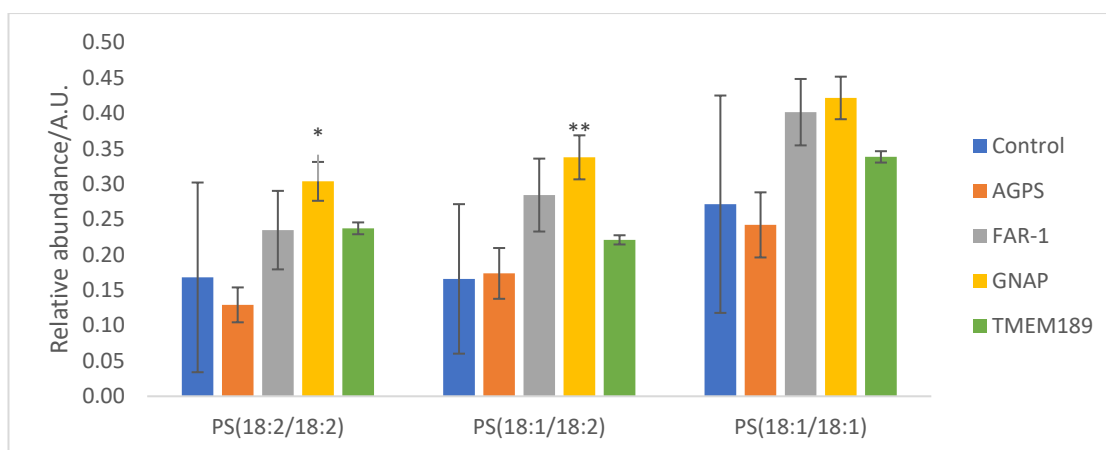


Figure 5-23- Phosphatidylserine (PS) group variance for control and each mutant strain observed. Significant variance between each of the mutant groups and the control group is denoted with p values as per \* $>0.05$  \*\* $>0.01$  \*\*\*  $>0.001$

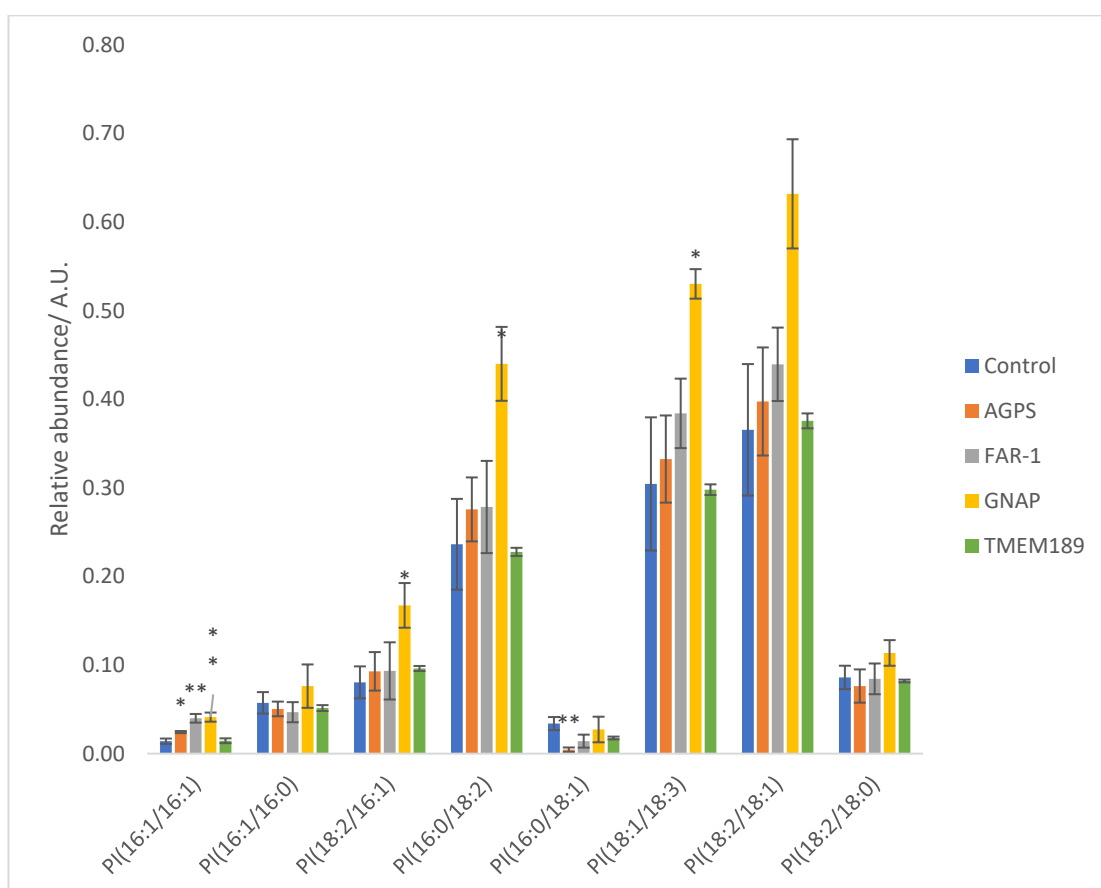


Figure 5-24- Subclass phosphatidylinositol (PI)'s relative abundance changes between control and mutant strains. Significant variance between each of the mutant groups and the control group is denoted with p values as per \* $>0.05$  \*\* $>0.01$  \*\*\*  $>0.001$

When looking at the average total presence of PE and PE-O measured in the mutant extracts there is a clear decrease in the overall PE and LPE content of the control group when compared to the mutant strains accompanied by a higher presence of PE-O.

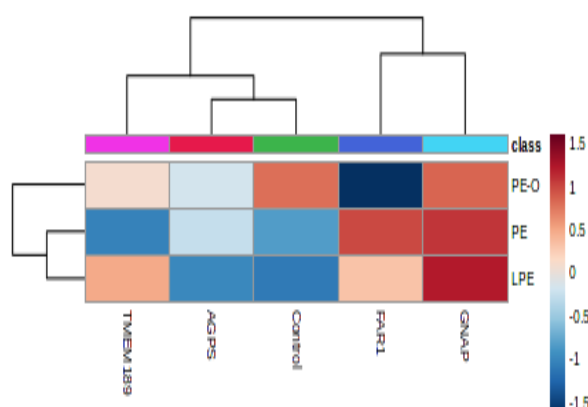


Figure 5-28- heatmap illustrating the fluctuations of average total phosphatidylethanolamine (PE), alkyl-phosphatidylethanolamines (PE-O) and ILPE present in each of the mutant and control sample extracts.

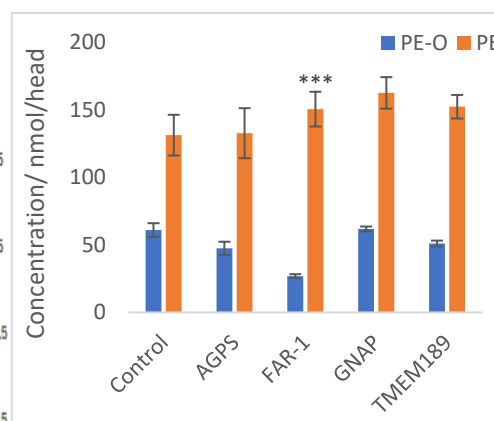


Figure 5-29- Average total abundance of phosphatidylethanolamines (PE-O) and phosphatidylethanolamine (PE) in each group under study with significant variation from the control group indicated with asterisk whereby \* $<0.05$ , \*\* $<0.01$  & \*\*\* $<0.001$ .

This trend is better illustrated when looking at the shift in PE:PEO, Figure 5-30, and changes in plasmalogen levels have knock on effects on the presence of other lipids as well. This was seen by a comparative elevation in PE-O and a reduction in LPE.

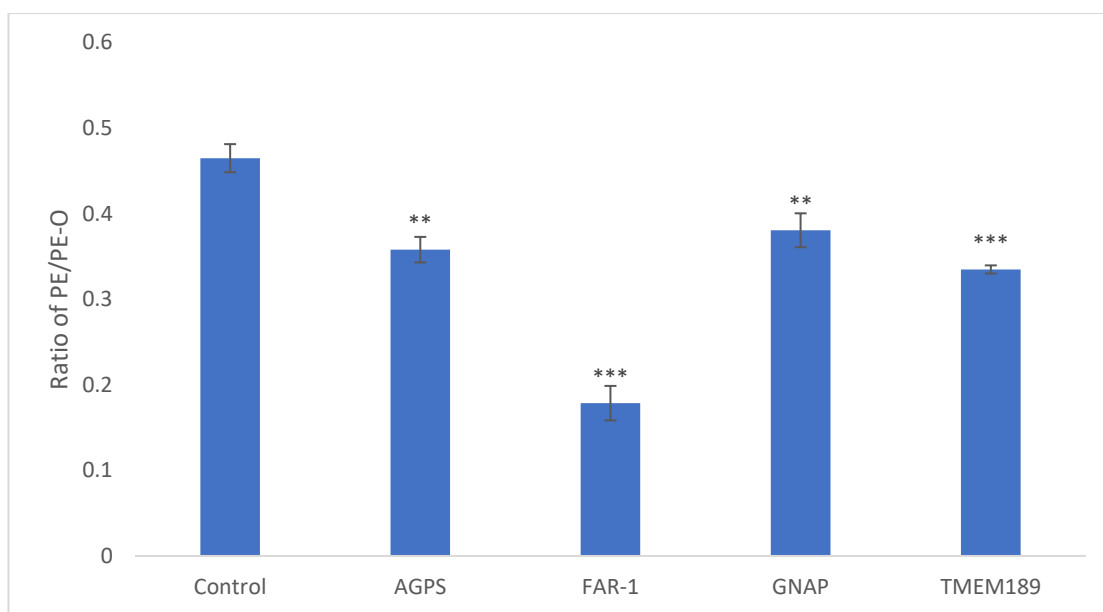


Figure 5-30- Average ratio of phosphatidylethanolamine (PE): phosphatidylethanolamines (PE-O) in each group under study with significant differences between the mutant strain and control group shown with asterisk \* $<0.05$ , \*\* $<0.01$  and \*\*\* $<0.001$

## 5.4- Discussion

Lipidomic alterations in various malignancies are very well documented, including alterations to plasmalogen levels in neurological disorders. Despite the vast changes seen, there is still a lack of understanding about how the dysregulation occurs, what broadly associated lipidomic changes occur and what their overall impact is. This study has analysed the lipidomic profiles of various mutant drosophila heads with known plasmalogen disrupted pathways, in order to document any changes that occur with particular attention paid to phosphatidylethanolamine lipids.

The first aim was to successfully extract and detect a range of lipids present in the heads of drosophila, including detection of plasmalogen based lipids, which to date have not been found to be present in the drosophila species. The study successfully characterised a number of PE, PC, PI and PS lipids including documenting the presence of plasmalogens, which included PE(P-18:0/20:4), PE(P-18:1/18:1), PE(P-18:0/18:2) and, PE(P-18:0/18:1). This is the first documentation of the presence of plasmalogen species in drosophila. This finding invites the potential for drosophila flies suitability as a model organism in future neurological disorder studies involving lipids, through proof of a better conservation of lipid species with humans than originally anticipated.

The success of plasmalogen detection is unfortunately limited in this study to one of qualitative value as unfortunately, during this study the concentrations present were either not large enough or were occluded by more dominating species with the same *m/z*. Despite this, there were key instances where a total absence of plasmalogens that were noted as being present in the control sample which indicated the loss of plasmalogen biosynthesis. These are known to have associations with Alzheimer's and other neurological conditions and so warrant a further investigation in the future.

Beyond plasmalogens, metabolic lipid markers upstream and downstream of plasmalogens were analysed and saw alterations in their abundances, signalling towards the anticipated alterations to plasmalogen levels. Numerous associated PE,

PakEtn and LysoPE were identified and analysed during this study as well as PI and PS, utilised as control lipids as these should show no or at most a limited lipid fluctuation as a result of the mutations under study.

The final aim of this chapter was to evaluate the different impacts the mutations had on lipidomic profiles in drosophila brains. Through analysing the impacts on PE, LPE and PE-O an understanding of the related impacts on plasmalogen synthesis can be drawn.

### **FAR1**

The FAR1 is the rate limiting enzyme in plasmalogen synthesis and the mutation occurs at one of the earliest points in the plasmalogen biosynthesis pathway (converting acyl-CoA into a fatty alcohol) and as such is expected to show the most dramatic profile changes in the lipidome. The overall changes noted in the FAR1 species when compared to the control sample was PE was largely unchanged, very slightly increased; PE-O was significantly decreased and LysoPE has increased. This change was to be expected as FAR1 will prevent the plasmalogen biosynthesis pathway progressing at its earliest point and as PE-O is a bi-product as well as metabolite in this process it will not only inhibit its formation but cause any reserves to get utilised as well. The down-stream affect of this would be expected to encompass a decrease in plasmalogen production as well as there being less availability of the upstream metabolites required.

As PE and LPE are analogues of PE-O and LPE(O) a lower amount of LPE(O) due to a lack of its substrate PEO may result in more LPE to counteract their effects.

Both PS and PI were slightly elevated in abundance in the FAR-1 mutant strain when compared to the healthy, though this was not significantly so.

### **GNPAT**

The GNPAT mutation occurs very early on in plasmalogen biosynthesis whereby acyl-CoA is converted into 1-acyl-DHAP ready for processing by AGPS. Unlike FAR-1, GNPAT is not the rate limiting step, but it still provides a key component in plasmalogen biosynthesis. The effects of the GNPAT mutant knockout are an increase

in PE, a very small increase in PE-O and an increase in LPE. In a very similar fashion to the FAR1 knockout mutation, the knockout of GNPAT will cause a hindered production of plasmalogens as it is one of the earliest enzymes in the biosynthetic pathway. The lack of change in PE-O in this instance suggests that 1-acyl-DHAP may be provided by/ compensated for by an alternative source. Studies have found GNPAT mutations to cause significant decreases in PE-O levels, however these can be more prominent in PC-O and so may be a cell specific effect (Spiegel, A., et al. 2022; Todt, H., et al., 2020)

### **AGPS**

The mutation in AGPS cause a small increase in PE, a decrease in PE-O and an increase in LPE. AGPS is a mutation associated with type 2 Rhizomelic Chondrodysplasia and believed to be the main cause of the disruption of plasmalogen biosynthesis in the condition. AGPS and GNPAT work together during the synthesis, whereby AGPS combines the products of GNPAT and FAR1 to form 1-o-alkyl-DHAP. Recent studies have shown that AGPS is required for GNPAT functionality (Itzkovitz, B., et al., 2012). The study further showed that absence of GNPAT with functioning AGPS does not completely stop plasmalogen synthesis, as was seen in this study with the PE-O not decreasing with the mutation, however the absence of AGPS with fully functioning GNPAT does impact the way in which GNPAT functions, this explains the stronger impact the AGPS activity had on PE-O production when compared to GNPAT.

### **TMEM-189**

The final mutation under evaluation occurs at the last point in plasmalogen synthesis PEDS-1 encoded by gene TMEM-189, which acts by putting the characteristic alk-1-enyl ether double bond upon the plasmalogen. Studies have previously shown that knockout of TMEM-189 causes complete depletion of plasmalogen and the resultant build-up of the plasmalogen precursor PE-O (Werner, E. R., Keller, et al., 2020). This



shares similar findings in the drosophila fly whereby we have noted an increase in PE-O, a decrease in PE and an increase in LPE, Figure 5-15.

The shift in PE:PEO, Figure 5-30, and changes in plasmalogen levels have knock on effects on the presence of other lipids as well. This was seen by a comparative elevation in PE-O and a reduction in LPE. Literature suggests that this may be due to the presence on higher levels of DG-O which compete with DG for the limited number of available activated headgroups of the CDP-ethanolamine and CDP-choline resulting in the synthesis of ether and non-ether analogues of PC and PE shifting the balance towards ether phospholipids at the expense of the non-ether equivalents.

### **Knock-on effects of these mutations**

During the final stage of PE/plasmalogen biosynthesis phosphoethanolamine is transferred to diacylglycerol/alkylacyl respectively via the CDP-ethanolamine pathway. If the availability of plasmalogen is decreased (as in the mutants studied) then the resultant amount of PE will increase due to a greater availability of substrate and enzyme Honsho, M., et al., (2010).

A further consequence of increased levels of PE is a resultant decrease in the amount of diacylglycerol (DAG). This is due to DAG being one of the precursors in PE biosynthesis via the Kennedy pathway described in the introduction. DAG's hold critical function structurally, as precursors for GpEtn and as mediators in signal transduction, second messengers in plasma membranes and additionally have a role in ion channel regulation and neurotransmitter release (Wood, P. L., et al., 2015 and Schuhmacher, M., et al., 2020). Alternatively there may in fact be an increase of DAG due to degradation of PE, or the up-regulations of LysoPE may result in more metabolism of Lyso-PE to DAG. Either of these will have an altered level of DAG presence with knock-on neurological consequences. Studies have shown significantly higher levels of DAG in the frontal cortex, gray matter, of brains of individuals suffering with mild cognitive impairment and also sufferers of dementia when compared to healthy controls and a comparative decrease in individuals with Alzheimer's disease whom show no cognitive impairment (Wood, P. L., et al., 2015).

All of the mutations under study resulted in an increase in the level of PE as a consequence of the plasmalogen biosynthesis impact. Higher amounts of PE can lead to an increase in  $\gamma$ -secretase activity resulting in the generation of A $\beta$  peptides. Further evidence of this came through studies that have shown that conversely a depletion in PE will lead to a decrease in  $\gamma$ -secretase activity and therefore reduce this accumulation of proteins (Calzada, E., et al., 2016). This is one of the hallmark traits of Alzheimer's disease and may in part account for the phenomenon seen. Additional to this almost all mutants had much larger amounts of LPE present which can further add to the elevated PE through LPEs ability to be converted into PE via the CDP-ethanolamine pathway (Calzada, E., et al., 2016). Another possible

explanation for the elevated levels of LPE noted in this study is that these are causative of plasmalogen degradation via oxidative stress. Oxidative cleavage of the plasmalogens distinctive vinyl-ether bond in the presence of  $H_2O_2$  has been shown to be a possible route of plasmalogen degradation. The presence of  $H_2O_2$  increases with age and disease including oxidative stress associated with inflammation. This can in turn lead to lowered levels of plasmalogens through the degradation route and consequently, higher levels of LPE (Senanayake, V., & Goodenowe, D. B. 2019). Degradation of plasmalogens in this way generates not only LPE but additionally long chain aldehydes which induce vesicle aggregation. Free aldehydes have been shown to be up-regulated in multiple sclerosis (ref) but there is little study on the further impacts that we have already discussed on the up-regulation of the LPE bi-products via this mechanism.

Both PE and LPE have a biosynthesis role in the production of PC which is another known important lipid present in the brain. Though PC has not been included in this study the mis-regulation observed in the drosophila models under study will have impacts on the PC production in the brains. PC's are known to hold critical roles in neurological homeostasis and functionality. They are known to be dysregulated in various neurological conditions including Alzheimer's disease.

PE phospholipids adopt a non-bilayer structure enabling its role in facilitating membrane fusion under physiological conditions, more so than other lipids such as PC. A result of this unique role is the larger influence that small alterations in PE can then hold over the membranes structure and functionality (Rodemer, C., et al., 2003). Studies looking at such impacts have proven the neurological transmission consequences through hindered synaptic vesicle fusion between the synaptic vesicles in axon terminals and the presynaptic membranes. This causes the release of neurotransmitters mediating interneuronal communication. This was further confirmed in PE deficient mice showing neurotransmission aberrations (Dorninger F, et al., 2019 and Dorninger F, et al., 2020). LPE's can have a similar consequence through their ability to manipulate membrane curvature. LPE has a positive impact on curvature whereas PE and DAG have a negative influence. Any impact to the membranes structure through altered presence of these structural lipids can have

detrimental consequences including signalling, trafficking as well as other cellular functions and activities that could all be associated with neurological dysregulation (McMahon, H. T., & Boucrot, E., 2015; Chan, R. B., et al., 2012)..

## 5.5- Conclusion and future perspectives

Overall, the study has provided good evidence of drosophila's suitability in neurological lipidomic studies going forward through its ability to retain and mirror findings from human and mammalian based studies. This study has provided further detail of the drosophila brains lipidome and measurable properties that can be achieved using Orbitrap analysis and can be further developed to cover larger profiles and SIM analysis going forward.

The study has also provided a better insight into the global changes that can occur within the brains lipidome that are accountable of plasmalogen dysregulation.

### 5.5.1- Future work

Due to the presence of occluding  $m/z$ 's we were not able to provide absolute quantitative values to the lipidomic presence analysed, nor were we able to provide full detail of all of the plasmalogens present in the brain samples under analysis. A further development looking at separation of these lipids via derivatisation of prior extraction techniques will be able to further add specific values to the findings already drawn or alternatively utilising larger sample sizes with the set-up of a UPLC method may alleviate this issue and highlight a wider range of lipidomic profile changes caused by mutations.

A further step to be added in this study is the correlation of these changes seen in each mutant strain to the physical and behavioural changes being seen in the mutant flies to provide a better understanding of physiological onset markers. This work is currently progressing and again should help give insight into which specific lipidomic changes are most accountable for different disease presentations.

A final limitation in this study involved the use of drosophila heads rather than brains specifically. The issue with this is that a lot of lipids are stored in the heads of

drosophila fly and so could occlude the brain profiles of interest in this study. The next step in the project involves the more specific extraction of brains from drosophila from the heads and see how these profiles differ from the whole heads profile.

## 5.6- References

Bleijerveld, O. B., Houweling, M., Thomas, M. J., & Cui, Z. (2006). Metabolipidomics: profiling metabolism of glycerophospholipid species by stable isotopic precursors and tandem mass spectrometry. *Analytical biochemistry*, 352(1), 1–14. <https://doi.org/10.1016/j.ab.2006.02.016>

Bozelli JC Jr, Epand RM. Plasmalogen Replacement Therapy. *Membranes* (Basel). 2021 Oct 29;11(11):838. doi: 10.3390/membranes11110838. PMID: 34832067; PMCID: PMC8620983.

Brites, P., Ferreira, A. S., da Silva, T. F., Sousa, V. F., Malheiro, A. R., Duran, M., Waterham, H. R., Baes, M., & Wanders, R. J. (2011). Alkyl-glycerol rescues plasmalogen levels and pathology of ether-phospholipid deficient mice. *PloS one*, 6(12), e28539. <https://doi.org/10.1371/journal.pone.0028539>

Calzada, E., Onguka, O., & Claypool, S. M. (2016). Phosphatidylethanolamine Metabolism in Health and Disease. *International review of cell and molecular biology*, 321, 29–88. <https://doi.org/10.1016/bs.ircmb.2015.10.001>

Carvalho, M., Sampaio, J. L., Palm, W., Brankatschk, M., Eaton, S., & Shevchenko, A. (2012). Effects of diet and development on the *Drosophila* lipidome. *Molecular systems biology*, 8, 600. <https://doi.org/10.1038/msb.2012.29>

Chakraborty, A., Praharaj, S. K., Prabhu, R.V.K., & Prabhu, M., (2020) Lipidomics and cognitive dysfunction – A Narrative review. *Turkish Journal of biochemistry*, 45,2, 109-119 <https://doi.org/10.1515/tjb-2020-0134>

Chan, R. B., Oliveira, T. G., Cortes, E. P., Honig, L. S., Duff, K. E., Small, S. A., Wenk, M. R., Shui, G., & Di Paolo, G. (2012). Comparative lipidomic analysis of mouse and

human brain with Alzheimer disease. *The Journal of biological chemistry*, 287(4), 2678–2688. <https://doi.org/10.1074/jbc.M111.274142>

Dementia UK. (2023) What is Alzheimer's disease. Retrieved from (<https://www.dementiauk.org/about-dementia/types-of-dementia/alzheimers-disease/>) Accessed on 23/01/2023

Dean, J. M., & Lodhi, I. J. (2018). Structural and functional roles of ether lipids. *Protein & cell*, 9(2), 196–206. <https://doi.org/10.1007/s13238-017-0423-5>

Dawson G. (2015). Measuring brain lipids. *Biochimica et biophysica acta*, 1851(8), 1026–1039. <https://doi.org/10.1016/j.bbalip.2015.02.007>

Dorninger, F., Forss-Petter, S., Wimmer, I., & Berger, J. (2020). Plasmalogens, platelet-activating factor and beyond - Ether lipids in signaling and neurodegeneration. *Neurobiology of disease*, 145, 105061. <https://doi.org/10.1016/j.nbd.2020.105061>

Dorninger F, König T, Scholze P, Berger ML, Zeitler G, Wiesinger C, Gundacker A, Pollak DD, Huck S, Just WW, Forss-Petter S, Pifl C, Berger J. Disturbed neurotransmitter homeostasis in ether lipid deficiency. *Hum Mol Genet*. 2019;28(12):2046–2061.

Enomoto, H., Furukawa, T., Takeda, S., Hatta, H., & Zaima, N. (2020). Unique Distribution of Diacyl-, Alkylacyl-, and Alkenylacyl-Phosphatidylcholine Species Visualized in Pork Chop Tissues by Matrix-Assisted Laser Desorption/Ionization-Mass Spectrometry Imaging. *Foods* (Basel, Switzerland), 9(2), 205. <https://doi.org/10.3390/foods9020205>

Farooqui A. A. (2010). Studies on plasmalogen-selective phospholipase A2 in brain. *Molecular neurobiology*, 41(2-3), 267–273. <https://doi.org/10.1007/s12035-009-8091-y>

Farooqui, A. A., & Horrocks, L. A. (2001). Plasmalogens: workhorse lipids of membranes in normal and injured neurons and glia. *The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry*, 7(3), 232–245. <https://doi.org/10.1177/107385840100700308>

Farooqui, A.A., Horrocks, L.A., Farooqui, T. (2009). Choline and Ethanolamine Glycerophospholipids. In: Lajtha, A., Tettamanti, G., Goracci, G. (eds) Handbook of Neurochemistry and Molecular Neurobiology. Springer, Boston, MA. [https://doi.org/10.1007/978-0-387-30378-9\\_2](https://doi.org/10.1007/978-0-387-30378-9_2)

Farooqui, A. A., Rapoport, S. I., & Horrocks, L. A. (1997). Membrane phospholipid alterations in Alzheimer's disease: deficiency of ethanolamine plasmalogens. *Neurochemical research*, 22(4), 523–527. <https://doi.org/10.1023/a:1027380331807>

Ferdinandusse, S., McWalter, K., Te Brinke, H., Ilst, L., Mooijer, P. M., Ruiter, J. P. N., van Lint, A. E. M., Pras-Raves, M., Wever, E., Millan, F., Guillen Sacoto, M. J., Begtrup, A., Tarnopolsky, M., Brady, L., Ladda, R. L., Sell, S. L., Nowak, C. B., Douglas, J., Tian, C., Ulm, E., ... Vaz, F. M. (2021). An autosomal dominant neurological disorder caused by de novo variants in FAR1 resulting in uncontrolled synthesis of ether lipids. *Genetics in medicine : official journal of the American College of Medical Genetics*, 23(4), 740–750. <https://doi.org/10.1038/s41436-020-01027-3>

Ginsberg, L., Rafique, S., Xuereb, J. H., Rapoport, S. I., & Gershfeld, N. L. (1995). Disease and anatomic specificity of ethanolamine plasmalogen deficiency in Alzheimer's disease brain. *Brain research*, 698(1-2), 223–226. [https://doi.org/10.1016/0006-8993\(95\)00931-f](https://doi.org/10.1016/0006-8993(95)00931-f)

Goodenowe, D. B., Cook, L. L., Liu, J., Lu, Y., Jayasinghe, D. A., Ahiahonu, P. W., Heath, D., Yamazaki, Y., Flax, J., Krenitsky, K. F., Sparks, D. L., Lerner, A., Friedland, R. P., Kudo, T., Kamino, K., Morihara, T., Takeda, M., & Wood, P. L. (2007). Peripheral ethanolamine plasmalogen deficiency: a logical causative factor in Alzheimer's disease and dementia. *Journal of lipid research*, 48(11), 2485–2498. <https://doi.org/10.1194/jlr.P700023-JLR200>

Grimm, M. O., Kuchenbecker, J., Rothhaar, T. L., Grösgen, S., Hundsdörfer, B., Burg, V. K., Friess, P., Müller, U., Grimm, H. S., Riemenschneider, M., & Hartmann, T. (2011). Plasmalogen synthesis is regulated via alkyl-dihydroxyacetonephosphate-synthase by amyloid precursor protein processing and is affected in Alzheimer's disease.

Journal of neurochemistry, 116(5), 916–925. <https://doi.org/10.1111/j.1471-4159.2010.07070.x>

Gu, J., Chen, L., Sun, R., Wang, J. L., Wang, J., Lin, Y., Lei, S., Zhang, Y., Lv, D., Jiang, F., Deng, Y., Collman, J. P., & Fu, L. (2022). Plasmalogens Eliminate Aging-Associated Synaptic Defects and Microglia-Mediated Neuroinflammation in Mice. *Frontiers in molecular biosciences*, 9, 815320. <https://doi.org/10.3389/fmolb.2022.815320>

Hammad, L. A., Cooper, B. S., Fisher, N. P., Montooth, K. L., & Karty, J. A. (2011). Profiling and quantification of *Drosophila melanogaster* lipids using liquid chromatography/mass spectrometry. *Rapid communications in mass spectrometry : RCM*, 25(19), 2959–2968. <https://doi.org/10.1002/rcm.5187>

Han, X., Holtzman, D. M., & McKeel, D. W., Jr (2001). Plasmalogen deficiency in early Alzheimer's disease subjects and in animal models: molecular characterization using electrospray ionization mass spectrometry. *Journal of neurochemistry*, 77(4), 1168–1180. <https://doi.org/10.1046/j.1471-4159.2001.00332.x>

Hartmann, T., Kuchenbecker, J., & Grimm, M. O. (2007). Alzheimer's disease: the lipid connection. *Journal of neurochemistry*, 103 Suppl 1, 159–170. <https://doi.org/10.1111/j.1471-4159.2007.04715.x>

Honsho, M., Asaoku, S., & Fujiki, Y. (2010). Posttranslational regulation of fatty acyl-CoA reductase 1, Far1, controls ether glycerophospholipid synthesis. *The Journal of biological chemistry*, 285(12), 8537–8542. <https://doi.org/10.1074/jbc.M109.083311>

Hornemann T. (2021). Mini review: Lipids in Peripheral Nerve Disorders. *Neuroscience letters*, 740, 135455. <https://doi.org/10.1016/j.neulet.2020.135455>

Itzkovitz, B., Jiralerspong, S., Nimmo, G., Loscalzo, M., Horovitz, D. D., Snowden, A., Moser, A., Steinberg, S., & Braverman, N. (2012). Functional characterization of novel mutations in GNPAT and AGPS, causing rhizomelic chondrodysplasia punctata (RCDP) types 2 and 3. *Human mutation*, 33(1), 189–197. <https://doi.org/10.1002/humu.21623>



- Kao, Y. C., Ho, P. C., Tu, Y. K., Jou, I. M., & Tsai, K. J. (2020). Lipids and Alzheimer's Disease. *International journal of molecular sciences*, 21(4), 1505. <https://doi.org/10.3390/ijms21041505>
- Lipimaps (2023). LIPID MAPS Structure database- browse classification. Retrieved from (<https://www.lipidmaps.org/databases/lmsd/browse>) Accessed on 23/01/2023
- McMahon, H. T., & Boucrot, E. (2015). Membrane curvature at a glance. *Journal of cell science*, 128(6), 1065–1070. <https://doi.org/10.1242/jcs.114454>
- Messias, M. C. F., Mecatti, G. C., Priolli, D. G., & de Oliveira Carvalho, P. (2018). Plasmalogen lipids: functional mechanism and their involvement in gastrointestinal cancer. *Lipids in health and disease*, 17(1), 41. <https://doi.org/10.1186/s12944-018-0685-9>
- Moloney, A., Sattelle, D. B., Lomas, D. A., & Crowther, D. C. (2010). Alzheimer's disease: insights from *Drosophila melanogaster* models. *Trends in biochemical sciences*, 35(4), 228–235. <https://doi.org/10.1016/j.tibs.2009.11.004>
- Moser, A. B., Steinberg, S. J., Watkins, P. A., Moser, H. W., Ramaswamy, K., Siegmund, K. D., Lee, D. R., Ely, J. J., Ryder, O. A., & Hacia, J. G. (2011). Human and great ape red blood cells differ in plasmalogen levels and composition. *Lipids in health and disease*, 10, 101. <https://doi.org/10.1186/1476-511X-10-101>
- Overgaard, J., Tomcala, A., Sørensen, J. G., Holmstrup, M., Krogh, P. H., Simek, P., & Kostál, V. (2008). Effects of acclimation temperature on thermal tolerance and membrane phospholipid composition in the fruit fly *Drosophila melanogaster*. *Journal of insect physiology*, 54(3), 619–629. <https://doi.org/10.1016/j.jinsphys.2007.12.011>
- Phan, N. T. N., Munem, M., Ewing, A. G., & Fletcher, J. S. (2017). MS/MS analysis and imaging of lipids across *Drosophila* brain using secondary ion mass spectrometry. *Analytical and bioanalytical chemistry*, 409(16), 3923–3932. <https://doi.org/10.1007/s00216-017-0336-4>

- Reiter, L. T., Potocki, L., Chien, S., Gribskov, M., & Bier, E. (2001). A systematic analysis of human disease-associated gene sequences in *Drosophila melanogaster*. *Genome research*, 11(6), 1114–1125. <https://doi.org/10.1101/gr.169101>
- Rietveld, A., Neutz, S., Simons, K., & Eaton, S. (1999). Association of sterol- and glycosylphosphatidylinositol-linked proteins with *Drosophila* raft lipid microdomains. *The Journal of biological chemistry*, 274(17), 12049–12054. <https://doi.org/10.1074/jbc.274.17.12049>
- Rodemer, C., Thai, T. P., Brugger, B., Kaercher, T., Werner, H., Nave, K. A., Wieland, F., Gorgas, K., & Just, W. W. (2003). Inactivation of ether lipid biosynthesis causes male infertility, defects in eye development and optic nerve hypoplasia in mice. *Human molecular genetics*, 12(15), 1881–1895. <https://doi.org/10.1093/hmg/ddg191>
- Schuhmacher, M., Grasskamp, A. T., Barahtjan, P., Wagner, N., Lombardot, B., Schuhmacher, J. S., Sala, P., Lohmann, A., Henry, I., Shevchenko, A., Coskun, Ü., Walter, A. M., & Nadler, A. (2020). Live-cell lipid biochemistry reveals a role of diacylglycerol side-chain composition for cellular lipid dynamics and protein affinities. *Proceedings of the National Academy of Sciences of the United States of America*, 117(14), 7729–7738. <https://doi.org/10.1073/pnas.1912684117>
- Senanayake, V., & Goodenowe, D. B. (2019). Plasmalogen deficiency and neuropathology in Alzheimer's disease: Causation or coincidence?. *Alzheimer's & dementia (New York, N. Y.)*, 5, 524–532. <https://doi.org/10.1016/j.trci.2019.08.003>
- Shen, L. R., Lai, C. Q., Feng, X., Parnell, L. D., Wan, J. B., Wang, J. D., Li, D., Ordovas, J. M., & Kang, J. X. (2010). *Drosophila* lacks C20 and C22 PUFAs. *Journal of lipid research*, 51(10), 2985–2992. <https://doi.org/10.1194/jlr.M008524>
- Shin, M., & Venton, B. J. (2018). Electrochemical Measurements of Acetylcholine-Stimulated Dopamine Release in Adult *Drosophila melanogaster* Brains. *Analytical chemistry*, 90(17), 10318–10325. <https://doi.org/10.1021/acs.analchem.8b02114>
- Spiegel, A., Lauber, C., Bachmann, M., Heninger, A. K., Klose, C., Simons, K., Sarov, M., & Gerl, M. J. (2022). A set of gene knockouts as a resource for global lipidomic

changes. *Scientific reports*, 12(1), 10533. <https://doi.org/10.1038/s41598-022-14690-0>

Su, X. Q., Wang, J., & Sinclair, A. J. (2019). Plasmalogens and Alzheimer's disease: a review. *Lipids in health and disease*, 18(1), 100. <https://doi.org/10.1186/s12944-019-1044-1>

Todt, H., Dorninger, F., Rothauer, P. J., Fischer, C. M., Schranz, M., Bruegger, B., Lüchtenborg, C., Ebner, J., Hilber, K., Koenig, X., Erdem, F. A., Gawali, V. S., & Berger, J. (2020). Oral batyl alcohol supplementation rescues decreased cardiac conduction in ether phospholipid-deficient mice. *Journal of inherited metabolic disease*, 43(5), 1046–1055. <https://doi.org/10.1002/jimd.12264>

Udagawa, J., & Hino, K. (2022). Plasmalogen in the brain: Effects on cognitive functions and behaviors attributable to its properties. *Brain research bulletin*, 188, 197–202. <https://doi.org/10.1016/j.brainresbull.2022.08.008>

Vance J. E. (2015). Phospholipid synthesis and transport in mammalian cells. *Traffic (Copenhagen, Denmark)*, 16(1), 1–18. <https://doi.org/10.1111/tra.12230>

Werner, E. R., Keller, M. A., Sailer, S., Lackner, K., Koch, J., Hermann, M., Coassin, S., Golderer, G., Werner-Felmayer, G., Zoeller, R. A., Hulo, N., Berger, J., & Watschinger, K. (2020). The TMEM189 gene encodes plasmalogen ethanolamine desaturase which introduces the characteristic vinyl ether double bond into plasmalogens. *Proceedings of the National Academy of Sciences of the United States of America*, 117(14), 7792–7798. <https://doi.org/10.1073/pnas.1917461117>

Whiley, L., Sen, A., Heaton, J., Proitsi, P., García-Gómez, D., Leung, R., Smith, N., Thambisetty, M., Kloszewska, I., Mecocci, P., Soininen, H., Tsolaki, M., Vellas, B., Lovestone, S., Legido-Quigley, C., & AddNeuroMed Consortium (2014). Evidence of altered phosphatidylcholine metabolism in Alzheimer's disease. *Neurobiology of aging*, 35(2), 271–278. <https://doi.org/10.1016/j.neurobiolaging.2013.08.001>

Wood, P. L., Medicherla, S., Sheikh, N., Terry, B., Phillipps, A., Kaye, J. A., Quinn, J. F., & Woltjer, R. L. (2015). Targeted Lipidomics of Frontal Cortex and Plasma Diacylglycerols (DAG) in Mild Cognitive Impairment and Alzheimer's Disease:

Validation of DAG Accumulation Early in the Pathophysiology of Alzheimer's Disease. Journal of Alzheimer's disease : JAD, 48(2), 537–546. <https://doi.org/10.3233/JAD-150336>

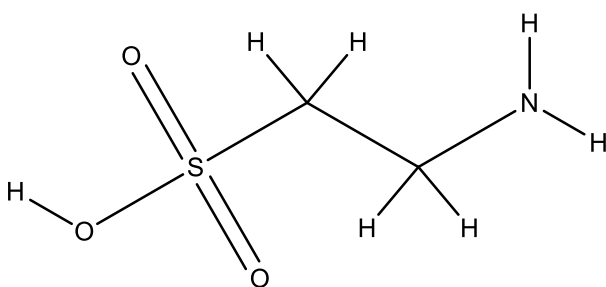
Xun, Z., Sowell, R. A., Kaufman, T. C., & Clemmer, D. E. (2007). Lifetime proteomic profiling of an A30P alpha-synuclein Drosophila model of Parkinson's disease. Journal of proteome research, 6(9), 3729–3738. <https://doi.org/10.1021/pr0700504>

Yamashita, S., Fujiwara, K., Tominaga, Y., Nguma, E., Takahashi, T., Otoki, Y., Yamamoto, A., Higuchi, O., Nakagawa, K., Kinoshita, M., & Miyazawa, T. (2021). Absorption Kinetics of Ethanolamine Plasmalogen and Its Hydrolysate in Mice. Journal of oleo science, 70(2), 263–273. <https://doi.org/10.5650/jos.ess20223>

## Chapter 6 – Analysing the thermoregulatory properties of taurine under heat stress

### 6.1- Introduction

Taurine, Figure 6-1, is one of the most prevalent free amino acids throughout the body tissue, comprising of between 50 and 60% of the free amino acid content, (Kurtz, J. A., et al., 2021). The proportion of taurine making up this amino acid pool exhibits substantial variability across different tissue structures. Notably, its most abundant concentrations are seen in the skeletal muscles, contributing to approximately 53% of its composition, thereby highlighting its integral role in facilitating muscular functionality.



Taurine  
2-aminoethanesulfonic acid  
125.15g/mol

Figure 6-1-- Structure of taurine

The quantities of taurine are lower and fluctuate amongst other tissues encompassing approximately 25% in the liver, whereas in the kidney it represents a more substantial 50% and 19% in the brain. Despite this high prevalence in tissue content, taurine only accounts for 3% of the amino acid pool in plasma sources, which emphasises its active uptake by tissues (Brosnon, J.T., and Bronson M.E., 2006 and Kurtz, J. A., et al., 2021).

### 6.1.2- Taurine intake

Taurine is one of the few sulphur containing  $\beta$ -amino acids and is described as being conditionally essential, due to it not being required in protein synthesis (Brosnon, J.T., and Bronson M.E., 2006, Galloway, S. D., et al., 2008 and Ripps, H., & Shen, W., 2012) and additionally non-essential as it can be derived from the de novo synthesis via the metabolism of Glycogen, methionine and more directly cysteine in the body when vitamin B6 is present, Figure 6-2 (Brosnon, J.T., and Bronson M.E., 2006 & Galloway, S. D., et al., 2008). Taurine can be synthesised in this way throughout various locations of the body, however the liver is a pinnacle processing site for taurine. Despite the capability to metabolically derive taurine there is still a dietary requirement for taurine to provide sufficient levels for function. Typical good sources of taurine in the diet include animal proteins, breast milk, sports drinks, or capsule form (de Luca A., et al., 2015; Tochtani S., 2022). Taurine is frequently utilised as a supplement in this capacity, most commonly being added to energy and sports drinks, to boost alertness and exercise performances.

When taken as a supplement, studies have shown that taurine levels in the plasma increase approximately ten minutes after ingestion with a peak then being seen at approximately 1 hour. Plasma content will then return to baseline levels

approximately six and a half hours later, which in the plasma are approximately 20-100 $\mu$ M (Kurtz, J. A., et al., 2021 & De Luca, A., et al., 2015).

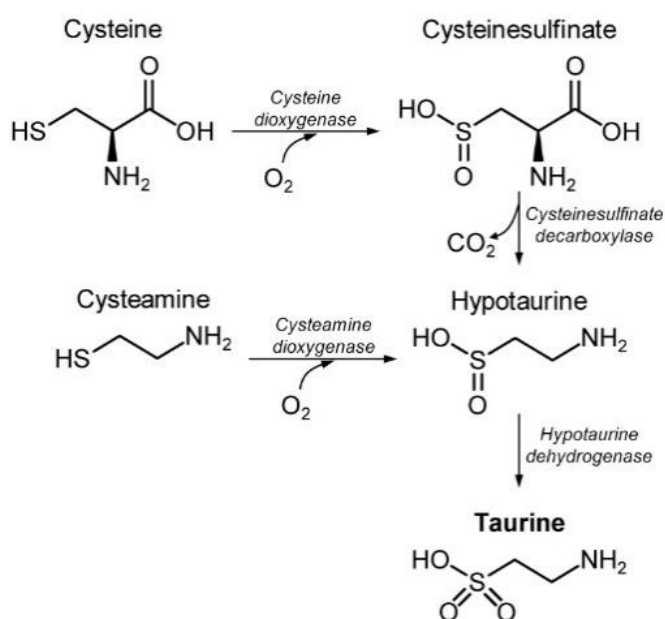


Figure 6-4- Synthesis of taurine from cysteine

### 6.1.3- Regulation and uptake

After ingestion taurine is absorbed by the small intestine, before entering the bloodstream. Some will be processed and excreted by the kidneys whereas the remainder will be carried to its target organs where its uptake is enabled through utilising the membrane transporters TauT and PAT1 (Lambert, I. H., et al., 2015; Rafiee Z et al., 2022). The former is the most dominant method for uptake, located on gene SLC6A6. The membrane channel works through active mechanism, requiring the additional movement of sodium chloride ions, and therefore their corresponding channels also need to be activated, to drive the active intracellular uptake, Figure 6-3 (De Luca, A., et al., 2015). The muscles are one of the largest uptakers of taurine via this mechanism, which is believed to be resultant of a need to regulate  $Ca^{2+}$  levels. Taurine also has a role in maintaining plasma volume, which is noted to be achieved through exiting the cells via depolarization of the cell during sodium entry or possibly co-released with water to maintain the plasma volume via homeostasis and coupled to  $Ca^{2+}$  regulation, additionally this depolarisation in turn allows a muscle to relax. This processing by the skeletal muscles is the likely reason a peak in plasma taurine

concentrations is noted after exercise Kurtz, J. A., et al., 2021 & De Luca, A., et al., 2015). The functional uptake of taurine by TauT can be inhibited by the presence of hypotaurine, GABA,  $\beta$ -Alanine and GPA (Grafe, F., et al., 2004; Ito T., et al., 2010).

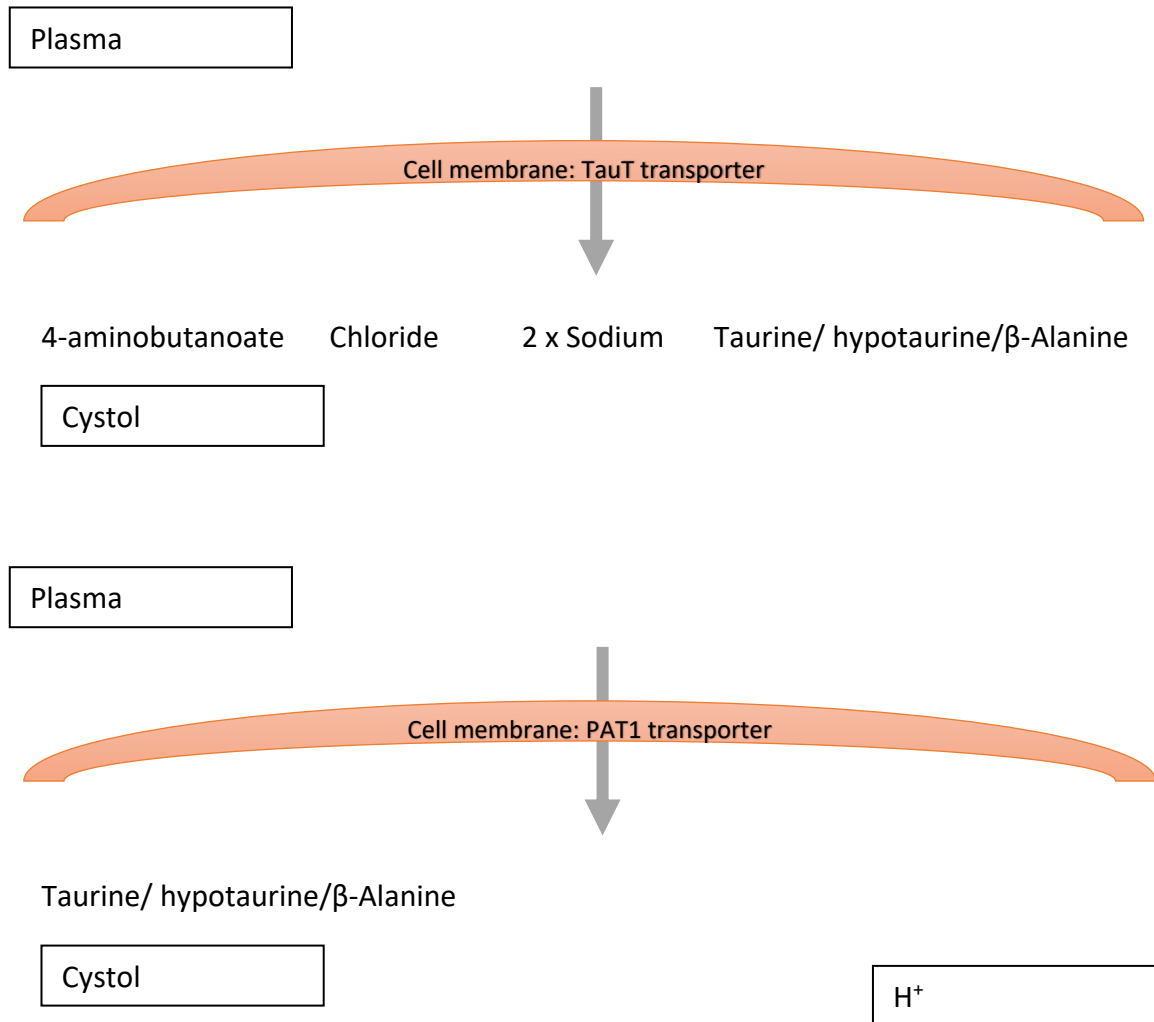


Figure 6-5- Mode of action of the taurine channels to enable to the uptake of taurine from the plasma into cellular space.

After kidney processing and regulation (this is managed through urination), taurine is excreted from the body via directly into the urine or as conjugated bile acids and is not further broken down.

#### 6.1.5- Roles

Taurine's high prevalence as an amino acid in the body gestures towards its level of functioning and importance within the body. Taurine holds multiple broad scale uses



within biological systems, though these are typically without direct metabolic participation, additionally there is limited knowledge on taurine's exact function, however it is generally known that taurine participates in various roles as shown in Figure 6-6

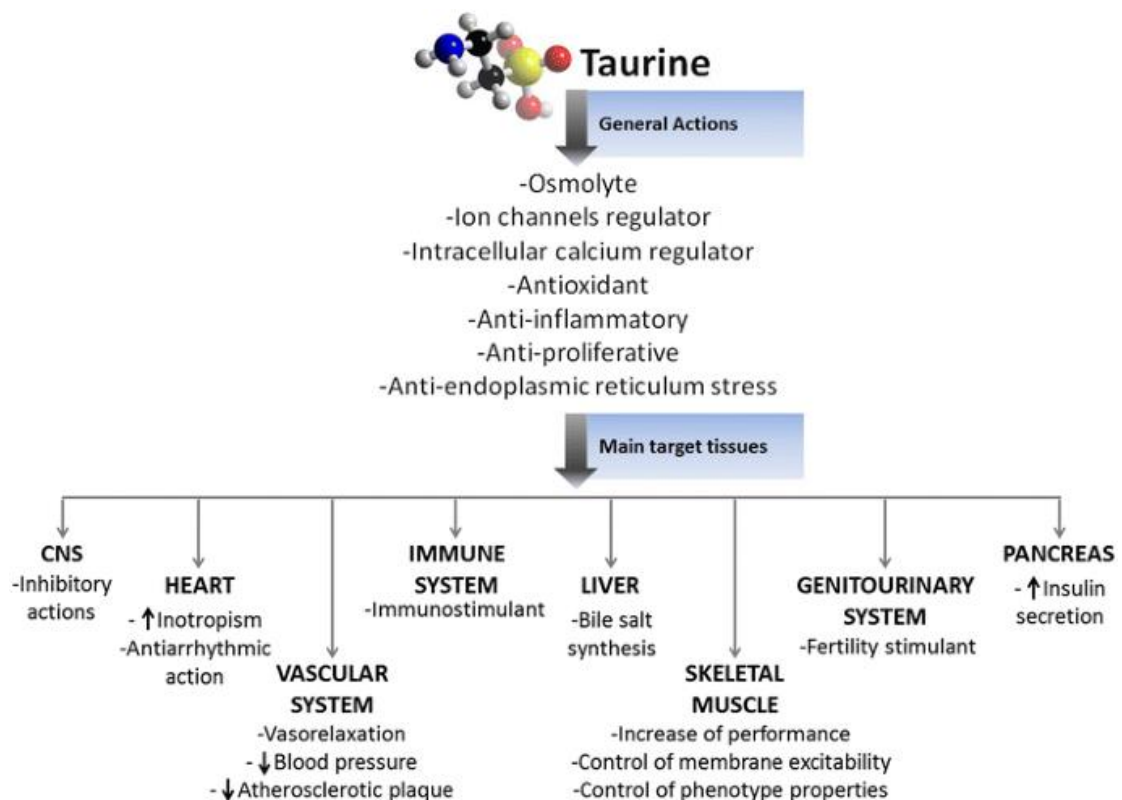


Figure 6-6- Functions of taurine in the body (De Luca, A., et al., 2015)

#### 6.1.6- Sport impact

Taurine is utilised for its known ability to enhance sports performance (Bingöl Diedhiou, A., et al., 2023, Carvalho, M. B., et al., 2020 & Kurtz J. A., et al., 2021) with observations of lowered blood pressure, improved vascular function (Sun Q., et al., 2016), Lowered heart rate accompanied with an improved sprint power (Waldron, M., et al., 2018) and improved overall physical endurance (Waldron M., et al., 2018b) achieved after taurine supplementation.

The understanding about how taurine functions to improve such attributes is only partly understood with current research pointing largely to its ability to modulate energy source. One such study has shown taurine supplementation can increase lipid metabolism, increasing rates of oxidation modifying the sources utilised for energy provisions (Carvalho, M. B., et al., 2020). Similarly, another study on taurine deficient mice whereby the Taurine transporter gene was knocked out, showed skeletal muscle wasting and drops in athletic performance with lowered running speeds and endurance. This was coupled with over three-fold increase in the levels of lactate quantitated in the blood when compared to WT mice and a heavier reliance on glucose for energy, depleting blood sources faster during exercise (Merckx, C., & De Paepe, B. 2022; Ito T et al., 2013). Additional proposed methods of metabolic alterations include a reduced glycolytic metabolism, enable a better regulation of cellular volumes (through use as an osmolyte), lower blood glucose as well as having a role in an antioxidant and anti-inflammatory capacity enabling a better muscular recovery and has been shown to be able to lower blood pressure though data is contradictory on its role in this capacity relating to exercise (Kurtz, J. A., et al., 2021; Grafe, F., et al., 2004; De Luca, A., et al., 2015;; Ripps, H., & Shen, W. 2012; Thirupathi, A., et al., 2020; Mou, S., et al., 2002). In addition to this, taurine acts as an agonist of GABA, being the main inhibitory neurotransmitter of the central nervous system with thermoregulatory roles including preventing the overheating of nerve cells (Ishiwata, T., et al., 2005; Miyazawa, T., et al., 2009).

While administering taurine orally does not lead to an elevation in resting muscle's taurine content, it does result in heightened plasma concentrations, as observed in the study by Galloway et al. in 2008. This phenomenon might impede its typical expulsion from skeletal cells, which serves an osmoregulatory role. This observation in conjunction to the vast content of taurine throughout different parts of the body implies that taurine's ergogenic impact on endurance stems from a combination of mechanisms and one not yet considered (Page L.K., et al., 2019). One such explanation may lie in thermoregulation.

#### 6.1.7- Role in thermoregulation

Maintaining complete bodily homeostasis is critical to human survival, one such mechanism requires the regulation of heat. Thermal strain is exacerbated by multiple factors including prolonged period of exercise, high environmental temperatures, high levels of humidity all of which cause a rise in core temperature beyond the homeostatic range. There are many mechanisms that the body utilises to achieve adjustments of thermal strain, but one of the most critical for heat reduction, particularly during periods of intense exercise is sweating with an overall aim to cool the skins surface. In addition to this the sweat glands play a crucial role in removing waste metabolic products from the body in a similar fashion to urine.

There are three different types of sweat gland; eccrine, apocrine and apoeccrine with the former being the major player. Eccrine glands are found across the human body's surface, though not evenly distributed. These are controlled by thermoreceptors feedback to the hypothalamus. As well as thermal stimuli they can also react to non-thermal queues including exercise through induction of the exercise pressor reflex, osmoreceptors and baroreceptors (Baker L. B., 2019; Shibasaki, M., et al., 2013).

In order to secrete, eccrine glands are triggered by their muscarinic receptors on the basolateral membrane resulting in a release of intracellular calcium stores increasing the concentration in the cytoplasm. This then results in an efflux of potassium chloride and cell shrinkage. This then results in an influx of sodium, potassium and chlorine accompanied by an opposing sodium and potassium efflux in addition to a chlorine efflux on the apical membrane. This generates an increased concentration of chlorine in the lumen, generating an electrochemical gradient, enabling sodium movement across the cellular junction. The net result is the generation of an osmotic gradient allowing the water movement into the lumen via aquaporin-5 channels. (Baker L. B. 2019)

Studies have suggested that taurine supplementation can reduce the amount of core temperature increase seen in heat-stressed rabbits that thermoneutral controls after intrathecal infusions of taurine (Frosini, M. et al., 2000). Furthermore taurine has a

known role as a vasorelaxant which could facilitate convective and evaporative cooling processes.

#### 6.1.8- Chapter aims

Overall the aims of this chapter are to:

- 1) Develop a method that can successfully measure taurine concentrations in plasma samples
- 2) Evaluate how the plasma taurine concentrations fluctuate pre and post exercise with and without taurine supplementation
- 3) Associate these taurine changes with any changes to sweating capability and impact on temperature changes experiences by participants.

#### 6.2- Methodology

##### 6.2.1- Patient samples

The work determining the sweat and thermoregulatory conditions was conducted by Jenny Peel, Swansea University with participants selected and treated as follows;

Thirteen participants (10 males and 3 females) were entered into the analysis. Participants were aged between 21 and 35 years of age with a mean age of 26.4(Standard deviation  $\pm 4.8$ ) years old, the average weight of participants was 77.2kg (Standard deviation  $\pm 8.0$ ) with an average VO<sub>2</sub> max of 52.1 mL/Kg/min (Standard deviation  $\pm 7.9$ ) all scoring average or above in this regard. There was no significant variation in VO<sub>2</sub> max between participants. It was ensured that participants had not been exposed to hot ambient temperatures prior to the study (2 months), no alcohol or caffeine was permitted 24 hours prior to participating in the study and no strenuous exercise for 48 hours before. Any and all dietary supplements were prohibited for the duration of the study. Participants arrived euhydrated (else were asked to drink water until this was achieved) and equipt with a rectal thermistor

ahead of the study. Participants then rested for 30 minutes in the seated position within the environmental chamber whilst skin thermistors were applied.

Supplementation was administered with a patients either being provided with a taurine supplement or a placebo treatment to be taken every morning for 8 days prior to entering the experiment, this was taken 1.5 hours prior to exercise on the day of the test. Taurine was administered at 50mg per kg of body mass.

During the test phase, participants were subjected to intense cardiovascular exercise starting from walking pace (for 45 minutes at a fixed humidity (34.2%) and then for 60 minutes at a ramped humidity (increased by 1.5-mmHg every 5 minutes for a further 60 minutes) through incremental (increasing pace by 1 km/h every 3 minutes) treadmill running until volatile exhaustion. In addition to increase in pace, there was also an increase in running gradients during the experimental period starting at 0% increasing to 5 and then 10% before increasing an additional 2% thereafter. Blood was drawn pre-exercise and post-exercise, being taken 2 and a half hours after the conclusion of the exercise. The room temperature was maintained at 37°C throughout. During the ramped exercises participants were exposed to a higher humidity level to inhibit ability to sweat during exercise. Participants were offered water 200mL before the exercise and 400mL before initiating the ramped exercise.

Bloods were frozen immediately before having the plasma extracted. This was again frozen at -80°C until ready for analysis.

Before and during the study participants were monitored for temperature fluctuations of skin and core, had plasma volume determined, additionally participant sweat production was measured using whole body sweat loss and, sweat gland activation and local sweat rate noted and critical pressure and heart rate monitored and noted. Significant variance between the placebo and taurine administered participants for each of these measures was then analysed.

### 6.2.2- Quantification using HPLC

In order to quantify taurine using HPLC, a derivitisation procedure was needed in order to generate a chromophore group that can be detected using the spectrophotometer. OPA reacts with Taurine in the presence of 2- mercaptoethanol to produce a substituted isoindole ring which has a short-term stability. The derivitisation can be done in an automated fashion ahead of injection using the pre-sequence methodology as follows:

#### 6.1.2.1- Plasma Taurine analysis

Plasma samples were thawed on ice before 100µL was first depleted through the addition of 400µL of methanol. Samples were kept on ice and vortexed for 10 minutes before being centrifuged at 3,000 rpm for 5 minutes. The supernatant was then transferred to a new vial and speed vacuum concentrated to dryness at 7°C. Once dried, the sample was reconstituted in 100µL of 0.4M (pH 9) sodium bicarbonate buffer before being spiked with 100nmoles of aspartic acid standard. Aspartic acid was utilised as a standard in this study as it was always found to be absent in plasma and sweat samples analysed, agreeing with literature (Dunstan, R. H., Set al., 2016) and would behave in the same manner as the other amino acids during the extraction and derivitisation processes. Additionally, aspartic was found to be one of the most distinct amino acids from Taurine Figure 6-10.

The samples were analysed for taurine content using an Agilent 1100 system utilising a pre-column derivatisation process as follows; 2.5µL of Borate buffer was mixed with 0.5µL of sample before being incubated for 0.5minutes. 0.5µL of OPA reagent was then added and thoroughly mixed again and injected. The OPA reagent used was made in house. 50mg of o-phthalaldehyde (OPA) was dissolved in 1.25mL of MeOH before 50µL of mercaptoethanol was added. 11.2mL of 0.4M Sodium Borate (pH 9.5) was then added to the finalised solution which was kept in a dark vial in the fridge for upto 2 weeks. Additional mercaptoethanol (10µL) was added to the solution every two days. The samples were separated using C18 columns.

The samples were run using a gradient elution of 40mM Sodium phosphate buffer (pH 7.8) and ACN:MeOH:H<sub>2</sub>O (45:45:10) at 40°C at a flow rate of 1mL/min. The gradient was run whereby 100% sodium phosphate buffer was used for the first 2 minutes before elution was encouraged through a gradient increasing the composition of ACN:MeOH:H<sub>2</sub>O mixture to 57% by 18.1 minutes. This then increased further to 100% ACN:MeOH:H<sub>2</sub>O by 18.8 minutes and held for 3.5 minutes before decreasing back to 100% sodium phosphate buffer by 23.2 minutes at which it remained until the stop time of 45 minutes.

The amino acids that were successfully derivatised were detected using the fluorescence detector excitation 240nm an emission 450nm with a PMT gain: 10 and a peak width of 0.5 minutes, with the measured peak heights being used for quantifications.

After every 5 samples 15% methanol was run through the column at 40°C for 30 minutes to help maintain the columns integrity.

## 6.3- Results

### 6.3.1- Optimization

#### Stability

The stability of the taurine standard was first tested to determine the best method of storing samples between extraction and analysis on the HPLC. Initially, this involved assessing the stability of taurine and other amino acids after derivitisation to establish how quickly a sample would need to be quantified, the best storage solution and storage conditions for extracted samples.

The key rate of degradation occurred between time 0 and 30 minutes, Figure 6-7, after derivitisation. This illustrated the need to develop a pre-column derivitisation protocol for multiple sample analysis as this allows for derivatisation immediately prior to injection as opposed to batch derivatisation followed by separation and

analysis (as carried out in previous chapters in relation to metabolite analysis by GCMS).

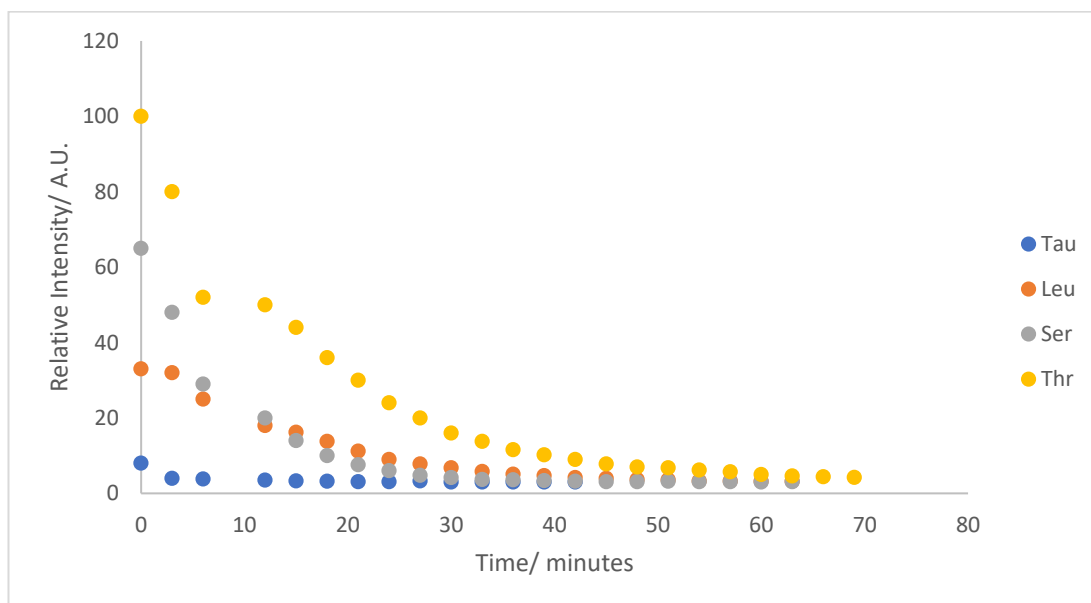


Figure 6-7- The decrease in taurine absorbance after derivatisation over time due to breakdown of fluorescence product illustrating the time critical element of the study between measurements and derivitisation

The next measure of taurine stability was undertaken in order to establish the stability of underivitisised taurine standard after being stored at different temperatures to assess how fast after extraction from plasma the samples would need to be analysed in order to determine taurine levels accurately. This was done through taking multiple aliquots from a taurine stock and storing them at room temperature, -7°C, -20°C and -80°C. The samples were held in this storage for 3 days before being analysed immediately on the HPLC. The data showed that for -80°C storage was the most preservative as shown in Figure 6-8.



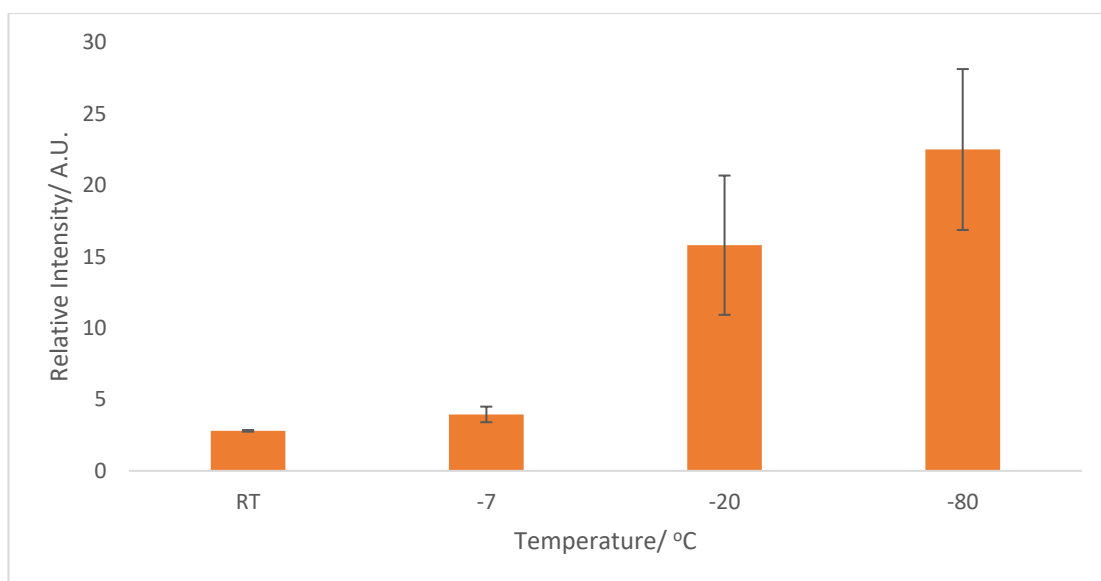


Figure 6-8- The change in taurine absorbance measured with changing storage temperatures

The best solvent to reconstitute the taurine sample in before analysis was also considered during the optimisation to ensure the best sample preservation was achieved. Different concentrations of taurine sample were reconstituted in 0.1% TCA, in water and sodium phosphate buffer showing a better preservation was achieved in sodium acetate buffer

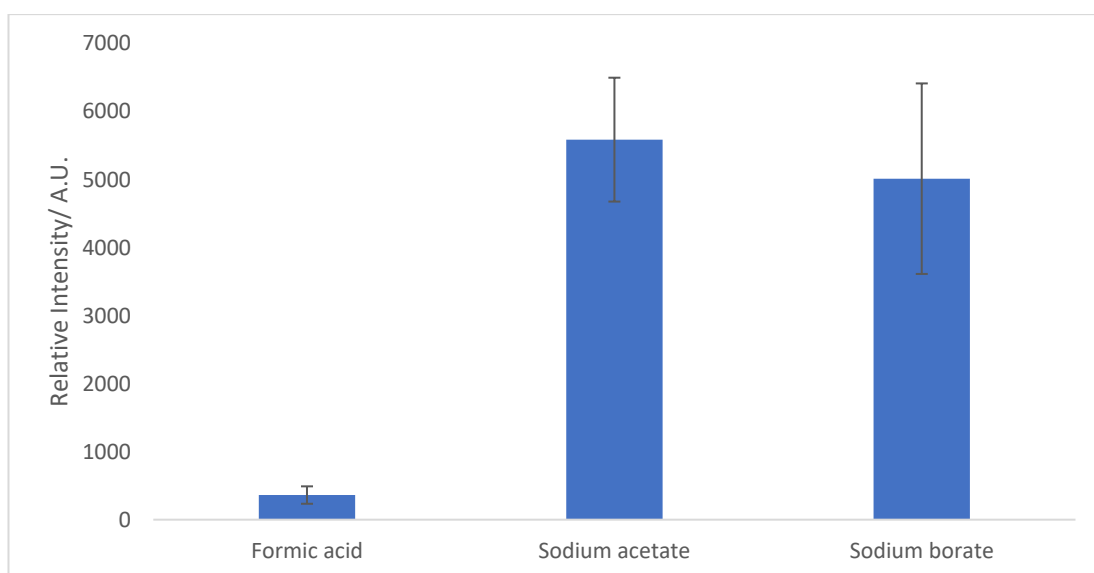


Figure 6-9- Detection of taurine after reconstitution in different buffered solutions including 0.1% formic acid, 0.025M (pH 6) sodium acetate and sodium borate (0.4M, pH 9.5)

The final optimisation consideration was to ensure a good separation of derivitised amino acids was achieved, making certain that none co-eluted, making extra assurances for Taurine. This was done by creating different amino acid stocks of the same concentration and running the planned experimental running conditions. The data shown in Figure 6-10 shows that a good level of separation was achieved in this way and both taurine and the internal standard (aspartic acid) are clearly resolved from the other amino acids derivatised.

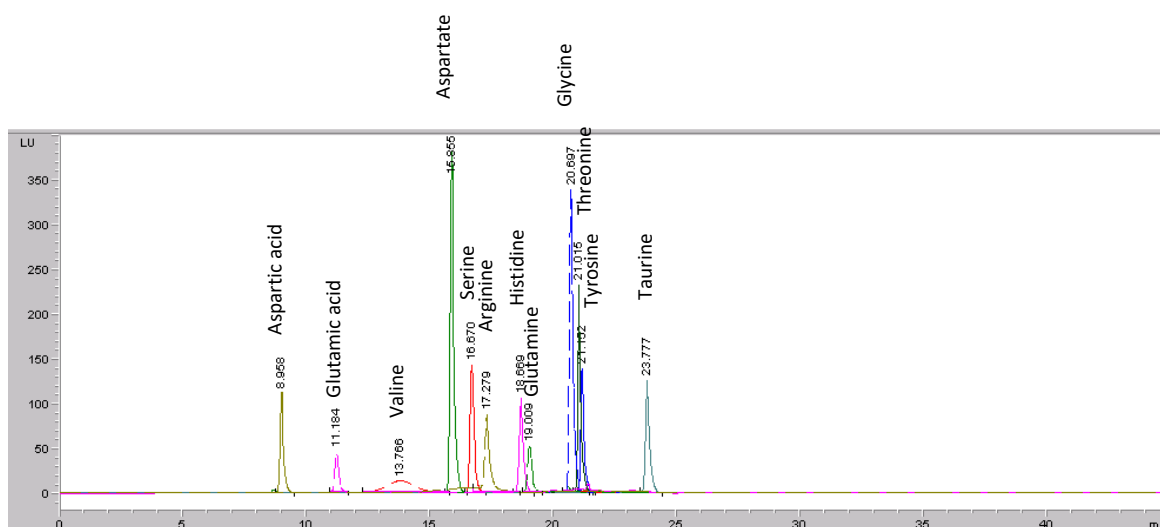


Figure 6-10- overlaid spectra from individually run amino acids showing the different elution times achieved on a c18 column. These included aspartic acid, glutamic acid, valine, asparagine, serine, arginine, histidine, glutamine, glycine, threonine, tyrosine and taurine. Cytidine, proline, methionine, phenyl alanine did not generate peaks.

The next step of this analysis looked at the detection and quantification of taurine.

### 6.3.1- Taurine quantification

Taurine standards were prepared using water as a solvent before being extracted using the same methodology utilised on the serum samples as described in 3.3.2 Metabolite extraction and analysis.

A standard curve was generated to quantify the taurine present in plasma samples and confirm the linear behaviour of taurine at working concentrations on the HPLC, shown in Figure 6-11 with triplicate data sets generated at each data point.

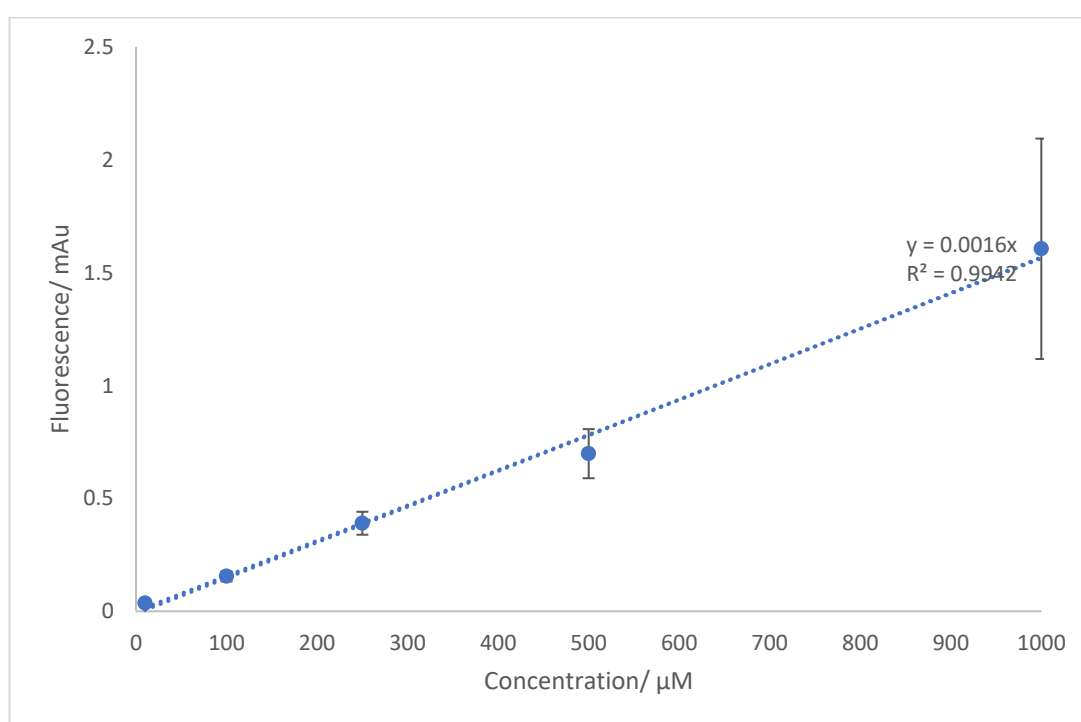


Figure 6-11- Standard curve showing the change in fluorescence absorbance of taurine with changes in taurine concentration in  $\mu\text{M}$  showing the standard deviation for each data point and a  $R^2$  value of 0.9972

The standard curve generated an  $R^2$  value of 0.9992 showing a very good level of reproducibility achievable within this range and utilising this methodological technique. Upon confirming the technique samples were then analysed and quantified using the equation of a straight line generated in the standard curve. The samples were then analysed and checks made to assure that all taurine concentrations reported fell within this linear dynamic range, Figure 6-12.

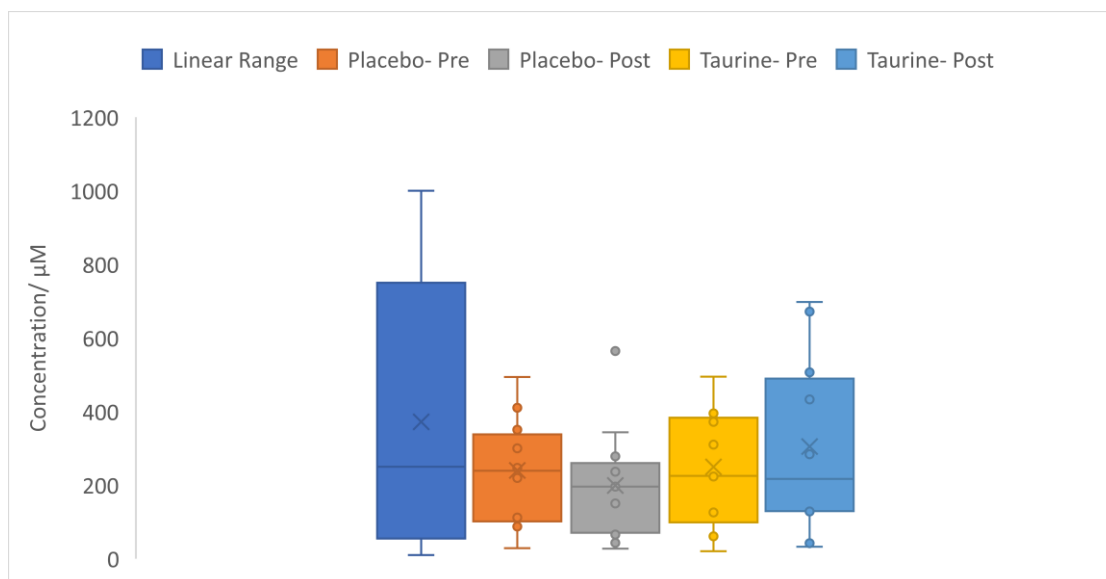


Figure 6-12- Illustration of the sample's concentration range vs the dynamic range of the standard curve generated with all samples falling within the working range of the graph.

Extracted plasma samples were then analysed using the HPLC methodology with concentrations of taurine then calculated for individual plasma samples taken from participants pre and post exercise after ingestion of placebo or taurine supplementation. The individual candidate samples are shown in Figure 6-13. A general decrease in plasma taurine concentrations was seen in participants after exercise after placebo treatment, whereas an increase was reported for taurine supplemented participants (with the exception of participants 4 and 14). For two candidates, one plasma sample was unable to be collected and so that data point (pre or post) was omitted from the study (candidate 13 and 18).

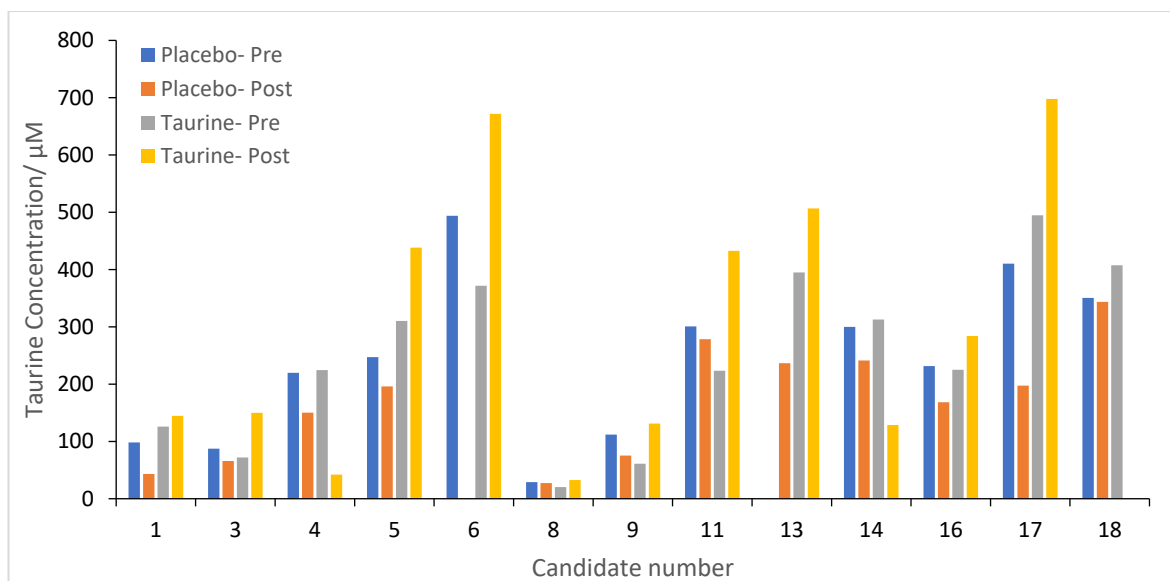


Figure 6-13- Plasma changes in taurine levels before and after exercise between candidates that took a placebo/ taurine supplement ahead of exercise.

The variation between taurine concentration within the placebo group pre and post exercise was found to significantly decrease (p. value= 0.04). The taurine supplemented group showed an increase in taurine concentration after exercise, however this was not found to be significant (p. value= 0.31).

When comparing the variation between the taurine and placebo group the pre exercise taurine concentrations were not found to vary significantly (p. value= 0.88) whereas the post-exercise did (p.value=0.03), Figure 6-14.

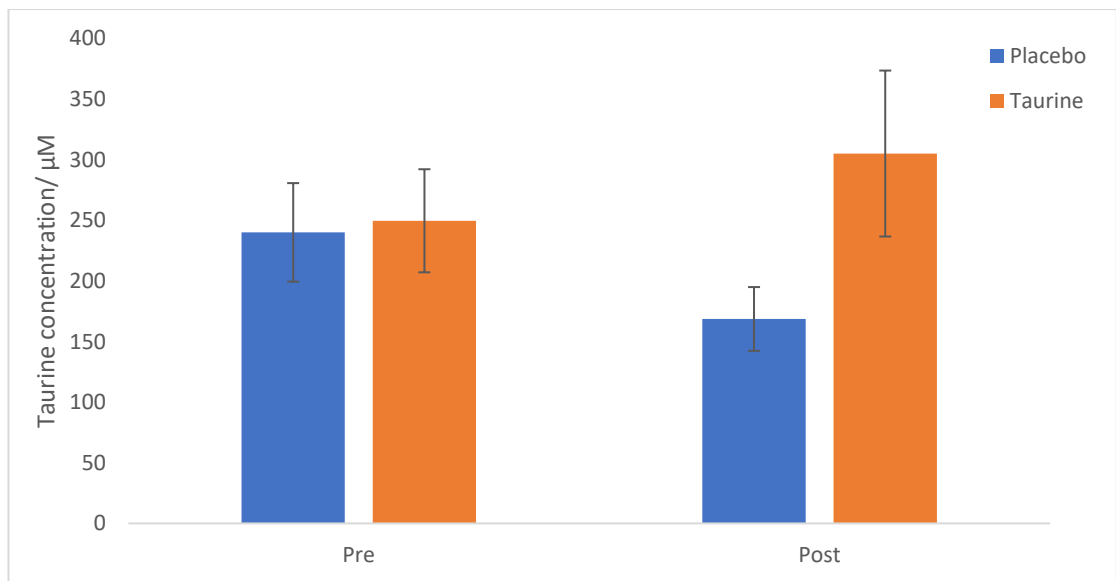


Figure 6-14- Average taurine concentration seen pre and post exercise in placebo and taurine administered participants with standard deviation bars plotted.

Additionally, the overall difference between pre and post exercise plasma taurine concentration's observed with and without taurine supplementation did show a significant change (p. value 0.02).

The data generated in this study showed a significant variance between the percentage change of taurine reported pre and post exercise in both placebo and taurine supplemented participants, achieving a p value of 0.003 as shown in Figure 6-15. Through interpreting the data in this way it is possible to develop a better understanding of the changes being seen in the analysis as it allows a level of normalisation between participants.

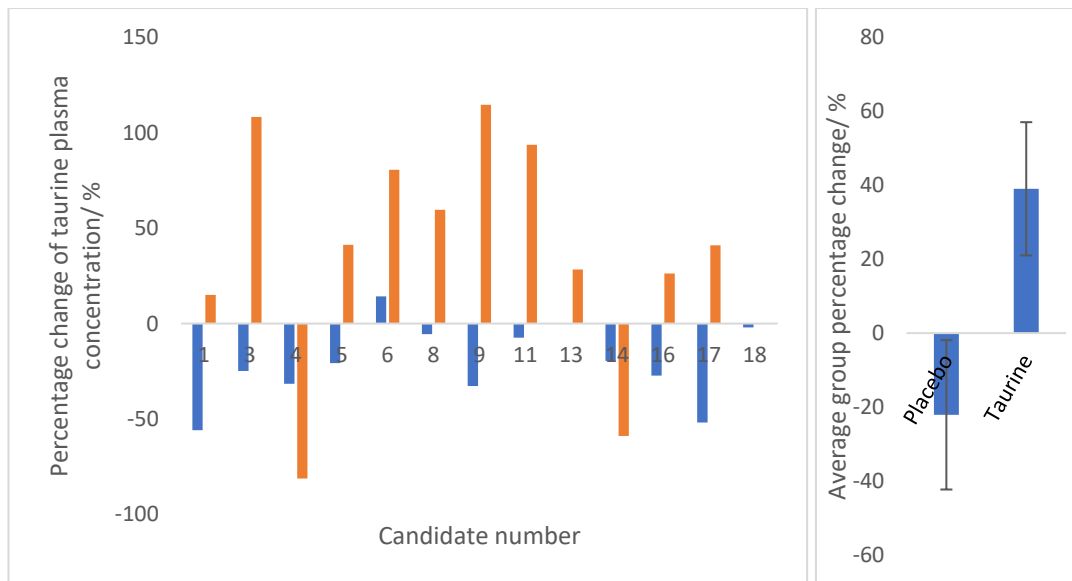


Figure 6-15- Percentage change of each participant's post exercise blood taurine content when compared to the pre-exercise taurine level.

A general trend was seen in the majority of candidates, whereby there was a significant decrease in post exercise taurine levels when considering the percentage change in candidates given a placebo supplement; whereas with taurine supplemented candidates a significant increase in taurine levels was seen as a general trend.

### 6.3.2- Thermoregulation participant measures- analysing sweating capability and impacts

In order to assess the thermoregulation of the different treatment states a number of candidate evaluations were conducted by Jenny Peel as follows after engagement with fixed intensity followed by ramped intensity workouts under the different supplementation states.

Both placebo and taurine candidates saw an increase in whole body sweat loss (WBSL) during ramped exercise in comparison to fixed intensity. This difference was

only found to be significant in the placebo group (p. value=0.001) but not in the taurine supplemented (p. value=0.083). When comparing the effect supplementation had on Fixed intensity and ramped exercises, both showed a higher rate of sweating present in the taurine supplemented groups. This was only found to be significantly different in the fixed intensity group (p. value= 0.021), Figure 6-16a.

When considering local sweat rates (LSR) a similar trend was seen to that of WBSL. There was an increase seen in the ramped intensity exercise when compared to the fixed intensity. In both instances of supplementation, taurine and placebo, this change was found to be significant (p. values  $4.98 \times 10^{-4}$  and  $2.98 \times 10^{-4}$  respectively). LSR was found to be higher in taurine supplemented candidates in both the fixed and ramped intensity exercise protocol but this change was not found to be significant in either case, Figure 6-16b.

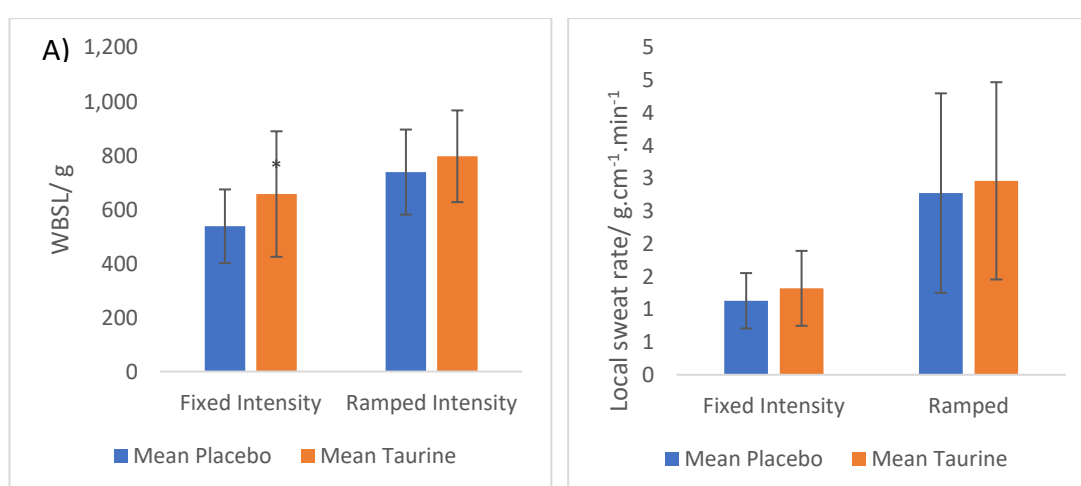


Figure 6-16- A) Whole body sweat rate (WBSL) and B) local body sweat rate (LSR) observed on participants under fixed and ramped conditions with significant variance between taurine/placebo supplementation noted with asterisks in each group, fixed or ramps, whereby \*<0.05, \*\*<0.01 & \*\*\*<0.001. Data generated by Jenny Peel.

Both supplementation groups had a great level of sweat gland activation (SGA) noted in ramped exercises when compared to fixed intensity, but in neither case was this found to be significant. There was however a significant variation when comparing the supplementation status of each different form of exercise. Taurine groups always



showed a significantly elevated level of SGA for both fixed and ramped exercises in both the 1x1 and 3x3 group testing. For fixed intensity groups the significant variation was p. value 0.02 and 0.002 for the 1x1 and 3x3 patch sizes whereas for the ramped group comparison only the 3x3 patch side showed a significant variance with a p.value of 0.008, Figure 6-17.

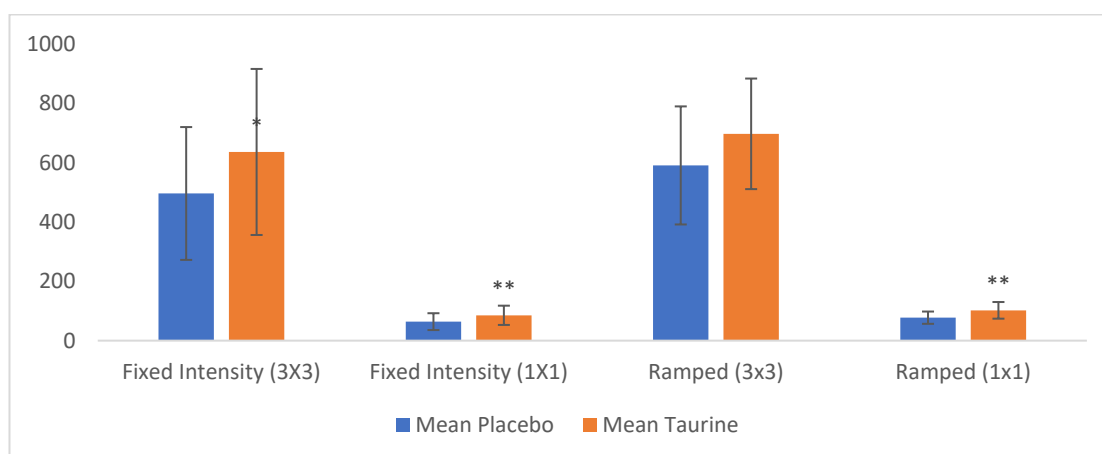


Figure 6-17- Sweat gland activation shown for fixed and ramped intensity levels with different measures made based on iodine paper size. Data generated by Jenny Peel.

When considering the thermal changes observed in the study a number of different measurements were taken to evaluate any changes that occurred with measured temperature.

Core temperature was one such measure used, with both fixed and ramped showing a significant increase in temperature between the start and end of exercise stage, as well as a significant increase across the entire experiment, in both the placebo and taurine supplemented groups. There was no significant variation seen at any of the time points of the experiment when comparing taurine and placebo supplemented in the core temperature observed in participants when comparing taurine and placebo supplemented subjects. There was however a slight overall lower core temperature noted throughout in taurine supplemented candidates. However, overall there was a decrease noted during fixed interval exercise, Figure 6-18a and b.

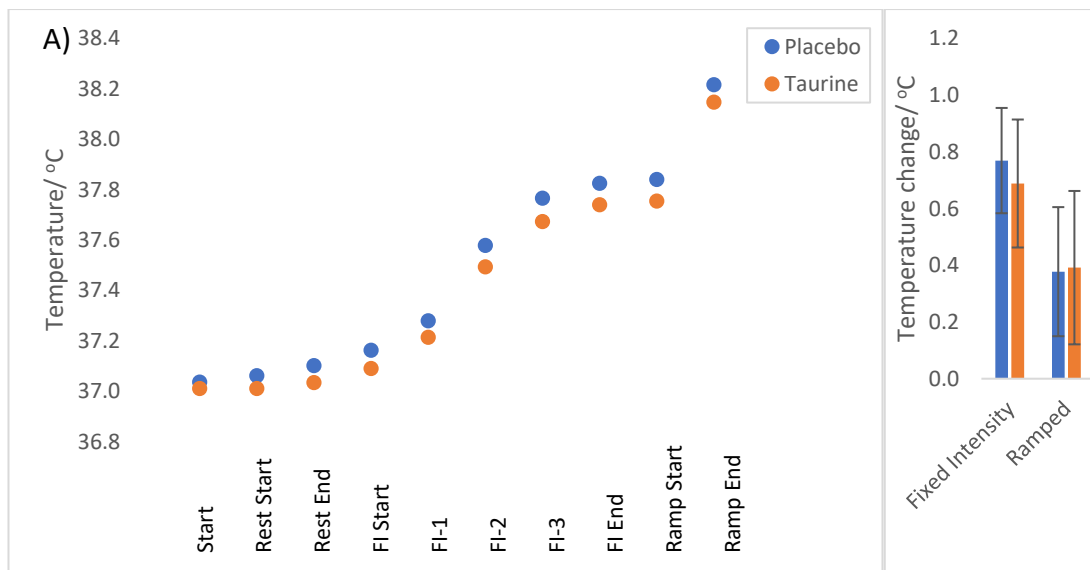


Figure 6-18- A) The gradual changes in core temperature observed over the course of the experiment including during fixed intensity (FI) at different time points and ramped activity b) Shows the average temperature change of each exercise type for both treatment groups. Data generated by Jenny Peel.

The next measure of temperature change looked at the skin surface temperature (TS). Both conditions, placebo and supplemented, showed a significant increase in skin temperature measured across both individual exercise categories and the experiment as a whole. There was no significant difference seen in skin temperature reported between taurine and placebo candidates subjected to fixed or ramped exercise at any time point, nor overall, Figure 6-19.

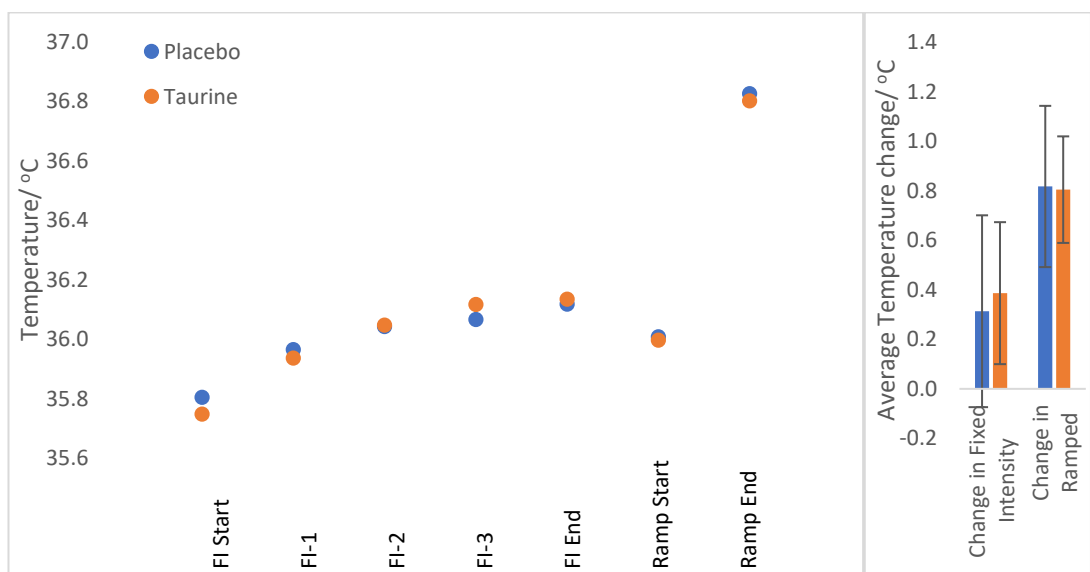


Figure 6-19- Skin temperature change observed during fixed intensity workouts with and without taurine supplementation. Data generated by Jenny Peel.

## **Perceptual**

The evaluation of heat ordinance during exercise was measuring the perceived heat capacity during the different stages and different treatments. This was done using three different measurements. The first was the perceived thermal comfort.

Perceived thermal comfort (TC), Figure 6-20, was reported using the Borg scale whereby -3 is much too cool, 0 is comfortable and 3 is much too warm. All candidates found the perceptual thermal comfort increased beyond warmer than comfortable as both exercises progressed. During fixed exercise taurine candidates felt slightly, but not significantly, cooler than placebo candidates during fixed intensity exercises, whereas the opposite was true during ramped.

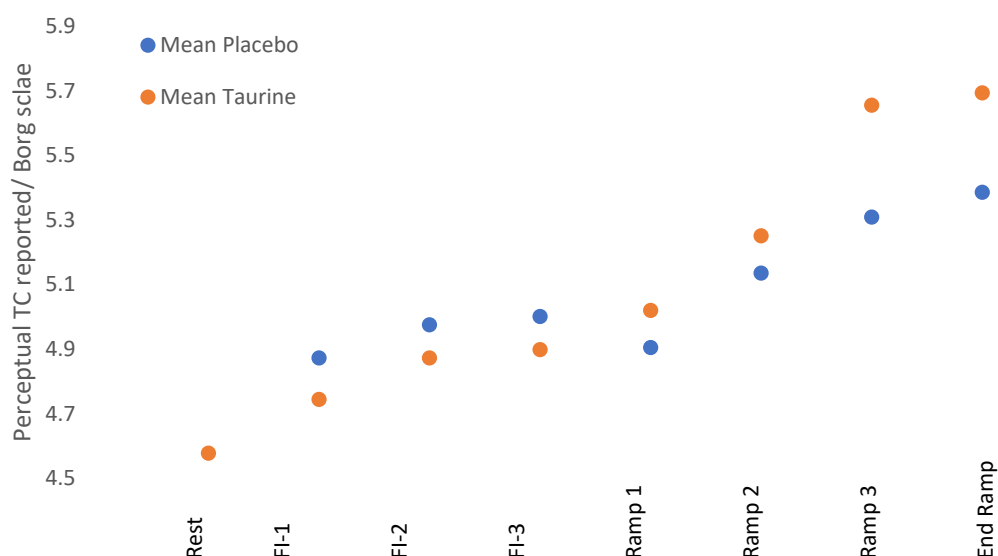


Figure 6-20- Average perceived thermal comfort reported by candidates during fixed Intensity (A) and Ramped Intensity (B) exercises with and without taurine supplementation. Data generated by Jenny Peel.

Perceived thermal sensation was also reported, Figure 6-21, in a similar fashion, but this time utilising a 9 point scale whereby -4 is very cold, 0 is neutral and 4 is very hot. In both cases the taurine supplemented group perceived their thermal sensation as

hotter than the control cases, significantly so towards the end of the ramped period, period 3 p.value 0.047 and end of ramp p.value 0.018.

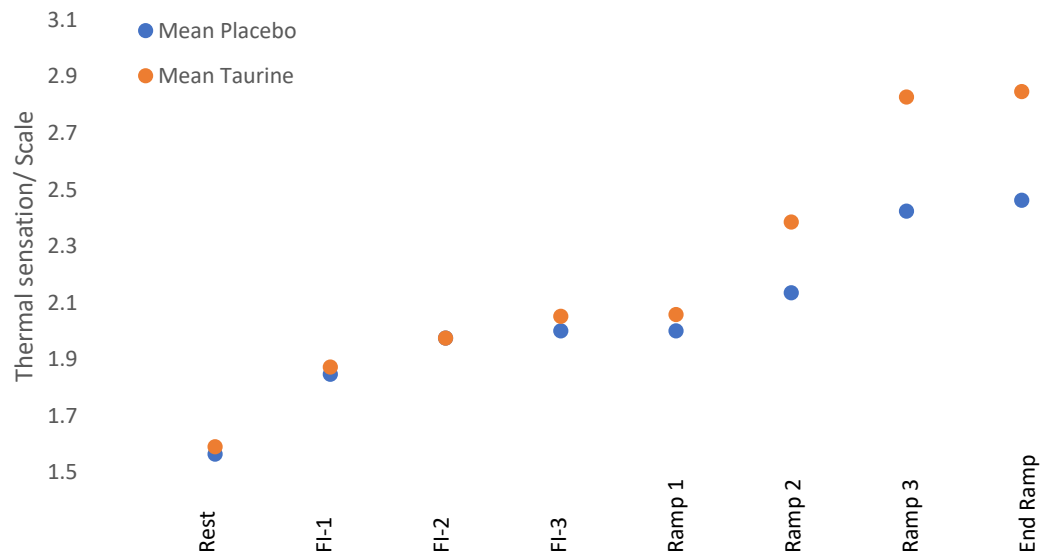


Figure 6-21- Average thermal sensation perceived during A) Fixed Intensity and B) Ramped intensity periods of exercise with and without taurine supplementation. Data generated by Jenny Peel.

The final perceived quality reported in the candidates was the rating of perceiving exertion (RPE), ranked on the Borg scale 6-20.

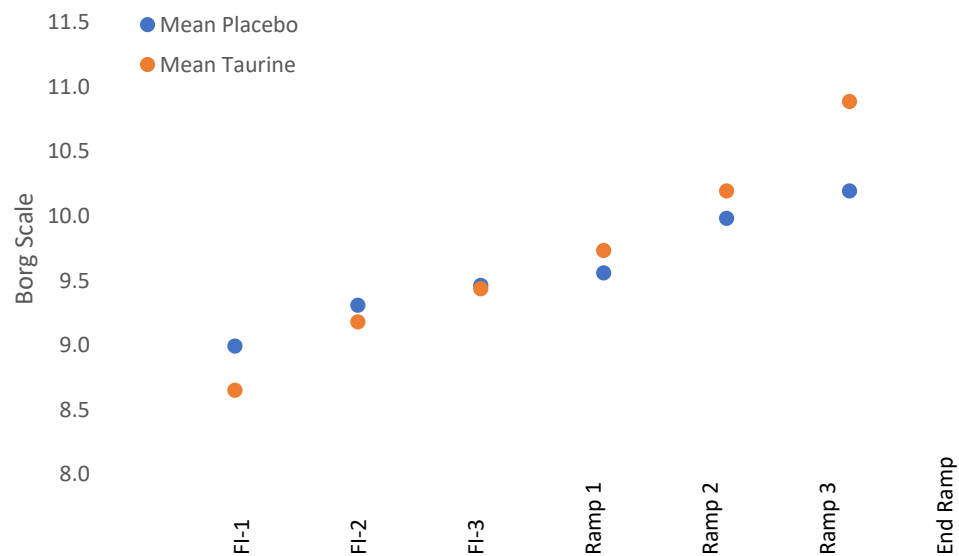
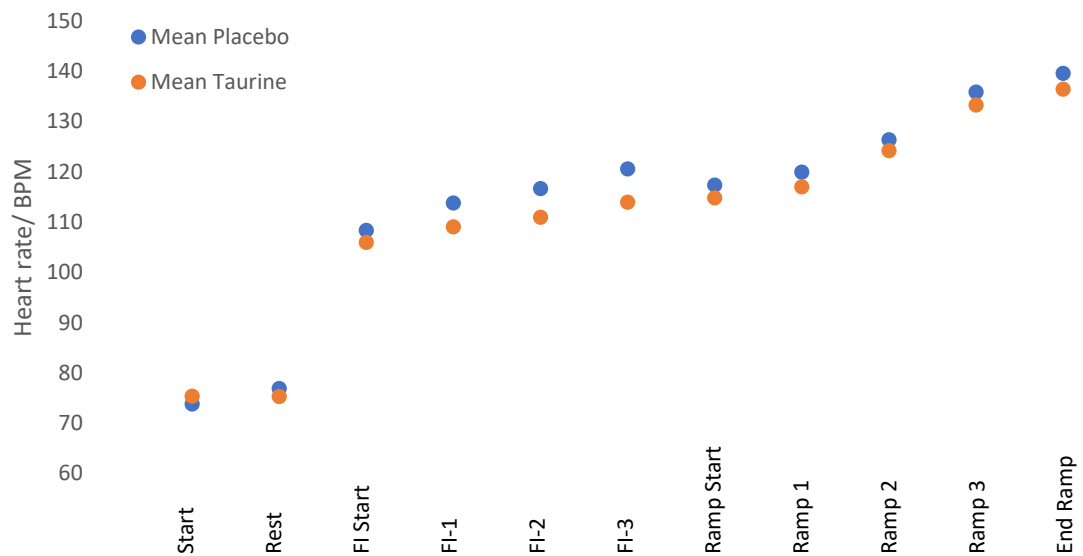


Figure 6-22- Average rating of perceived exertion observed during fixed (A) and ramped (B) exercises for placebo and taurine supplemented candidate. Data generated by Jenny Peel.

There was no observed difference in heart rate during either activities with or without taurine supplementation. Both groups showed a steady increase in average heart rate as the exercise period progressed.



There were no noteworthy changes in performance noted through  $VO_2$  or heart rate changes reported under each experimental condition that needs to be accounted for during thermoregulation evaluation.

## 6.4- Discussion

This study successfully evaluated plasma taurine concentrations in candidates pre and post exercise after consumption of taurine supplement or a placebo. The study found that taurine levels showed a significant decrease in taurine concentration ( $p$ -value = 0.04) when comparing pre and post exercise serum content; whereas with taurine supplementation the concentration was found to increase after exercise, though this change was not found to be significant. There was found to be a lot of sample variation between different participants, some of which surprisingly went against this trend, but the overall variation was still found to be considered significant.

Additionally, the difference in taurine concentration post exercise was found to be significantly up-regulated with taurine supplementation when compared to placebo groups. These findings corroborate findings by (Carvalho, M. B., et al., 2020; Cuisinier, C., et al., 2001) whereby supplementation with taurine shows an elevation of taurine presence both pre and post exercise.

Taurine is abundant in skeletal muscles, this study instead looks at the presence in the plasma reflecting what has been taken up and lost by the muscular tissues. Exercise up-regulates the presence of cotransports along muscle cell membranes this can result in an elevated release or up-take from muscles, and therefore alterations in presence in plasma. This alteration is most likely a consequence of cellular alterations in salt content during exercise as well as taurines utilisation in ion modulation and involvement in excitation-contraction coupling mechanisms (Ito T et al., 2010). Another explanation for the increase is taurine's use as a mechanism for controlling calcium homeostasis resulting in a net loss from the muscle cells. A final potential explanation for the increase observed in plasma lies in taurine's role reducing oxidative injury and being an anti-inflammatory factor, which will be required more during intense levels of exercise. In order to correct for this the body may stimulate the release of more taurine (Kurtz J.A., et al., 2021; Ishikura K., et al.,

2013; Ito T et al., 2010). The greater changes observed after supplementation may simply be an explanation of the greater availability of taurine to perform these roles.

There were a number of measurements conducted to evaluate how temperature changes were impacted during exercise with and without taurine supplementation. Core body temperature was evaluated and showed a slight, but not significant, decrease in participants that had taken a taurine supplement prior to exercise when compared to their placebo counterparts during fixed intensity exercises. There was a very slight increase in core temperature with taurine supplementation observed with the ramped exercise. However, the changes observed during ramped exercises are far more discrete and as this form of exercise is done towards the end of the practise already start at a higher point. Skin temperature was first reported as being slightly lower on average in the taurine supplemented group than in the placebo group, and though both groups reached an almost identical maximal skin temperature result, there was a resultant greater net change seen in the taurine group. This variation in any instance was not found to be significant. When considering the perceived temperature of candidates, in this case the taurine group did report feeling cooler relating to their thermal comfort during the fixed intensity exercise, but warmer during the ramped. This was only found to be a significant variance in the ramped group comparisons. Thermal sensation Taurine was reported as being experienced slightly higher when compared to placebo groups in every instance, significantly so in the ramped group. Rating of perceived exertion (RPE) was comparable between both study groups with no significant differences noted.

Overall, this data gives indication that taurine supplementation gives very little improvement to relating to heat production, skin cooling or the perceived heat experienced during exercise during the exercise regime undertaken in this trial.

A significant finding in this experiment however was in the variation observed in the sweat production of candidates after taurine supplementation when compared to their placebo counterparts. The whole body sweat loss was found to be significantly elevated during the fixed intensity exercises and also increased during the ramped.

The local sweat rate was found to similarly be more pronounced in both exercise formats and most noticeably was the significant increase in sweat gland activation observed in taurine supplemented groups when compared to placebo groups in every instance. This finding highlights a significant impact of taurine in a thermoregulatory capacity relating to exercise. This finding reflects what was observed by Page L.K. et al., 2019 but now can also be attributed to the amount of taurine present in the plasma.

There are three different types of sweat gland; eccrine, apocrine and apoeccrine with the former being the major player. Eccrine glands are found across the human body's surface, though not evenly distributed. These are controlled by thermoreceptors feedback to the hypothalamus. As well as thermal stimuli they can also react to non-thermal queues including exercise, through induction of the exercise pressor reflex, osmoreceptors and baroreceptors (Baker L. B., 2019; Shibasaki, M., et al., 2013). In order to secrete, eccrine glands are triggered by their muscarinic receptors on the basolateral membrane resulting in a release of intracellular calcium stores increasing the concentration in the cytoplasm. This then results in an efflux of potassium chloride and cell shrinkage. This then results in an influx of sodium, potassium and chlorine accompanied by an opposing sodium and potassium efflux in addition to a chlorine efflux on the apical membrane. This generates an increased concentration of chlorine in the lumen, generating an electrochemical gradient, enabling sodium movement across the cellular junction. The net result is the generation of an osmotic gradient allowing the water movement into the lumen via aquaporin-5 channels. (Baker L. B. 2019)

Sweat is produced by eccrine sweat glands throughout the skin of the body, with the sweat production being defined by the total number of activated sweat glands and the output of sweat per individual gland (Gagnon D., et al., 1985) and is one of the most prominent mechanisms of thermoregulation in humans during instances of heat stress. Maintaining complete bodily homeostasis is critical to human survival, one such mechanism requires the regulation of heat. There are many mechanism that the



body utilises to achieve this, but one of the most critical for heat reduction, particularly during periods of intense exercise is sweating with an overall aim to cool the skins surface. In addition to this the sweat glands play a crucial role in removing waste metabolic products from the body in a similar fashion to urine.

After ingestion, taurine is absorbed by the small intestine and before entering into the bloodstream were some is processed and excreted by the kidneys and some continues into the plasma as free amino acids until it is taken up by target organs (Wu G., 2020). Taurine is able to cross the blood brain barrier via Tau T transporter where it is known to have effects including osmoregulation, neuromodulation, appetite regulation, resistance to hypoxia and anti-oxidation (Rafiee, Z., et al., 2022). Studies have shown taurine can interact with GABA receptors in the brain which is believed to be a key aspect of their thermoregulator impacts, such interactions in the brain have been shown to induce of hyperthermia in rabbits (Frosini, M., et al., 2003). Beyond this, the exact mechanism by which taurine modulates sweating is poorly understood. It could also relate to its role in osmoregulation through its impacts on the balance of potassium and calcium ions (Rafiee, Z., et al., 2022).

A limitation of this study is in not seeing this increased sweat production have a cooling affect on candidates partaking in the exercise. Further evaluations increasing the duration of exercises, perhaps to the point of exhaustion, may enable an improved understanding of whether this type of increased sweating does eventually result in a reduction in thermal temperature and can that then be attributed to improvements in performances.

## 6.5- Conclusion

This study has provided novel insights into how taurine can impact upon thermoregulation. Previous studies have shown that taurine can improve overall sports performance but the exact mechanism in which is achieves this is unknown. This study has shown a possible attribute of taurine is through increased sweating which during prolonged periods of exercise can help to reduce heat experienced and

therefore enable candidates to exercise for longer or more intensely before reaching exhaustion. Further expansion of this study to see how the impacts of this sweating has on time taken to reach exhaustion or improvements in exercise intensity that may be consequential of the increased sweating reported in this study would help to better understand taurines involvement in thermoregulation.

## 6.5- References

- Baker L. B. (2019). Physiology of sweat gland function: The roles of sweating and sweat composition in human health. *Temperature* (Austin, Tex.), 6(3), 211–259. <https://doi.org/10.1080/23328940.2019.1632145>
- Bingöl Diedhiou, A., Milanović, Z., Can Eser, M., Şahin, F. N., Hamlin, M., & Can Yıldırım, U. (2023). The effects of taurine ingestion on anaerobic and physiological performance in female rugby players. *Research in sports medicine* (Print), 1–10. Advance online publication. <https://doi.org/10.1080/15438627.2023.2198129>
- Brosnan, J. T., & Brosnan, M. E. (2006). The sulfur-containing amino acids: an overview. *The Journal of nutrition*, 136(6 Suppl), 1636S–1640S. <https://doi.org/10.1093/jn/136.6.1636S>
- Carvalho, M. B., Brandao, C. F. C., Fassini, P. G., Bianco, T. M., Batitucci, G., Galan, B. S. M., De Carvalho, F. G., Vieira, T. S., Ferriolli, E., Marchini, J. S., Silva, A. S. R. D., & de Freitas, E. C. (2020). Taurine Supplementation Increases Post-Exercise Lipid Oxidation at Moderate Intensity in Fasted Healthy Males. *Nutrients*, 12(5), 1540. <https://doi.org/10.3390/nu12051540>
- Cuisinier, C., Ward, R. J., Francaux, M., Sturbois, X., & de Witte, P. (2001). Changes in plasma and urinary taurine and amino acids in runners immediately and 24h after a marathon. *Amino acids*, 20(1), 13–23. <https://doi.org/10.1007/s007260170062>
- De Luca, A., Pierno, S. & Camerino, D.C. Taurine: the appeal of a safe amino acid for skeletal muscle disorders. *J Transl Med* 13, 243 (2015). <https://doi.org/10.1186/s12967-015-0610-1>
- Dunstan, R. H., Sparkes, D. L., Dascombe, B. J., Macdonald, M. M., Evans, C. A., Stevens, C. J., Crompton, M. J., Gottfries, J., Franks, J., Murphy, G., Wood, R., & Roberts, T. K. (2016). Sweat Facilitated Amino Acid Losses in Male Athletes during Exercise at 32–34°C. *PloS one*, 11(12), e0167844. <https://doi.org/10.1371/journal.pone.0167844>

- Frosini, M., Sesti, C., Palmi, M., Valoti, M., Fusi, F., Mantovani, P., ... Sgaragli, G. (2000). Heat-stress-induced hyperthermia alters CSF osmolality and composition in conscious rabbits. *American Journal of Physiology-Regulatory, Integrative & Comparative Physiology*, 279, 2095–2103. doi: 10.1152/ajpregu.2000.279.6.R2095
- Frosini, M., Sesti, C., Saponara, S., Ricci, L., Valoti, M., Palmi, M., Machetti, F., & Sgaragli, G. (2003). A specific taurine recognition site in the rabbit brain is responsible for taurine effects on thermoregulation. *British journal of pharmacology*, 139(3), 487–494. <https://doi.org/10.1038/sj.bjp.0705274>
- Gagnon, D., Ganio, M. S., Lucas, R. A., Pearson, J., Crandall, C. G., & Kenny, G. P. (2012). Modified iodine-paper technique for the standardized determination of sweat gland activation. *Journal of applied physiology (Bethesda, Md. : 1985)*, 112(8), 1419–1425. <https://doi.org/10.1152/jappphysiol.01508.2011>
- Galloway, S. D., Talanian, J. L., Shoveller, A. K., Heigenhauser, G. J., & Spriet, L. L. (2008). Seven days of oral taurine supplementation does not increase muscle taurine content or alter substrate metabolism during prolonged exercise in humans. *Journal of applied physiology (Bethesda, Md. : 1985)*, 105(2), 643–651. <https://doi.org/10.1152/jappphysiol.90525.2008>
- Grafe, F., Wohlrab, W., Neubert, R. H., & Brandsch, M. (2004). Functional characterization of sodium- and chloride-dependent taurine transport in human keratinocytes. *European journal of pharmaceutics and biopharmaceutics : official journal of Arbeitsgemeinschaft fur Pharmazeutische Verfahrenstechnik e.V*, 57(2), 337–341. <https://doi.org/10.1016/j.ejpb.2003.10.010>
- Ishikura, K., Ra, S., and Ohmori, H. (2013) Exercise-induced changes in amino acid levels in skeletal muscle and plasma. *J Phys Fitness Sports Med*, 2(3): 301-310
- Ishiwata, T., Saito, T., Hasegawa, H., Yazawa, T., Kotani, Y., Otokawa, M., & Aihara, Y. (2005). Changes of body temperature and thermoregulatory responses of freely moving rats during GABAergic pharmacological stimulation to the preoptic area and anterior hypothalamus in several ambient temperatures. *Brain research*, 1048(1-2), 32–40. <https://doi.org/10.1016/j.brainres.2005.04.027>

- Ito, T., Oishi, S., Takai, M., Kimura, Y., Uozumi, Y., Fujio, Y., Schaffer, S. W., & Azuma, J. (2010). Cardiac and skeletal muscle abnormality in taurine transporter-knockout mice. *Journal of biomedical science*, 17 Suppl 1(Suppl 1), S20. <https://doi.org/10.1186/1423-0127-17-S1-S20>
- Ito, T., Yoshikawa, N., Schaffer, S. W., & Azuma, J. (2014). Tissue taurine depletion alters metabolic response to exercise and reduces running capacity in mice. *Journal of amino acids*, 2014, 964680. <https://doi.org/10.1155/2014/964680>
- Kurtz, J. A., VanDusseldorp, T. A., Doyle, J. A., & Otis, J. S. (2021). Taurine in sports and exercise. *Journal of the International Society of Sports Nutrition*, 18(1), 39. <https://doi.org/10.1186/s12970-021-00438-0>
- Lambert, I. H., Kristensen, D. M., Holm, J. B., & Mortensen, O. H. (2015). Physiological role of taurine--from organism to organelle. *Acta physiologica (Oxford, England)*, 213(1), 191–212. <https://doi.org/10.1111/apha.12365>
- Merckx, C., & De Paepe, B. (2022). The Role of Taurine in Skeletal Muscle Functioning and Its Potential as a Supportive Treatment for Duchenne Muscular Dystrophy. *Metabolites*, 12(2), 193. <https://doi.org/10.3390/metabo12020193>
- Miyazawa, T., Kawabata, T., Suzuki, T., Imai, D., Hamamoto, T., Yoshikawa, T., & Miyagawa, T. (2009). Effect of oral administration of GABA on temperature regulation in humans during rest and exercise at high ambient temperature. *Osaka city medical journal*, 55(2), 99–108.
- Mou, S., Ding, X., & Liu, Y. (2002). Separation methods for taurine analysis in biological samples. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences*, 781(1-2), 251–267. [https://doi.org/10.1016/s1570-0232\(02\)00619-0](https://doi.org/10.1016/s1570-0232(02)00619-0)
- Page, L. K., Jeffries, O., & Waldron, M. (2019). Acute taurine supplementation enhances thermoregulation and endurance cycling performance in the heat. *European journal of sport science*, 19(8), 1101–1109. <https://doi.org/10.1080/17461391.2019.1578417>

- Rafiee, Z., García-Serrano, A. M., & Duarte, J. M. N. (2022). Taurine Supplementation as a Neuroprotective Strategy upon Brain Dysfunction in Metabolic Syndrome and Diabetes. *Nutrients*, 14(6), 1292. <https://doi.org/10.3390/nu14061292>
- Ripps, H., & Shen, W. (2012). Review: taurine: a "very essential" amino acid. *Molecular vision*, 18, 2673–2686.
- Shibasaki, M., Kondo, N., & Crandall, C. G. (2003). Non-thermoregulatory modulation of sweating in humans. *Exercise and sport sciences reviews*, 31(1), 34–39. <https://doi.org/10.1097/00003677-200301000-00007>
- Sun, Q., Wang, B., Li, Y., Sun, F., Li, P., Xia, W., Zhou, X., Li, Q., Wang, X., Chen, J., Zeng, X., Zhao, Z., He, H., Liu, D., & Zhu, Z. (2016). Taurine Supplementation Lowers Blood Pressure and Improves Vascular Function in Prehypertension: Randomized, Double-Blind, Placebo-Controlled Study. *Hypertension (Dallas, Tex. : 1979)*, 67(3), 541–549. <https://doi.org/10.1161/HYPERTENSIONAHA.115.06624>
- Thirupathi, A., Pinho, R. A., Baker, J. S., István, B., & Gu, Y. (2020). Taurine Reverses Oxidative Damages and Restores the Muscle Function in Overuse of Exercised Muscle. *Frontiers in physiology*, 11, 582449. <https://doi.org/10.3389/fphys.2020.582449>
- Tochitani S. (2022). Taurine: A Maternally Derived Nutrient Linking Mother and Offspring. *Metabolites*, 12(3), 228. <https://doi.org/10.3390/metabo12030228>
- Waldron, M., Knight, F., Tallent, J. et al. The effects of taurine on repeat sprint cycling after low or high cadence exhaustive exercise in females. *Amino Acids* 50, 663–669 (2018b). <https://doi.org/10.1007/s00726-018-2554-2>
- Waldron, M., Patterson, S. D., Tallent, J., & Jeffries, O. (2018). The Effects of an Oral Taurine Dose and Supplementation Period on Endurance Exercise Performance in Humans: A Meta-Analysis. *Sports medicine (Auckland, N.Z.)*, 48(5), 1247–1253. <https://doi.org/10.1007/s40279-018-0896-2>

Wu G. (2020). Important roles of dietary taurine, creatine, carnosine, anserine and 4-hydroxyproline in human nutrition and health. *Amino acids*, 52(3), 329–360.  
<https://doi.org/10.1007/s00726-020-02823-6>

## Chapter 7 – Conclusions and future perspectives

Each chapter of this thesis utilised some form of metabolite analysis to study a different biological area with the view to generating some insight into each particular area as well as to evaluate the application of the different metabolomic analysis techniques. Individual conclusions within the specific results chapters have been used to bring together the data acquired and determine what this can tell us with regard to each chapters aims. The overall aim of this thesis was to investigate a number of metabolomic associated diseases in order to establish mass spectrometry-based techniques capable of providing usable biomarkers for diagnosis. These biomarkers were captured from various conditions and numerous different sample sources, with the aim of tailoring and optimising the most appropriate and simplistic methodology for each very different circumstance and so the discussion element of this thesis will be used to reflect on the utilisation of these techniques and consider their successes alongside areas in which other techniques exhibit superior abilities.

GCMS: was found to be a good all-round metabolic technique capable of identifying a broad range of markers, though primarily amino acid based markers from a full range of sample sources. It was a relatively fast through-put technique though did have the caveat of a more complex analytical process and sample preparation than some of the other techniques under study due to its requirement for knowledge of spectral details and need for sample extraction and derivitisation methodologies to be employed. During the thesis it successfully identified biomarkers capable of identifying cancerous from non-cancerous cell lines, markers for endometriosis and PCOS as well as faecal markers of infertility status in elephants illustrating the broad robustness of the technique. Through its capability in both targeted and untargeted analysis, a further advantage of GCMS methodology was the additional detail the technique provided in order to better understand the ailments under evaluation providing further potential for diagnostic and therapeutic understanding in future



evaluations. It also meant that no prior knowledge regarding suspected different metabolic markers was required when initiating the analysis.

HPLC was another very useful and broad scale methodology capable of distinguishing different biomarkers from a range of different bio-sources. During this thesis it was successfully utilised for modified nucleoside biomarkers in urinary samples for use in pancreatic cancer diagnosis as well as being used to monitor the alterations in taurine concentrations between pre and post exercise serums to Establish a connection with thermoregulation. In this sense, LCMS was more reliant on prior knowledge in a targeted based analysis to best establish the methodology to be undertaken. Though the sample processing required was a lot less intense than that needed in GCMS based analysis.

MALDI analysis was the fastest technique to be undertaken in the thesis, being utilised for both serum lipidomic studies for breast-milk analysis for comparisons to serum-based profiles, cell culture evaluation for biomarkers of pancreatic and lung cancer. Whilst the techniques had a narrower conference of metabolic changes, primarily being used for phospholipid verifications the analysis itself was untargeted and broad-scale.

Orbitrap was the most sensitive analytical method utilised in this study, being used to capture lipidomic changes in fly heads to establish a connection with markers for Alzheimer's disease. The study was targeted, and the level of sample processing very small.

FTIR was the most simplistic and fast method to be used for profile comparisons of serum markers distinguishing demographical groups and profile changes visible in cancerous and non-cancerous cell lines. Whilst the level of detail was very minimal, with the goal being purely one of diagnostic potential this methodology provided an extremely fast process with virtually no sample preparation required and a small amount of time required for data acquisition.

In conclusion mass spectrometry has been proven to be a successful technique to utilise in metabolomic analysis of disease biomarkers providing both diagnostic utilisation as well as an advancement in the understanding of disease progression and pathogenesis. In this way it can allow researchers a key tool to help better understand and attribute disease signatures to different disease statuses, progressions and prognosis.

Challenges still exist in this field with a greater push and requirement for faster and cheaper methodologies to keep up with the greater demands on health-care facilities as well as development for a broader knowledge base needed to better resolve some of the most complex and challenging ailments such as those discussed in this thesis.

In spite of these challenges, mass spectrometry has made significant way for the elucidation of disease underpinnings and developing a more tailored methodology enabling more targeted biomarkers capable of diagnosing more people at early disease stages, ultimately with the aim of providing the best provisions in healthcare for sufferers with the best prognosis and outcome possible.